Beniamino Murgante Osvaldo Gervasi
Andrés Iglesias David Taniar
Bernady O. Apduhan (Eds.)

LNCS 6785

Computational Science and Its Applications – ICCSA 2011

International Conference
Santander, Spain, June 2011
Proceedings, Part IV

Springer

Lecture Notes in Computer Science 6785

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Beniamino Murgante Osvaldo Gervasi
Andrés Iglesias David Taniar
Bernady O. Apduhan (Eds.)

Computational Science and Its Applications - ICCSA 2011

International Conference
Santander, Spain, June 20-23, 2011
Proceedings, Part IV

Springer

Volume Editors

Beniamino Murgante
Basilicata University Potenza, Italy
E-mail: beniamino.murgante@unibas.it

Osvaldo Gervasi
University of Perugia, Italy
E-mail: osvaldo@unipg.it

Andrés Iglesias
University of Cantabria, Santander, Spain
E-mail: iglesias@unican.es

David Taniar
Monash University, Clayton, VIC, Australia
E-mail: david.taniar@infotech.monash.edu.au

Bernady O. Apduhan
Kyushu Sangyo University
Fukuoka, Japan
E-mail: bob@is.kyusan-u.ac.jp

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21897-2 e-ISBN 978-3-642-21898-9
DOI 10.1007/978-3-642-21898-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929636

CR Subject Classification (1998): C.2, H.4, F.2, H.3, D.2, C.2.4, F.1, H.5

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

These multiple volumes (LNCS volumes 6782, 6783, 6784, 6785 and 6786) consist of the peer-reviewed papers from the 2011 International Conference on Computational Science and Its Applications (ICCSA 2011) held in Santander, Spain during June 20-23, 2011. ICCSA 2011 was a successful event in the International Conferences on Computational Science and Its Applications (ICCSA) conference series, previously held in Fukuoka, Japan (2010), Suwon, South Korea (2009), Perugia, Italy (2008), Kuala Lumpur, Malaysia (2007), Glasgow, UK (2006), Singapore (2005), Assisi, Italy (2004), Montreal, Canada (2003), and (as ICCS) Amsterdam, The Netherlands (2002) and San Francisco, USA (2001).

Computational science is a main pillar of most of the present research, as well as industrial and commercial activities and plays a unique role in exploiting ICT innovative technologies. The ICCSA conferences have been providing a venue to researchers and industry practitioners to discuss new ideas, to share complex problems and their solutions, and to shape new trends in computational science.

Apart from the general tracks, ICCSA 2011 also included 31 special sessions and workshops, in various areas of computational science, ranging from computational science technologies to specific areas of computational science, such as computer graphics and virtual reality. We accepted 52 papers for the general track, and 210 in special sessions and workshops. These represent an acceptance rate of 29.7%. We would like to show our appreciations to the Workshop and Special Session Chairs and co-Chairs.

The success of the ICCSA conference series, in general, and ICCSA 2011, in particular, is due to the support of many people: authors, presenters, participants, keynote speakers, Session Chairs, Organizing Committee members, student volunteers, Program Committee members, International Liaison Chairs, and people in other various roles. We would like to thank them all. We would also like to thank Springer for their continuous support in publishing ICCSA conference proceedings.

June 2011

Osvaldo Gervasi
David Taniar

Message from the ICCSA 2011 General Chairs

These five volumes contain an outstanding collection of refereed papers selected for the 11th International Conference on Computational Science and Its Applications, ICCSA 2011, held in Santander (Spain), June 20-23, 2011. We cordially invite you to visit the ICCSA website `http://www.iccsa.org` where you can find all relevant information about this interesting and exciting event.

ICCSA 2011 marked the beginning of the second decade of this conference series. Previous editions in this series of highly successful International Conferences on Computational Science and Its Applications (ICCSA) were held in Fukuoka, Japan (2010), Suwon, Korea (2009), Perugia, Italy (2008), Kuala Lumpur, Malaysia (2007), Glasgow, UK (2006), Singapore (2005), Assisi, Italy (2004), Montreal, Canada (2003), and (as ICCS) Amsterdam, The Netherlands (2002) and San Francisco, USA (2001).

As we enter the second decade of ICCSA, we realize the profound changes and spectacular advances in the world of computational science. This discipline plays a unique role in fostering new technologies and knowledge, and is crucial for most of the present research, and industrial and commercial activities. We believe that ICCSA has contributed to this change by offering a real opportunity to explore innovative approaches and techniques to solve complex problems. Reciprocally, the computational science community has enthusiastically embraced the successive editions of ICCSA, thus contributing to making ICCSA a focal meeting point for those interested in innovative, cutting-edge research about the latest and most exciting developments in the field. We are grateful to all those who have contributed to the current success of ICCSA with their continued support over the past ten years.

ICCSA 2011 would not have been made possible without the valuable contribution from many people. We would like to thank all session organizers for their diligent work, which further enhanced the conference levels and all reviewers for their expertise and generous effort which led to a very high quality event with excellent papers and presentations. We especially recognize the contribution of the Program Committee and Local Organizing Committee members for their tremendous support and for making this congress a very sucessful event.

We would like to sincerely thank our keynote speakers, who willingly accepted our invitation and shared their expertise through illuminating talks, helping us to fully meet the conference objectives.

We highly appreciate the University of Cantabria for their enthusiastic acceptance to host the conference on its main campus, their logistic assistance and additional financial support. The conference was held in the Faculty of Sciences of the University of Cantabria. We thank the Dean of the Faculty of Sciences, Ernesto Anabitarte, for his support before and during the congress, and for providing the venue of the conference and the use of all needed facilities.

ICCSA 2011 was jointly organized by the Department of Applied Mathematics and Computational Sciences and the Department of Mathematics, Statistics and Computation of the University of Cantabria, Spain. We thank both departments for their encouranging support of this conference from the very beginning. We would like to express our gratitude to the Local Organizing Committee for their persistent and enthusiastic work towards the success of this conference.

We owe special thanks to all our sponsors: the Faculty of Sciences, the University of Cantabria, the Municipality of Santander, the Regional Government of Cantabria and the Spanish Ministry of Science and Innovation, for their continuous support without which this conference would not be possible. We also thank our publisher, Springer, for their acceptance to publish the proceedings and for their kind assistance and cooperation during the editing process.

Finally, we thank all authors for their submissions and all conference attendants for making ICCSA 2011 truly an excellent forum on computational science, facilitating exchange of ideas, fostering new collaborations and shaping the future of this exciting field. Last, but certainly not least, we wish to thank our readers for their interest in these proceedings. We really hope you find in these pages interesting material and fruitful ideas for your future work.

June 2011

Andrés Iglesias
Bernady O. Apduhan

The Wisdom of Ancient Masters

In 1879, Marcelino Sanz de Sautuola and his young daugther María incidentally noticed that the ceiling of the Altamira cave was covered by images of bisons and other animals, some as old as between 25,000 and 35,000 years. They had discovered what came to be called the Sistine Chapel of Paleolithic Art. When the discovery was first made public in 1880, many experts rejected it under the belief that prehistoric man was unable to produce such beautiful and ellaborated paintings. Once their authenticity was later confirmed, it changed forever our perception of prehistoric human beings.

Today, the cave of Altamira and its paintings are a symbol of the wisdom and ability of our ancient ancestors. They remind us that our current technological development is mostly based on the work, genius and efforts of our predecessors over many generations.

The cave of Altamira (UNESCO World Heritage Site) is located in the region of Cantabria, near the city of Santander (ICCSA 2011 conference venue). The original cave is closed to the public for preservation, but conference attendees visited the "Neocave", an exact reproduction of the original space with all its cracks and textures and the permanent exhibition "The Times of Altamira", which introduces visitors to the prehistory of the peninsula and rupestrian art.

"After Altamira, all is decadence" (Pablo Picasso, famous Spanish painter)

ICCSA 2011 Welcome Message

Welcome to the proceedings of the 11th International Conference on Computational Science and Its Applications, ICCSA 2011, held in Santander, Spain.

The city of Santander is located in the self-governed region of Cantabria, on the northern coast of Spain between Asturias and the Basque Country. This beautiful region of half a million inhabitants is on the shores of the Cantabrian Sea and is crossed by a mountain range. The shores and inland valleys offer a wide variety of landscapes as a consequence of the mild, moist climate of so-called Green Spain. The coastal landscape of beaches, bays and cliffs blends together with valleys and highland areas. All along the coast there are typical traditional fishing ports and innumerable diverse beaches of soft white sand.

However, Cantabria's attractions are not limited to its natural treasures. History has provided a rich artistic and cultural heritage found in towns and villages that are outstanding in their own right. The archaeological remains and historic buildings bear the mark of a unique history starting with the world-famous Altamira cave paintings, a veritable shrine to the prehistoric age. In addition, there are remarkable remains from the Romans, the Mozarabic presence and the beginnings of the Reconquest of Spain, along with an artistic heritage of Romanesque, Gothic and Baroque styles. Examples include the Prehistoric Era (the Altamira and Puente Viesgo Caves), Roman ruins such as those of Julióbriga, medieval settlements, such as Santillana del Mar, and several examples of the civil and religious architecture of the nineteenth and twentieth centuries.

The surrounding natural landscape and the historical importance of many of its villages and buildings make this region very appealing for tourism, especially during the spring and summer seasons, when the humid, mild weather gives the region a rich and varied nature of woods and meadows. At the time of the conference, attendees enjoyed the gentle climate (with temperatures averaging 18-20 degrees Celsius) and the longest days of the year. They found themselves waiting for sunset at the beach at about 11 pm!

Capital of the autonomous region of Cantabria, the city of Santander is also a very popular destination for tourism. Based around one of the most beautiful bays in the world, this modern city is famous for its sparkling beaches of yellow sand and clean water, the hospitality of its people and the high reputation of its celebrated gastronomy, mostly based on fish and shellfish. With a population of about 200,000 inhabitants, Santander is a very safe city, with a vibrant tourist scene filled with entertainment and a high quality of life, matching the best standards in the world. The coastal side of the city boasts a long string of top–quality beaches and recreational areas, such as the Magdalena Peninsula, the Sardinero and Mataleñas Park. There are several beaches and harbors limiting the city on the northern side, toward the southern part there is the old city

center and a bit further on the green mountains. We proudly say that Santander is between the blue and the green.

The University of Cantabria (in Spanish, *the Universidad de Cantabria, UC*) is the only public university in Cantabria, Spain. It was founded in 1972 and is organized in 12 faculties and schools. With about 13,000 students and 1,000 academic staff, the University of Cantabria is one of the most reputed universities in the country, ranking in the highest positions of Spanish universities in relation to its size. Not surprisingly, it was selected as a Campus of International Excellence by the Spanish Government in 2009.

Besides the technical sessions and presentations, ICCSA 2011 provided an interesting, must-attend social program. It started with a Welcome Reception at the Royal Palace of the Magdalena (Sunday 19), the most emblematic building of Santander and also the most visited place in the city. The royal family used the palace during the period 1913–1930 as a base for numerous recreational and sporting activities, and the king sometimes also held government meetings at the property. Conference delegates had the wonderful opportunity to visit this splendid palace, enjoy the magnificent views and see some rooms where royalty lived. The Gala Dinner (Tuesday 21) took place at the Grand Casino, in the "Sardinero" area, a regal, 1920's building with large windows and spacious terraces offering superb views of the Sardinero beach. The Casino was King Alfonso XIII and Queen Victoria Eugenia's main place of entertainment during their summer holidays in the city between 1913 and 1930. The gala also included some cultural and musical events. Finally, a half-day conference tour (Wednesday 22) covered the "live museum" of the Middle Ages, Santillana del Mar (a medieval town with cobbled streets, declared "Site of Artistic and Historical Importance" and one of the best-known cultural and tourist centers in Cantabria) and the Altamira Neocave, an exact reproduction of the original Altamira cave (now closed to the public for preservation) with all its cracks and textures and the permanent exhibition "The Times of Altamira", which introduces visitors to the prehistory of the peninsula and rupestrian art.

To close the conference, attendees could join the people of Santander for St. John's day, celebrated in the night between June 23 and 24 to commemorate the summer solstice with bonfires on the beach.

We believe that all these attractions made the conference an unforgettable experience.

On behalf of the Local Organizing Committee members, I thank all attendees fot their visit.

June 2011 Andrés Iglesias

Message from the Chairs of the Session: 6th International Workshop on "Geographical Analysis, Urban Modeling, Spatial Statistics" (GEOG-AN-MOD 2011)

During the past few decades the main problem in geographical analysis was the lack of spatial data availability. Nowadays the wide diffusion of electronic devices containing geo-referenced information generates a great production of spatial data. Volunteered geographic information activities (e.g., Wikimapia, OpenStreetMap), public initiatives (e.g., spatial data infrastructures, geo-portals) and private projects (e.g., Google Earth, Microsoft Virtual Earth, etc.) produced an overabundance of spatial data, which, in many cases, do not help the efficiency of decision processes. The increase of geographical data availability has not been fully coupled by an increase of knowledge to support spatial decisions.

The inclusion of spatial simulation techniques in recent GIS software favored the diffusion of these methods, but in several cases led to mechanisms based on which buttons have to be pressed without having geography or processes in mind. Spatial modeling, analytical techniques and geographical analyses are therefore required in order to analyze data and to facilitate the decision process at all levels, with a clear identification of the geographical information needed and reference scale to adopt. Old geographical issues can find an answer thanks to new methods and instruments, while new issues are developing, challenging researchers for new solutions. This workshop aims at contributing to the development of new techniques and methods to improve the process of knowledge acquisition.

Conference themes include:

Geostatistics and spatial simulation
Agent-based spatial modeling
Cellular automata spatial modeling
Spatial statistical models
Space-temporal modeling
Environmental modeling
Geovisual analytics, geovisualization, visual exploratory data analysis
Visualization and modeling of track data
Spatial optimization
Interaction simulation models
Data mining, spatial data mining
Spatial data warehouse and spatial OLAP
Integration of spatial OLAP and spatial data mining
Spatial decision support systems

Spatial multicriteria decision analysis
Spatial rough set
Spatial extension of fuzzy set theory
Ontologies for spatial analysis
Urban modeling
Applied geography
Spatial data analysis
Dynamic modeling
Simulation, space-time dynamics, visualization and virtual reality.

Giuseppe Borruso
Beniamino Murgante
Stefania Bertazzon

Message from the Chairs of the Session: "Cities, Technologies and Planning" (CTP 2011)

'Share' term has turned into a key issue of many successful initiatives in recent times. Following the advent of Web 2.0, positive experiences based on mass collaboration generated "Wikinomics" hnd ave become 'Socialnomics", where 'Citizens are voluntary sensors'.

During the past few decades, the main issue in GIS implementation has been the availability of sound spatial information. Nowadays, the wide diffusion of electronic devices providing geo-referenced information resulted in the production of extensive spatial information datasets. This trend has led to "GIS wikification", where mass collaboration plays a key role in the main components of spatial information frameworks (hardware, software, data, and people).

Some authors (Goodchild, 2007) talk about 'volunteered geographic information' (VGI), as the harnessing of tools to create, assemble, and disseminate geographic information provided by individuals voluntarily creating their own contents by marking the locations of occurred events or by labeling certain existing features not already shown on a map. The term "neogeography" is often adopted to describe peoples activities when using and creating their own maps, geo-tagging pictures, movies, websites, etc. It could be defined as a new bottom up approach to geography prompted by users, therefore introducing changes in the roles of traditional' geographers and consumers' of geographical contents themselves. The volunteered approach has been adopted by important American organizations, such as US Geological Survey, US Census Bureau, etc. While technologies (e.g. GPS, remote sensing, etc.) can be useful in producing new spatial data, volunteered activities are the only way to update and describe such data. If spatial data have been produced in various ways, remote sensing, sensor networks and other electronic devices generate a great flow of relevant spatial information concerning several aspects of human activities or of environmental phenomena monitoring. This 'information-explosion era' is characterized by a large amount of information produced both by human activities and by automated systems; the capturing and the manipulation of this information leads to 'urban computing' and represents a sort of bridge between computers and the real world, accounting for the social dimension of human environments. This technological evolution produced a new paradigm of urban development, called 'u-City'. Such phenomena offer new challenges to scholars (geographers, engineers, planners, economists, sociologists, etc.) as well as to spatial planners in addressing spatial issues and a wealth of brand-new, updated data, generally created by people who are interested in geographically related phenomena. As attention is to date dedicated to visualization and content creation, little has been done from the spatial analytical point of view and in involving users as citizens in participatory geographical activities.

Conference themes include:

SDI and planning
Planning 2.0, participation 2.0
Urban social networks, urban sensing
E-democracy, e-participation, participatory GIS
Technologies for e-participation, policy modeling, simulation and visualization
Second Life and participatory games
Ubiquitous computing environment; urban computing; ubiquitous-city
Neogeography
Collaborative mapping
Geotagging
Volunteered geographic information
Crowdsourcing
Ontologies for urban planning
City Gml
Geo-applications for mobile phones
Web 2.0, Web 3.0
Wikinomics, socialnomics
WikiCities
Maps mash up
Tangible maps and planning
Augmented reality,
Complexity assessment and mapping

Giuseppe Borruso
Beniamino Murgante

Message from the Chairs of the Session: 11^{th} Annual International Workshop on "Computational Geometry and Applications" (CGA 2011)

The 11th International Workshop on Computational Geometry and Applications CGA 2011, held in conjunction with the International Conference on Computational Science and Applications, took place in Santander, Spain. The workshop has run annually since it was founded in 2001, and is intended as an international forum for researchers in computational geometry and related areas, with the goal of advancing the state of research in computational geometry and related disciplines. This year, the workshop was chaired for 11th year by CGA workshop series Founding Chair Marina Gavrilova, University of Calgary, joined by co-Chair Ovidiu Daescu, University of Texas at Dallas. Selected papers from the previous CGA Workshops have appeared in special issues in the following highly regarded journals: *International Journal of Computational Geometry and Applications*, Springer (three special issues), *International Journal of Computational Science and Engineering* (IJCSE), Journal of CAD/CAM , *Transactions on Computational Sciences*, Springer. A special issue comprising best papers presented at CGA 2011 is currently being planned.

The workshop attracts international attention and receives papers presenting high-quality original research in the following tracks:

- Theoretical computational geometry
- Applied computational geometry
- Optimization and performance issues in geometric algorithms implementation Workshop topics of interest include:
- Design and analysis of geometric algorithms
- Geometric algorithms in path planning and robotics
- Computational geometry in biometrics
- Intelligent geometric computing
- Geometric algorithms in computer graphics and computer vision
- Voronoi diagrams and their generalizations
- 3D Geometric modeling
- Geometric algorithms in geographical information systems
- Algebraic geometry
- Discrete and combinatorial geometry
- Implementation issues and numerical precision
- Applications in computational biology, physics, chemistry, geography, medicine, education
- Visualization of geometric algorithms

CGA 2011 was located in beautiful Santander, Cantabria, Spain. Santander, the capital city of Cantabria, is located on the northern coast of Spain, between Asturias and the Basque Country overlooking the Cantabrian Sea, and is surrounded by beaches. The conference preceded the Spanish Meeting on Computational Geometry, which took place in Madrid, facilitating interested researchers to attend both events. The 14 articles presented in this Springer LNCS proceeding volume represent papers selected fromi a large number of submissions to this year's workshop. We would like to express our sincere gratitude to the following International Program Committee members who performed their duties diligently and provided constructive feedback for authors to improve on their presentation:

Tetsuo Asano (Japan Advanced Institute of Science and Technology, Japan)
Sergei Bereg (University of Texas at Dallas, USA)
Karoly Bezdek (University of Calgary, Canada)
Ovidiu Daescu (University of Texas at Dallas, USA)
Mirela Damian (Villanova University, USA)
Tamal Dey (Ohio State University, USA)
Marina L. Gavrilova (University of Calgary, Canada)
Christopher Gold (University of Glamorgan, UK)
Hisamoto Hiyoshi (Gunma University, Japan)
Andrés Iglesias (University of Cantabria, Spain)
Anastasia Kurdia (Smith College, USA)
Deok-Soo Kim (Hanyang University, Korea)
Ivana Kolingerova (Unversity of West Bohemia, Czech Republic)
Nikolai Medvedev (Novosibirsk Russian Academy of Science, Russia)
Asish Mukhopadhyay (University of Windsor, Canada)
Dimitri Plemenos (Université de Limoges, France)
Val Pinciu (Southern Connecticut State University, USA)
Jon Rokne (University of Calgary, Canada)
Carlos Seara (Universitat Politecnica de Catalunya, Spain)
Kokichi Sugihara (University of Tokyo, Japan)
Vaclav Skala (University of West Bohemia, Czech Republic)
Muhammad Sarfraz (KFUPM, Saudi Arabia)
Alexei Sourin (Nanyang Technological University, Singapore)
Ryuhei Uehara (Japan Advanced Institute of Science and Technology, Japan)
Chee Yap (New York University, USA)
Kira Vyatkina (Sanct Petersburg State University, Russia)

We also would like to acknowledge the independent referees, ICCSA 2011 organizers, sponsors, volunteers, and Springer for their continuing collaboration and support.

Marina C. Gavrilova
Ovidiu Daescu

Message from the Chair of the Session: 3^{rd} International Workshop on "Software Engineering Processes and Applications" (SEPA 2011)

The Third International Workshop on Software Engineering Processes and Applications (SEPA 2011) covered the latest developments in processes and applications of software engineering. SEPA includes process models, agile development, software engineering practices, requirements, system and design engineering including architectural design, component level design, formal methods, software modeling, testing strategies and tactics, process and product metrics, Web engineering, project management, risk management, and configuration management and all those areas which are related to the processes and any type of software applications. This workshop attracted papers from leading researchers in the field of software engineering and its application areas. Seven regular research papers were accepted as follows.

Sanjay Misra, Ibrahim Akman and Ferid Cafer presented a paper on "A Multi-Paradigm Complexity Metric(MCM)" The authors argued that there are not metrics in the literature for multi-paradigm. MCM is developed by using function points and procedural and object–oriented language's features. In this view, MCM involves most of the factors which are responsible for the complexity of any multi-paradigm language. MCM can be used for most programming paradigms, including both procedural and object–oriented languages.

Mohamed A. El-Zawawy's paper entitled 'Flow Sensitive-Insensitive Pointer Analysis Based Memory Safety for Multithreaded Programs' presented approaches for the pointer analysis and memory safety of multithreaded programs as simply structured type systems. The author explained that in order to balance accuracy and scalability, the proposed type system for pointer analysis of multithreaded programs is flow-sensitive, which invokes another flow-insensitive type system for parallel constructs.

Cesar Pardo, Francisco Pino, Felix Garcia, Francisco Romero, Mario Piattini, and Maria Teresa Baldassarre presented their paper entitled 'HProcessTOOL: A Support Tool in the Harmonization of Multiple Reference Models'. The authors have developed the tool HProcessTOOL, which guides harmonization projects by supporting specific techniques, and supports their management by controlling and monitoring the resulting harmonization projects. The validation of the tool is performed by two case studies.

Wasi Haider Butt, Sameera Amjad and Farooque Azam presented a paper on 'Requirement Conflicts Resolution: Using Requirement Filtering and Analysis'. The authors presented a systematic approach toward resolving software requirements spanning from requirement elicitation to the requirement analysis

activity of the requirement engineering process. The authors developed a model 'conflict resolution strategy' (CRS) which employs a requirement filter and an analysis strategy for resolving any conflict arising during software development. They also implemented their model on a real project.

Rajesh Prasad, Suneeta Agarwal, Anuj Kumar Sharma, Alok Singh and Sanjay Misra presented a paper on 'Efficient Algorithm for Detecting Parameterized Multiple Clones in a Large Software System'. In this paper the authors have tried to solve the word length problem in a bit-parallel parameterized matching by extending the BLIM algorithm of exact string matching. The authors further argued that the extended algorithm is also suitable for searching multiple patterns simultaneously. The authors presented a comparison in support of their algorithm.

Takahiro Uchiya and Tetsuo Kinoshita presented the paper entitled 'Behavior Analyzer for Developing Multiagent Systems on Repository-Based Multiagent Framework'. In this paper the authors proposed an interactive design environment of agent system (IDEA) founded on an agent-repository-based multiagent framework. They focused on the function of the behavior analyzer for developing multiagent systems and showed the effectiveness of the function.

Jose Alfonso Aguilar, Irene Garrigos, and Jose-Norberto Mazon presented a paper on 'Impact Analysis of Goal-Oriented Requirements in Web Engineering'. This paper argues that Web developers need to know dependencies among requirements to ensure that Web applications finally satisfy the audience. The authors developed an algorithm to deal with dependencies among functional and non-functional requirements so as to understand the impact of making changes when developing a Web application.

Sanjay Misra

Message from the Chair of the Session: 2^{nd} International Workshop on "Software Quality" (SQ 2011)

Following the success of SQ 2009, the Second International Workshop on "Software Quality" (SQ 2011) was organized in conjunction with ICCSA 2011. This workshop extends the discussion on software quality issues in the modern software development processes. It covers all the aspects of process and product quality, quality assurance and standards, quality planning, quality control and software quality challenges. It also covers the frontier issues and trends for achieving the quality objectives. In fact this workshop covers all areas, that are concerned with the quality issue of software product and process. In this workshop, we featured nine articles devoted to different aspects of software quality.

Roberto Espinosa, Jose Zubcoff, and Jose-Norberto Mazon's paper entitled "A Set of Experiments to Consider Data Quality Criteria in Classification Techniques for Data Mining" analyzed data–mining techniques to know the behavior of different data quality criteria from the sources. The authors have conducted a set of experiments to assess three data quality criteria: completeness, correlation and balance of data.

In their paper, Ivaylo Spassov, Valentin Pavlov, Dessislava Petrova-Antonova, and Sylvia Ilieva's have developed a tool "DDAT: Data Dependency Analysis Tool for Web Service Business Processes". The authors have implemented and shown experimental results from the execution of the DDAT over BPEL processes.

Filip Radulovic and Raul Garca-Castro presented a paper on "Towards a Quality Model for Semantic Technologies". The authors presented some well-known software quality models, after which a quality model for semantic technologies is designed by extending the ISO 9126 quality model.

Luis Fernandez-Sanz and Sanjay Misra authored the paper "Influence of Human Factors in Software Quality and Productivity". The authors first analyzed the existing contributions in the area and then presented empirical data from specific initiatives to know more about real practices and situations in software organizations.

Eudisley Anjos, and Mario Zenha-Rela presented a paper on "A Framework for Classifying and Comparing Software Architecture Tools for Quality Evaluation". This framework identifies the most relevant features for categorizing different architecture evaluation tools according to six different dimensions. The authors reported that the attributes that a comprehensive tool should support include: the ability to handle multiple modeling approaches, integration with the industry standard UML or specific ADL, support for trade–off analysis of

competing quality attributes and the reuse of knowledge through the build-up of new architectural patterns.

Hendrik Decker presented a paper on "Causes of the Violation of Integrity Constraints for Supporting the Quality of Databases". He presented a quality metric with the potential of more accuracy by measuring the causes. He further argued that such measures also serve for controlling quality impairment across updates.

Csaba Nagy, Laszlo Vidacs , Rudolf Ferenc, Tibor Gyimothy Ferenc Kocsis, and Istvan Kovacs's presented a paper on "Complexity measures in a 4GL environment". The authors discussed the challenges in adopting the metrics from 3GL environments. Based on this, they presented a complexity measure in 4GL environments. They performed the experimentations and demonstrated the results.

Lukasz Radlinski's paper on "A Framework for Integrated Software Quality Prediction Using Bayesian Nets" developed a framework for integrated software quality prediction. His framework is developed and formulated using a Bayesian net, a technique that has already been used in various software engineering studies. The author argues that his model may be used in decision support for software analysts and managers.

Seunghun Park, Sangyoon Min, and Doohwan Bae authored the paper entitled "Process Instance Management Facilities Based on the Meta-Process Models". Based on the metar-process models, the authors proposed a process model and two types of process instance models: the structural instance model and the behavioral instance model. The authors' approach enables a project manager to analyze structural and behavioral properties of a process instance and allows a project manager to make use of the formalism for management facilities without knowledge of the formalism.

Sanjay Misra

Message from the Chairs of the Session: "Remote sensing Data Analysis, Modeling, Interpretation and Applications: From a Global View to a Local Analysis" (RS 2011)

Remotely sensed data provide temporal and spatial consistent measurements useful for deriving information on the dynamic nature of Earth surface processes (sea, ice, land, atmosphere), detecting and identifying land changes, discovering cultural resources, studying the dynamics of urban expansions. Thanks to the establishment and maintenance of long-term observation programs, presently a huge amount of multiscale and multifrequency remotely sensed data are available.

To fully exploit such data source for various fields of application (environmental, cultural heritage, urban analysis, disaster management) effective and reliable data processing, modeling and interpretation are required. This session brought together scientists and managers from the fields of remote sensing, ICT, geospatial analysis and modeling, to share information on the latest advances in remote sensing data analysis, product development, validation and data assimilation.

Main topics included:

Remotely sensed data – Multispectral satellite : from medium to very high spatial resolution; airborne and spaceborne Hypespectral data; open data source (Modis, Vegetation, etc..); airborne Laser Scanning; airborne and spaceborne Radar imaging; thermal imaging; declassified Intelligence Satellite Photographs (Corona, KVR); ground remote sensing

Methods and procedures – change detection; classification Data fusion / Data integration; data mining; geostatistics and Spatial statistics; image processing; image interpretation; linear and on linear statistical analysis; segmentation Pattern recognition and edge detection; time space modeling

Fields of application and products – archaeological site discovery; cultural Heritage management; disaster management; environmental sciences; mapping Landscape and digital elevation models; land cover analysis; open source softwares; palaeoenvironmental studies; time series

Nicola Masini
Rosa Lasaponara

Message from the Chairs of the Session: "Approximation, Optimization and Applications" (AOA 2011)

The objective of the session Approximation, Optimization and Applications during the 11th International Conference on Computational Science and Its Applications was to bring together scientists working in the areas of Approximation Theory and Numerical Optimization, including their applications in science and engineering.

Hypercomplex function theory, renamed Clifford analysis in the 1980s, studies functions with values in a non-commutative Clifford algebra. It has its roots in quaternionic analysis, developed as another generalization of the classic theory of functions of one complex variable compared with the theory of functions of several complex variables. It is well known that the use of quaternions and their applications in sciences and engineering is increasing, due to their advantages for fast calculations in 3D and for modeling mathematical problems. In particular, quasi-conformal 3D-mappings can be realized by regular (monogenic) quaternionic functions. In recent years the generalization of classical polynomials of a real or complex variable by using hypercomplex function theoretic tools has been the focus of increased attention leading to new and interesting problems. All these aspects led to the emergence of new software tools in the context of quaternionic or, more generally, Clifford analysis.

Irene Falcão
Ana Maria A.C. Rocha

Message from the Chair of the Session: "Symbolic Computing for Dynamic Geometry" (SCDG 2011)

The papers comprising in the Symbolic Computing for Dynamic Geometry technical session correspond to talks delivered at the conference. After the evaluation process, six papers were accepted for oral presentation, according to the recommendations of the reviewers. Two papers, "Equal bisectors at a vertex of a triangle" and "On Equivalence of Conditions for a Quadrilateral to Be Cyclica", study geometric problem by means of symbolic approaches.

Another contributions deal with teaching ("Teaching geometry with TutorMates" and "Using Free Open Source Software for Intelligent Geometric Computing"), while the remaining ones propose a framework for the symbolic treatment of dynamic geometry ("On the Parametric Representation of Dynamic Geometry Constructions") and a formal library for plane geometry ("A Coq-based Library for Interactive and Automated Theorem Proving in Plane Geometry").

Francisco Botana

Message from the Chairs of the Session: "Computational Design for Technology Enhanced Learning" (CD4TEL 2011)

Providing computational design support for orchestration of activities, roles, resources, and systems in technology-enhanced learning (TEL) is a complex task. It requires integrated thinking and interweaving of state-of-the-art knowledge in computer science, human–computer interaction, pedagogy, instructional design and curricular subject domains. Consequently, even where examples of successful practice or even standards and specifications like IMS learning design exist, it is often hard to apply and (re)use these efficiently and systematically. This interdisciplinary technical session brought together practitioners and researchers from diverse backgrounds such as computer science, education, and cognitive sciences to share their proposals and findings related to the computational design of activities, resources and systems for TEL applications.

The call for papers attracted 16 high-quality submissions. Each submission was reviewed by three experts. Eventually, five papers were accepted for presentation. These contributions demonstrate different perspectives of research in the CD4TEL area, dealing with standardization in the design of game-based learning; the integration of individual and collaborative electronic portfolios; the provision of an editing environment for different actors designing professional training; a simplified graphical notation for modeling the flow of activities in IMS learning design units of learning; and a pattern ontology-based model to support the selection of good-practice scripts for designing computer–supported collaborative learning.

Michael Derntl
Manuel Caeiro-Rodríguez
Davinia Hernández-Leo

Message from the Chair of the Session: "Chemistry and Materials Sciences and Technologies" (CMST 2011)

The CMST workshop is a typical example of how chemistry and computer science benefit from mutual interaction when operating within a grid e-science environment. The scientific contributions to the workshop, in fact, contain clear examples of chemical problems solved by exploiting the extra power offered by the grid infrastructure to the computational needs of molecular scientists when trying to push ahead the frontier of research and innovation.

Ideal examples of this are the papers on the coulomb potential decomposition in the multiconfiguration time–dependent Hartree method, on the extension of the grid–empowered simulator GEMS to the a priori evaluation of the crossed beam measurements and on the evaluation of the concentration of pollutants when using a box model version of the Community Multiscale Air Quality Modeling System 4.7. Another example of such progress in computational molecular science is offered by the paper illustrating the utilization of a fault–tolerant workflow for the DL-POLY package for molecular dynamics studies.

At the same time molecular science studies are an excellent opportunity for investigating the use of new (single or clustered) GPU chips as in the case of the papers related to their use for computationally demanding quantum calculations of atom diatom reactive scattering. In addition, of particular interest are the efforts spent to develop tools for evaluating user and service quality to the end of promoting collaborative work within virtual organizations and research communities through the awarding and the redeeming of credits.

Antonio Laganà

Message from the Chairs of the Session: "Cloud for High Performance Computing" (C4HPC 2011)

On behalf of the Program Committee, it is a pleasure for us to introduce the proceedings of this First International Workshop on Cloud for High–Performance Computing held in Santander (Spain) in 2011 during the 11th International Conference on Computational Science and Its Applications. The conference joined high quality researchers around the world to present the latest results in the usage of cloud computing for high–performance computing.

High–performance computing, or HPC, is a great tool for the advancement of science, technology and industry. It intensively uses computing resources, both CPU and storage, to solve technical or scientific problems in minimum time. It also uses the most advanced techniques to achieve this objective and evolves along with computing technology as fast as possible. During the last few years we have seen the introduction of new hardware isuch as multi-core and GPU representing a formidable challenge for the scientific and technical developers that need time to absorb these additional characteristics. At the same time, scientists and technicians have learnt to make faster and more accurate measurements, accumulating a large set of data which need more processing capacity. While these new paradigms were entering the field of HPC, virtualization was suddenly introduced in the market, generating a new model for provisioning computing capacity: the cloud. Although conceptually the cloud is not completely new, because it follows the old dream of computing as a utility, it has introduced new characteristics such as elasticity, but at the cost of losing some performance.

Consequently, HPC has a new challenge: how to tackle or solve this reduction in performance while adapting methods to the elasticity of the new platform. The initial results show the feasibility of using cloud infrastructures to execute HPC applications. However, there is also some consensus that the cloud is not the solution for grand challenges, which will still require dedicated supercomputers. Although recently a cluster of more than 4000 CPUs has been deployed, there are still many technical barriers to allow technicians to use it frequently. This is the reason for this workshop which we had the pleasure of introducing.

This First International Workshop on Cloud for High–Performance Computing was an original idea of Osvaldo Gervasi. We were working on the proposal of a COST action devoted to the cloud for HPC which would link the main researchers in Europe. He realized that the technical challenges HPC has to solve in the next few years to use the Cloud efficiently, need the collaboration of as many scientists and technicians as possible as well as to rethink the way the applications are executed.

This first workshop, which deserves in the next ICCSA conferences, joined together experts in the field that presented high quality research results in the area. They include the first approximations of topology methods such as cellular data system to cloud to be used to process data. Data are also the main issue for the TimeCloud front end, an interface for time series analysis based on Hadop and Hbase, designed to work with massive datasets. In fact, cloud can generate such a large amount of data when accurate information about its infrastructure and executing applications is needed. This is the topic of the third paper which introduce LISA algorithm to tackle the problem of information retrieval in cloud environment where the methods must adapt to the elasticity, scalability and possibility of failure. In fact, to understand Cloud infrastructures, researchers and technicians will need these series of data as well as the usage of tools that allow better knowledge to be gained. In this sense, iCanCloud, a novel simulator of cloud infrastructures, is introduced presenting its results for the most used and cited service: Amazon.

We strongly believe that the reader will enjoy the selected papers, which represent only a minimal, but important, part of the effervescent activity in Cloud for HPC. This selection was only possible thanks to the members of the Program Committee, all of them supporting actively the initiative. We appreciate their commitment to the workshop. Also, we want to thank all of the reviewers who kindly participated in the review of the papers and, finally, to all the scientists who submitted a paper, even if it was not accepted. We hope that they will have the opportunity to join us in the next editions.

Andrés Gomez
Osvaldo Gervasi

ICCSA 2011 Invited Speakers

Ajith Abraham
Machine Intelligence Research Labs, USA

Marina L. Gavrilova
University of Calgary, Canada

Yee Leung
The Chinese University of Hong Kong, China

Evolving Future Information Systems: Challenges, Perspectives and Applications

Ajith Abraham

Machine Intelligence Research Labs, USA
ajith.abraham@ieee.org

Abstract

We are blessed with the sophisticated technological artifacts that are enriching our daily lives and society. It is believed that the future Internet is going to provide us with the framework to integrate, control or operate virtually any device, appliance, monitoring systems, infrastructures etc. The challenge is to design intelligent machines and networks that could communicate and adapt according to the environment. In this talk, we first present the concept of a digital ecosystem and various research challenges from several application perspectives. Finally, we present some real–world applications.

Biography

Ajith Abraham received a PhD degree in Computer Science from Monash University, Melbourne, Australia. He is currently the Director of Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, USA, which has members from more than 75 countries. He serves/has served the editorial board of over 50 international journals and has also guest edited 40 special issues on various topics. He has authored/co-authored more than 700 publications, and some of the works have also won best paper awards at international conferences. His research and development experience includes more than 20 years in industry and academia. He works in a multidisciplinary environment involving machine intelligence, network security, various aspects of networks, e-commerce, Web intelligence, Web services, computational grids, data mining, and their applications to various real-world problems. He has given more than 50 plenary lectures and conference tutorials in these areas.

Dr. Abraham is the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing. He is a Senior Member of the IEEE, the IEEE Computer Society, the Institution of Engineering and Technology (UK) and the Institution of Engineers Australia (Australia). He is actively involved in the Hybrid Intelligent Systems (HIS), Intelligent Systems Design and Applications (ISDA), Information Assurance and Security (IAS), and Next–Generation Web Services Practices (NWeSP) series of international conferences, in addition to other conferences. More information can be found at: http://www.softcomputing.net.

Recent Advances and Trends in Biometric

Marina L. Gavrilova

Department of Computer Science, University of Calgary
marina@cpsc.ucalgary.ca

Extended Abstract

The area of biometric, without a doubt, is one of the most dynamic areas of interest, which recently has displayed a gamut of broader links to other fields of sciences. Among those are visualization, robotics, multi-dimensional data analysis, artificial intelligence, computational geometry, computer graphics, e-learning, data fusion and data synthesis. The theme of this keynote is reviewing the state of the art in multi-modal data fusion, fuzzy logic and neural networks and its recent connections to advanced biometric research.

Over the past decade, multimodal biometric systems emerged as a feasible and practical solution to counterweight the numerous disadvantages of single biometric systems. Active research intoi the design of a multimodal biometric system has started, mainly centered around: types of biometrics, types of data acquisition and decision-making processes. Many challenges originating from non-uniformity of biometric sources and biometric acquisition devices result in significant differences on which information is extracted, how is it correlated, the degree of allowable error, cost implications, ease of data manipulation and management, and also reliability of the decisions being made. With the additional demand of computational power and compact storage, more emphasis is shifted toward database design and computational algorithms.

One of the actively researched areas in multimodal biometric systems is information fusion. Which information needs to be fused and what level is needed to obtain the maximum recognition performance is the main focus of current research. In this talk I concentrate on an overview of the current trends in recent multimodal biometric fusion research and illustrate in detail one fusion strategy: rank level fusion. From the discussion, it is seen that rank level fusion often outperforms other methods, especially combined with powerful decision models such as Markov chain or fuzzy logic.

Another aspect of multi-modal biometric system development based on neural networks is discussed further. Neural networks have the capacity to simulate learning processes of a human brain and to analyze and compare complex patters, which can originate from either single or multiple biometric sources, with amazing precision. Speed and complexity have been the downsides of neural networks, however, recent advancements in the area, especially in chaotic neural networks, allow these barriers to be overcome.

The final part of the presentation concentrates on emerging areas utilizing the above developments, such as decision making in visualization, graphics, e-learning, navigation, robotics, and security of web-based and virtual worlds. The extent to which biometric advancements have an impact on these emerging areas makes a compelling case for the bright future of this area.

References

1. Ross, A., Nandakumar, K., and Jain, A.K., Handbook of multibiometrics, New York, Springer (2006).
2. Jain, A.K., Ross, A., Prabhakar, S., An introduction to biometric recognition, IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, 14 (1): 420 (2004)
3. Nandakumar, K., Jain, A.K., Ross, A., Fusion in multibiometric identification systems: What about the missing data?, in LNCS 5558: 743752, Springer (2009).
4. Monwar, M. M., and Gavrilova, M.L., A multimodal biometric system using rank level fusion approach, IEEE Trans. SMC - B: Cybernetics, 39(4): 867-878 (2009).
5. Monwar, M. M., and Gavrilova, M.L., Secured access control through Markov chain based rank level fusion method, in proc. of 5th Int. Conf. on Computer Vision Theory and Applications (VISAPP), 458-463, Angres, France (2010).
6. Monwar, M. M., and Gavrilova, M.L., FES: A system of combining face, ear and signature biometrics using rank level fusion, in proc. 5th IEEE Int. Conf. IT: New Generations, pp 922-927, (2008).
7. Wang, C., Gavrilova, M.L., Delaunay Triangulation Algorithm for Fingerprint Matching. ISVD'2006. pp.208 216
8. Wecker, L., Samavati, F.F., Gavrilova, M.L., Iris synthesis: a reverse subdivision application. GRAPHITE'2005. pp.121 125
9. Anikeenko, A.V., Gavrilova, M.L., Medvedev, N.N., A Novel Delaunay Simplex Technique for Detection of Crystalline Nuclei in Dense Packings of Spheres. ICCSA (1)'2005. pp.816 826
10. Luchnikov, V.A., Gavrilova, M.L., Medvedev, N.N., Voloshin, V. P., The Voronoi-Delaunay approach for the free volume analysis of a packing of balls in a cylindrical container. Future Generation Comp. Syst., 2002: 673 679
11. Frischholz, R., and Dieckmann, U., BioID: A multimodal biometric identification system, IEEE Computer, 33 (2): 64-68 (2000).
12. Latifi, S., Solayappan, N. A survey of unimodal biometric methods, in proc. of Int. Conf. on Security & Management, 57-63, Las Vegas, USA (2006).
13. Dunstone, T., and Yager, N., Biometric system and data analysis: Design, evaluation, and data mining. Springer, New York (2009).
14. Ho, T.K., Hull, J.J., and Srihari, S.N., Decision combination in multiple classifier systems, IEEE Trans. on Pattern Analysis and Machine Intelligence, 16 (1): 66-75 (1994)

Biography

Marina L. Gavrilova is an Associate Professor in the Department of Computer Science, University of Calgary. Prof. Gavrilova's research interests lie in the area of computational geometry, image processing, optimization, spatial and biometric modeling. Prof. Gavrilova is founder and co-director of two innovative research laboratories: the Biometric Technologies Laboratory: Modeling and Simulation and the SPARCS Laboratory for Spatial Analysis in Computational Sciences. Prof. Gavrilova publication list includes over 120 journal and conference papers, edited special issues, books and book chapters, including World Scientific Bestseller of the Month (2007) *Image Pattern Recognition: Synthesis and Analysis in Biometric* and the Springer book Computational Intelligence: A Geometry-Based Approach. Together with Dr. Kenneth Tan, Prof. Gavrilova founded the ICCSA series of successful international events in 2001. She founded and chaired the International Workshop on Computational Geometry and Applications for over ten years, was co-Chair of the International Workshop on Biometric Technologies BT 2004, Calgary, served as Overall Chair of the Third International Conference on Voronoi Diagrams in Science and Engineering (ISVD) in 2006, was Organizing Chair of WADS 2009 (Banff), and general chair of the International Conference on Cyberworlds CW2011 (October 4-6, Banff, Canada). Prof. Gavrilova is an Editor-in-Chief of the successful LNCS Transactions on Computational Science Journal, Springer-Verlag since 2007 and serves on the Editorial Board of the International Journal of Computational Sciences and Engineering, CAD/CAM Journal and Journal of Biometrics. She has been honored with awards and designations for her achievements and was profiled in numerous newspaper and TV interviews, most recently being chosen together with other outstanding Canadian scientists to be featured in the National Museum of Civilization and National Film Canada production.

Theories and Applications of Spatial-Temporal Data Mining and Knowledge Discovery

Yee Leung

The Chinese University of Hong Kong, China
yeeleung@cuhk.edu.hk

Abstract

Basic theories of knowledge discovery in spatial and temporal data are examined in this talk. Fundamental issues in the discovery of spatial structures and processes will first be discussed. Real-life spatial data mining problems are employed as the background on which concepts, theories and methods are scrutinized. The unraveling of land covers, seismic activities, air pollution episodes, rainfall regimes, epidemics, patterns and concepts hidden in spatial and temporal data are employed as examples to illustrate the theoretical arguments and algorithms performances. To round up the discussion, directions for future research are outlined.

Biography

Yee Leung is currently Professor of Geography and Resource Management at The Chinese University of Hong Kong. He is also the Associate Academic Director of the Institute of Space and Earth Information Science of The Chinese University of Hong Kong. He is adjunct professor of several universities in P.R. China. Professor Leung had also served on public bodies including the Town Planning Board and the Environmental Pollution Advisory Committee of Hong Kong SAR. He is now Chair of The Commission on Modeling Geographical Systems, International Geographical Union, and Chair of The Commission on Quantitative and Computational Geography of The Chinese Geographical Society. He serves on the editorial board of several international journals such as *Annals of Association of American Geographers*, *Geographical Analysis*, *GeoInformatica*, *Journal of Geographical Systems*, *Acta Geographica Sinica*, *Review of Urban and Regional Development Studies*, etc. Professor Leung is also Council member of The Society of Chinese Geographers.

Professor Leung carried out pioneer and influential research in imprecision/uncertainty analysis in geography, intelligent spatial decision support systems, geocomputation (particularly on fuzzy sets, rough sets, spatial statistics,

fractal analysis, neural networks and genetic algorithms), and knowledge discovery and data mining. He has obtained more than 30 research grants, authored and co-authored six books and over 160 papers in international journals and book chapters on geography, computer science, and information engineering. His landmark books are: *Spatial Analysis and Planning under Imprecision* (Elsevier, 1988), *Intelligent Spatial Decision Support Systems* (Springer-Verlag, 1997), and *Knowledge Discovery in Spatial Data* (Springer-Verlag, 2010).

Organization

ICCSA 2011 was organized by the University of Cantabria (Spain), Kyushu Sangyo University (Japan), the University of Perugia (Italy), Monash University (Australia) and the University of Basilicata (Italy).

Honorary General Chairs

Antonio Laganà	University of Perugia, Italy
Norio Shiratori	Tohoku University, Japan
Kenneth C.J. Tan	Qontix, UK

General Chairs

Bernady O. Apduhan	Kyushu Sangyo University, Japan
Andrés Iglesias	University of Cantabria, Spain

Program Committee Chairs

Osvaldo Gervasi	University of Perugia, Italy
David Taniar	Monash University, Australia

Local Arrangements Chairs

Andrés Iglesias	University of Cantabria, Spain (Chair)
Akemi Gálvez	University of Cantabria, Spain
Jaime Puig-Pey	University of Cantabria, Spain
Angel Cobo	University of Cantabria, Spain
José L. Montaña	University of Cantabria, Spain
César Otero	University of Cantabria, Spain
Marta Zorrilla	University of Cantabria, Spain
Ernesto Anabitarte	University of Cantabria, Spain
Unal Ufuktepe	Izmir University of Economics, Turkey

Workshop and Session Organizing Chair

Beniamino Murgante	University of Basilicata, Italy

International Liaison Chairs

Jemal Abawajy	Deakin University, Australia
Marina L. Gavrilova	University of Calgary, Canada
Robert C.H. Hsu	Chung Hua University,Taiwan
Tai-Hoon Kim	Hannam University, Korea
Takashi Naka	Kyushu Sangyo University, Japan

Awards Chairs

Wenny Rahayu	LaTrobe University, Australia
Kai Cheng	Kyushu Sangyo University, Japan

Workshop Organizers

Approaches or Methods of Security Engineering (AMSE 2011)

Tai-hoon Kim	Hannam University, Korea

Approximation, Optimization and Applications (AOA 2011)

Ana Maria A.C. Rocha	University of Minho, Portugal
Maria Irene Falcao	University of Minho, Portugal

Advances in Web–Based Learning (AWBL 2011)

Mustafa Murat Inceoglu	Ege University (Turkey)

Computational Aspects and Methods in Renewable Energies (CAMRE 2011)

Maurizio Carlini	University of Tuscia, Italy
Sonia Castellucci	University of Tuscia, Italy
Andrea Tucci	University of Tuscia, Italy

Computer–Aided Modeling, Simulation, and Analysis (CAMSA 2011)

Jie Shen	University of Michigan, USA

Computer Algebra Systems and Applications (CASA 2011)

Andrés Iglesias	University of Cantabria (Spain)
Akemi Gálvez	University of Cantabria (Spain)

Computational Design for Technology–Enhanced Learning: Methods, Languages, Applications and Tools (CD4TEL 2011)

Michael Derntl	University of Vienna, Austria
Manuel Caeiro-Rodriguez	University of Vigo, Spain
Davinia Hernandez-Leo	Universitat Pompeu Fabra, Spain

Computational Geometry and Applications (CGA 2011)

Marina L. Gavrilova	University of Calgary, Canada

Computer Graphics and Virtual Reality (CGVR 2011)

Osvaldo Gervasi	University of Perugia, Italy
Andrés Iglesias	University of Cantabria, Spain

Chemistry and Materials Sciences and Technologies (CMST 2011)

Antonio Laganà	University of Perugia, Italy

Consulting Methodology and Decision Making for Security Systems (CMDMSS 2011)

Sangkyun Kim	Kangwon National University, Korea

Cities, Technologies and Planning (CTP 2011)

Giuseppe Borruso	University of Trieste, Italy
Beniamino Murgante	University of Basilicata, Italy

Cloud for High–Performance Computing (C4HPC 2011)

Andrés Gomez	CESGA, Santiago de Compostela, Spain
Osvaldo Gervasi	University of Perugia, Italy

Future Information System Technologies and Applications (FISTA 2011)

Bernady O. Apduhan	Kyushu Sangyo University, Japan
Jianhua Ma	Hosei University, Japan
Qun Jin	Waseda University, Japan

Geographical Analysis, Urban Modeling, Spatial Statistics (GEOG-AN-MOD 2011)

Stefania Bertazzon	University of Calgary, Canada
Giuseppe Borruso	University of Trieste, Italy
Beniamino Murgante	University of Basilicata, Italy

International Workshop on Biomathematics, Bioinformatics and Biostatistics (IBBB 2011)

Unal Ufuktepe	Izmir University of Economics, Turkey
Andrés Iglesias	University of Cantabria, Spain

International Workshop on Collective Evolutionary Systems (IWCES 2011)

Alfredo Milani	University of Perugia, Italy
Clement Leung	Hong Kong Baptist University, China

Mobile Communications (MC 2011)

Hyunseung Choo	Sungkyunkwan University, Korea

Mobile Sensor and Its Applications (MSA 2011)

Moonseong Kim	Korean Intellectual Property Office, Korea

Mobile Systems and Applications (MoSA 2011)

Younseung Ryu	Myongji University, Korea
Karlis Kaugars	Western Michigan University, USA

Logical, Scientific and Computational Aspects of Pulse Phenomena in Transitions (PULSES 2011)

Carlo Cattani	University of Salerno, Italy
Cristian Toma	Corner Soft Technologies, Romania
Ming Li	East China Normal University, China

Resource Management and Scheduling for Future–Generation Computing Systems (RMS 2011)

Jemal H. Abawajy	Deakin University, Australia

Remote Sensing Data Analysis, Modeling, Interpretation and Applications: From a Global View to a Local Analysis (RS 2011)

Rosa Lasaponara	IRMMA, CNR, Italy
Nicola Masini	IBAM, CNR, Italy

Symbolic Computing for Dynamic Geometry (SCDG 2011)

Francisco Botana	Vigo University, Spain

Software Engineering Processes and Applications (SEPA 2011)

Sanjay Misra Atilim University, Turkey

Software Quality (SQ 2011)

Sanjay Misra Atilim University, Turkey

Tools and Techniques in Software Development Processes (TTSDP 2011)

Sanjay Misra Atilim University, Turkey

Virtual Reality in Medicine and Surgery (VRMS 2011)

Giovanni Aloisio University of Salento, Italy
Lucio T. De Paolis University of Salento, Italy

Wireless and Ad-Hoc Networking (WADNet 2011)

Jongchan Lee Kunsan National University, Korea
Sangjoon Park Kunsan National University, Korea

WEB 2.0 and Social Networks (Web2.0 2011)

Vidyasagar Potdar Curtin University of Technology, Australia

Workshop on Internet Communication Security (WICS 2011)

Josè Maria Sierra Camara University of Madrid, Spain

Wireless Multimedia Sensor Networks (WMSN 2011)

Vidyasagar Potdar Curtin University of Technology, Australia
Yan Yang Seikei University, Japan

Program Committee

Jemal Abawajy Deakin University, Australia
Kenneth Adamson Ulster University, UK
Michela Bertolotto University College Dublin, Ireland
Sandro Bimonte CEMAGREF, TSCF, France
Rod Blais University of Calgary, Canada
Ivan Blecic University of Sassari, Italy
Giuseppe Borruso Università degli Studi di Trieste, Italy
Martin Buecker Aachen University, Germany
Alfredo Buttari CNRS-IRIT, France
Yves Caniou Lyon University, France
Carlo Cattani University of Salerno, Italy
Mete Celik Erciyes University, Turkey

Alexander Chemeris	National Technical University of Ukraine "KPI", Ukraine
Min Young Chung	Sungkyunkwan University, Korea
Rosa Coluzzi	National Research Council, Italy
Stefano Cozzini	National Research Council, Italy
Josè A. Cardoso e Cunha	Universidade Nova de Lisboa, Portugal
Alfredo Cuzzocrea	University of Calabria, Italy
Frank Dévai	London South Bank University, UK
Rodolphe Devillers	Memorial University of Newfoundland, Canada
Pasquale Di Donato	Sapienza University of Rome, Italy
Carla Dal Sasso Freitas	UFRGS, Brazil
Prabu Dorairaj	NetApp, India/USA
Cherry Liu Fang	U.S. DOE Ames Laboratory, USA
Josè-Jesus Fernandez	National Centre for Biotechnology, CSIS, Spain
Francesco Gabellone	National Research Council, Italy
Akemi Galvez	University of Cantabria, Spain
Marina Gavrilova	University of Calgary, Canada
Jerome Gensel	LSR-IMAG, France
Andrzej M. Goscinski	Deakin University, Australia
Shanmugasundaram Hariharan	B.S. Abdur Rahman University, India
Hisamoto Hiyoshi	Gunma University, Japan
Fermin Huarte	University of Barcelona, Spain
Andres Iglesias	University of Cantabria, Spain
Peter Jimack	University of Leeds, UK
Qun Jin	Waseda University, Japan
Farid Karimipour	Vienna University of Technology, Austria
Baris Kazar	Oracle Corp., USA
Ivana Kolingerova	University of West Bohemia, Czech Republic
Dieter Kranzlmueller	LMU & LRZ Munich, Germany
Domenico Labbate	University of Basilicata, Italy
Antonio Laganà	University of Perugia, Italy
Rosa Lasaponara	National Research Council, Italy
Maurizio Lazzari	National Research Council, Italy
Cheng Siong Lee	Monash University, Australia
Sangyoun Lee	Yonsei University, Korea
Jongchan Lee	Kunsan National University, Korea
Clement Leung	Hong Kong Baptist University, Hong Kong
Chendong Li	University of Connecticut, USA
Ming Li	East China Normal University, China
Xin Liu	University of Calgary, Canada
Savino Longo	University of Bari, Italy
Tinghuai Ma	NanJing University of Information Science and Technology, China
Sergio Maffioletti	University of Zurich, Switzerland

Nicola Masini	National Research Council, Italy
Nirvana Meratnia	University of Twente, The Netherlands
Alfredo Milani	University of Perugia, Italy
Sanjay Misra	Atilim University, Turkey
Josè Luis Montaña	University of Cantabria, Spain
Beniamino Murgante	University of Basilicata, Italy
Jiri Nedoma	Academy of Sciences of the Czech Republic, Czech Republic
Laszlo Neumann	University of Girona, Spain
Kok-Leong Ong	Deakin University, Australia
Belen Palop	Universidad de Valladolid, Spain
Marcin Paprzycki	Polish Academy of Sciences, Poland
Eric Pardede	La Trobe University, Australia
Kwangjin Park	Wonkwang University, Korea
Vidyasagar Potdar	Curtin University of Technology, Australia
David C. Prosperi	Florida Atlantic University, USA
Wenny Rahayu	La Trobe University, Australia
Jerzy Respondek	Silesian University of Technology Poland
Alexey Rodionov	Institute of Computational Mathematics and Mathematical Geophysics, Russia
Jon Rokne	University of Calgary, Canada
Octavio Roncero	CSIC, Spain
Maytham Safar	Kuwait University, Kuwait
Haiduke Sarafian	The Pennsylvania State University, USA
Qi Shi	Liverpool John Moores University, UK
Dale Shires	U.S. Army Research Laboratory, USA
Carmelo Torre	Polytechnic of Bari, Italy
Giuseppe A. Trunfio	University of Sassari, Italy
Unal Ufuktepe	Izmir University of Economics, Turkey
Mario Valle	Swiss National Supercomputing Centre, Switzerland
Piero Giorgio Verdini	INFN Pisa and CERN, Italy
Andrea Vittadini	University of Padova, Italy
Koichi Wada	University of Tsukuba, Japan
Krzysztof Walkowiak	Wroclaw University of Technology, Poland
Robert Weibel	University of Zurich, Switzerland
Roland Wismüller	Universität Siegen, Germany
Markus Wolff	University of Potsdam, Germany
Mudasser Wyne	National University, USA
Chung-Huang Yang	National Kaohsiung Normal University, Taiwan
Xin-She Yang	National Physical Laboratory, UK
Salim Zabir	France Telecom Japan Co., Japan
Albert Y. Zomaya	University of Sydney, Australia

Sponsoring Organizations

ICCSA 2011 would not have been possible without tremendous support of many organizations and institutions, for which all organizers and participants of ICCSA 2011 express their sincere gratitude:

- The Department of Applied Mathematics and Computational Sciences, University of Cantabria, Spain
- The Department of Mathematics, Statistics and Computation, University of Cantabria, Spain
- The Faculty of Sciences, University of Cantabria, Spain
- The Vicerrector of Research and Knowledge Transfer, University of Cantabria, Spain
- The University of Cantabria, Spain
- The University of Perugia, Italy
- Kyushu Sangyo University, Japan
- Monash University, Australia
- The University of Basilicata, Italy
- Cantabria Campus Internacional, Spain
- The Municipality of Santander, Spain
- The Regional Government of Cantabria, Spain
- The Spanish Ministry of Science and Innovation, Spain
- GeoConnexion (`http://www.geoconnexion.com/`)
- Vector1 Media (`http://www.vector1media.com/`)

Table of Contents – Part IV

Workshop on Computer Aided Modeling, Simulation, and Analysis (CAMSA 2011)

Workshop on Mobile Sensor and Its Applications (MSA 2011)

Workshop on Computational Aspects and Methods in Renewable Energies (CAMRE 2011)

Workshop on Symbolic Computing for Dynamic Geometry (SCDG 2011)

Workshop on Wireless and Ad Hoc Networking (WADNet 2011)

Session on Computational Design for Technology Enhanced Learning: Methods, Languages, Applications and Tools (CD4TEL 2011)

Session on Virtual Reality in Medicine and Surgery (VRMS 2011)

Session on Logical, Scientific and Computational Aspects of Pulse Phenomena in Transitions (PULSES 2011)

On the Stability of Fully-Explicit Finite-Difference Scheme for Two-Dimensional Parabolic Equation with Nonlocal Conditions

Svajūnas Sajavičius

[1] Vilnius University, Naugarduko str. 24, LT-03225 Vilnius, Lithuania
Svajunas.Sajavicius@mif.vu.lt
http://www.mif.vu.lt/~svajunas
[2] Mykolas Romeris University, Ateities str. 20, LT-08303 Vilnius, Lithuania
Svajunas@mruni.eu

Abstract. We construct and analyse a fully-explicit finite-difference scheme for a two-dimensional parabolic equation with nonlocal integral conditions. The main attention is paid to the stability of the scheme. We apply the stability analysis technique which is based on the investigation of the spectral structure of the transition matrix of a finite-difference scheme and demonstrate that depending on the parameters of nonlocal conditions the proposed method can be stable or unstable. The results of numerical experiment with one test problem are also presented and they validate theoretical results.

Keywords: parabolic equation, nonlocal integral conditions, fully-explicit finite-difference scheme, stability

1 Introduction

We consider the two-dimensional parabolic equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f(x,y,t), \quad 0 < x < 1, \quad 0 < y < 1, \quad 0 < t \leqslant T, \tag{1}$$

subject to nonlocal integral conditions

$$u(0,y,t) = \gamma_1 \int_0^1 u(x,y,t)\mathrm{d}x + \mu_1(y,t), \tag{2}$$

$$u(1,y,t) = \gamma_2 \int_0^1 u(x,y,t)\mathrm{d}x + \mu_2(y,t), \quad 0 < y < 1, \quad 0 < t \leqslant T, \tag{3}$$

boundary conditions

$$u(x,0,t) = \mu_3(x,t), \quad u(x,1,t) = \mu_4(x,t), \quad 0 < x < 1, \quad 0 < t \leqslant T, \tag{4}$$

and initial condition

$$u(x,y,0) = \varphi(x,y), \quad 0 \leqslant x \leqslant 1, \quad 0 \leqslant y \leqslant 1, \tag{5}$$

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 1–10, 2011.

where $f(x,y,t)$, $\mu_1(y,t)$, $\mu_2(y,t)$, $\mu_3(x,t)$, $\mu_4(x,t)$, $\varphi(x,y)$ are given functions, γ_1, γ_2 are given parameters, and function $u(x,y,t)$ is unknown. We assume that for all t, $0 < t \leqslant T$, nonlocal integral conditions (2), (3) and boundary conditions (4) are compatible, i.e., the following compatibility conditions are satisfied:

$$\gamma_1 \int_0^1 \mu_3(x,t)\mathrm{d}x + \mu_1(0,t) = \mu_3(0,t),$$
$$\gamma_1 \int_0^1 \mu_4(x,t)\mathrm{d}x + \mu_1(1,t) = \mu_4(0,t),$$
$$\gamma_2 \int_0^1 \mu_3(x,t)\mathrm{d}x + \mu_2(0,t) = \mu_3(1,t),$$
$$\gamma_2 \int_0^1 \mu_4(x,t)\mathrm{d}x + \mu_2(1,t) = \mu_4(1,t).$$

The present paper is devoted to the numerical solution of the two-dimensional differential problem (1)–(5). We construct the fully-explicit finite-difference scheme and analyse its stability. In order to efficiently apply the stability analysis technique which has been applied for other types of finite-difference schemes, we formulate the proposed numerical method as splitting finite-difference scheme.

The stability of finite-difference schemes for the corresponding one-dimensional parabolic problems with nonlocal integral conditions similar to conditions (2), (3) or with more general integral conditions is investigated by M. Sapagovas [12,13], Ž. Jesevičiūtė and M. Sapagovas [4] and other authors. The alternating direction implicit (ADI) and locally one-dimensional (LOD) methods for the two-dimensional differential problem (1)–(5) has been proposed and the stability of these methods has been analysed by S. Sajavičius [5,6,7]. The LOD technique for two-dimensional parabolic problems with nonlocal integral condition (the specification of mass/energy) has been investigated by M. Dehghan [3]. The paper [14] deals with the ADI method for the two-dimensional parabolic equation (1) with Bitsadze-Samarskii type nonlocal boundary condition. We use the similar technique and argument in order to construct the fully-explicit finite-difference scheme for the differential problem (1)–(5) and to investigate the stability of that method.

The paper is organized as follows. In Sect. 2, the details of the fully-explicit finite-difference scheme are described. Section 3 reviews the well-known stability analysis technique which is based on the spectral structure of the transition matrix of a finite-difference scheme, and discusses possibilities to use such technique in order to analyse the stability of the proposed finite-difference scheme. The results of numerical experiment with a particular test problem are presented in Sect. 4. Some remarks in Sect. 5 conclude the paper.

2 The Fully-Explicit Finite-Difference Scheme

To solve the two-dimensional differential problem (1)–(5) numerically, we apply the finite-difference technique [11]. Let us define discrete grids with uniform steps,

$$\omega_{h_1} = \{x_i = ih_1, i = 1, 2, \ldots, N_1 - 1, N_1 h_1 = 1\},$$
$$\overline{\omega}_{h_1} = \omega_{h_1} \cup \{x_0 = 0, x_{N_1} = 1\},$$
$$\omega_{h_2} = \{y_j = jh_2, j = 1, 2, \ldots, N_2 - 1, N_2 h_2 = 1\},$$
$$\overline{\omega}_{h_2} = \omega_{h_2} \cup \{y_0 = 0, y_{N_2} = 1\},$$
$$\omega = \omega_{h_1} \times \omega_{h_2}, \quad \overline{\omega} = \overline{\omega}_{h_1} \times \overline{\omega}_{h_2},$$
$$\omega^\tau = \{t^k = k\tau, k = 1, 2, \ldots, M, M\tau = T\}, \quad \overline{\omega}^\tau = \omega^\tau \cup \{t^0 = 0\}.$$

We use the notation $U_{ij}^k = U(x_i, y_j, t^k)$ for functions defined on the grid $\overline{\omega} \times \overline{\omega}^\tau$ or its parts, and the notation $U_{ij}^{k+1/2} = U(x_i, y_j, t^k + 0.5\tau)$ (some of the indices can be omitted). We define one-dimensional discrete operators

$$\Lambda_1 U_{ij} = \frac{U_{i-1,j} - 2U_{ij} + U_{i+1,j}}{h_1^2}, \quad \Lambda_2 U_{ij} = \frac{U_{i,j-1} - 2U_{ij} + U_{i,j+1}}{h_2^2}.$$

Now we explain the main steps of the fully-explicit (splitting) finite-difference scheme for the numerical solution of problem (1)–(5). First of all, we replace the initial condition (5) by equations

$$U_{ij}^0 = \varphi_{ij}, \quad (x_i, y_j) \in \overline{\omega}. \tag{6}$$

Then, for any k, $0 \leqslant k < M - 1$, the transition from the kth layer of time to the $(k + 1)$th layer can be carried out by splitting it into two stages and solving one-dimensional finite-difference subproblems in each of them. The both of subproblems are fully-explicit. The values $U_{ij}^{k+1/2}$, $x_i \in \omega_{h_1}$, can be computed from the first subproblem:

$$\frac{U_{ij}^{k+1/2} - U_{ij}^k}{\tau} = \Lambda_1 U_{ij}^k, \quad y_j \in \omega_{h_2}, \tag{7}$$
$$U_{i0}^{k+1/2} = (\overline{\mu}_3)_i, \tag{8}$$
$$U_{iN_2}^{k+1/2} = (\overline{\mu}_4)_i, \tag{9}$$

where

$$\overline{\mu}_3 = (\mu_3)^k + \tau\Lambda_1(\mu_3)^k, \quad \overline{\mu}_4 = (\mu_4)^k + \tau\Lambda_1(\mu_4)^k.$$

In the second subproblem, nonlocal integral conditions (2), (3) are approximated by the trapezoidal rule and from this subproblem we can compute the values U_{ij}^{k+1}, $y_j \in \omega_{h_2}$:

$$\frac{U_{ij}^{k+1} - U_{ij}^{k+1/2}}{\tau} = \Lambda_2 U_{ij}^{k+1/2} + f_{ij}^{k+1}, \quad x_i \in \omega_{h_1}, \tag{10}$$
$$U_{0j}^{k+1} = \gamma_1(1, U)_j^{k+1} + (\mu_1)_j^{k+1}, \tag{11}$$
$$U_{N_1 j}^{k+1} = \gamma_2(1, U)_j^{k+1} + (\mu_2)_j^{k+1}, \tag{12}$$

where

$$(1,U)_j^{k+1} = h_1\Big(\frac{U_{0j}^{k+1}+U_{N_1j}^{k+1}}{2} + \sum_{i=1}^{N_1-1} U_{ij}^{k+1}\Big).$$

Every transition is finished by computing

$$U_{i0}^{k+1} = (\mu_3)_i^{k+1}, \quad U_{iN_2}^{k+1} = (\mu_4)_i^{k+1}, \quad x_i \in \overline{\omega}_{h_1}. \tag{13}$$

Thus, the procedure of numerical solution can be stated as follows:

procedure The Fully-Explicit Finite-Difference Scheme
begin
  Compute U_{ij}^0, $(x_i, y_j) \in \overline{\omega}$, from (6);
  for $k = 0, 1, \ldots, M-1$
    for each $x_i \in \omega_{h_1}$
      Compute $U_{ij}^{k+1/2}$, $y_j \in \overline{\omega}_{h_2}$, from (7)–(9);
    end for
    for each $y_j \in \omega_{h_2}$
      Compute U_{ij}^{k+1}, $x_i \in \overline{\omega}_{h_1}$, from (10)–(12);
    end for
    Compute U_{i0}^{k+1} and $U_{iN_2}^{k+1}$, $x_i \in \overline{\omega}_{h_1}$, from (13);
  end for
end

It is known [11] that the finite-difference scheme (7)–(12) approximates the two-dimensional differential problem (1)–(5) with error $\mathrm{O}(\tau + h_1 + h_2)$.

Now let us transform the finite-difference scheme (7)–(12) to the matrix form. From (11) and (12) we obtain

$$U_{0j}^{k+1} = \overline{\alpha}\sum_{i=1}^{N_1-1} U_{ij}^{k+1} + (\overline{\mu}_1)_j^{k+1},$$
$$U_{N_1j}^{k+1} = \overline{\beta}\sum_{i=1}^{N_1-1} U_{ij}^{k+1} + (\overline{\mu}_2)_j^{k+1},$$

where

$$\overline{\alpha} = \frac{\gamma_1 h_1}{D}, \quad \overline{\beta} = \frac{\gamma_2 h_1}{D},$$
$$(\overline{\mu}_1)_j^{k+1} = \Big(\frac{1}{D} - \frac{\overline{\beta}}{2}\Big)(\mu_1)_j^{k+1} + \frac{\overline{\alpha}}{2}(\mu_2)_j^{k+1},$$
$$(\overline{\mu}_2)_j^{k+1} = \frac{\overline{\beta}}{2}(\mu_1)_j^{k+1} + \Big(\frac{1}{D} - \frac{\overline{\alpha}}{2}\Big)(\mu_2)_j^{k+1},$$
$$D = 1 - \frac{h_1}{2}(\gamma_1+\gamma_2).$$

If $M_1 = \max\{|\gamma_1|, |\gamma_2|\} < \infty$ and the grid step $h_1 < 1/M_1$, then $D > 0$.

Let us introduce $(N_1-1)\times(N_1-1)$ and $(N_2-1)\times(N_2-1)$ matrices

$$\widetilde{\Lambda}_1 = h_1^{-2}\begin{pmatrix} -2+\overline{\alpha} & 1+\overline{\alpha} & \overline{\alpha} & \cdots & \overline{\alpha} & \overline{\alpha} & \overline{\alpha} \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ \overline{\beta} & \overline{\beta} & \overline{\beta} & \cdots & \overline{\beta} & 1+\overline{\beta} & -2+\overline{\beta} \end{pmatrix}$$

and

$$\widetilde{\Lambda}_2 = h_2^{-2}\begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix}.$$

Now we can define matrices of order $(N_1-1)\cdot(N_2-1)$,

$$A_1 = -E_{N_2-1}\otimes\widetilde{\Lambda}_1, \quad A_2 = -\widetilde{\Lambda}_2\otimes E_{N_1-1},$$

where E_N is the identity matrix of order N and $A\otimes B$ denotes the Kronecker (tensor) product of matrices A and B. We can directly verify that A_1 and A_2 are commutative matrices, i.e.,

$$A_1A_2 = A_2A_1 = \widetilde{\Lambda}_2\otimes\widetilde{\Lambda}_1.$$

Introducing the matrices A_1 and A_2 allow us to rewrite the finite-difference scheme (7)–(12) in the following form:

$$U^{k+1/2} = \big(E-\tau A_1\big)U^k, \tag{14}$$

$$U^{k+1} = \big(E-\tau A_2\big)U^{k+1/2}+\tau F^{k+1}, \tag{15}$$

where E is the identity matrix of order $(N_1-1)\cdot(N_2-1)$,

$$U = \big(\widetilde{U}_1,\widetilde{U}_2,\ldots,\widetilde{U}_j,\ldots,\widetilde{U}_{N_2-1}\big)^T, \quad \widetilde{U}_j = \big(U_{1j},U_{2j},\ldots,U_{ij},\ldots,U_{N_1-1,j}\big)^T,$$

and

$$F^{k+1} = \big(F_1^{k+1},F_2^{k+1},\ldots,F_j^{k+1},\ldots,F_{N_2-1}^{k+1}\big)^T,$$

$$F_j^{k+1} = \left(\frac{(\overline{\mu}_1)_j^{k+1}}{h_1^2}+f_{1j}^{k+1},f_{2j}^{k+1},\ldots,f_{N_1-2,j}^{k+1},\frac{(\overline{\mu}_2)_j^{k+1}}{h_1^2}+f_{N_1-1,j}^{k+1}\right)^T.$$

From (14) and (15) it follows that

$$U^{k+1} = SU^k + \overline{F}^k, \tag{16}$$

where

$$S = \big(E - \tau A_2\big)\big(E - \tau A_1\big), \quad \overline{F}^k = \tau F^{k+1}.$$

3 Analysis of the Stability

Let us recall some facts related with the stability of the finite-difference schemes. The finite-difference scheme (16) is called stepwise stable [2] if for all fixed τ and h_1, h_2 there exists a constant $C = C(\tau, h_1, h_2)$ such that $|U_{ij}^k| \leqslant C$, $(x_i, y_j) \in \overline{\omega}$, $k = 0, 1, \ldots$. We know [11] that a sufficient stability condition for the finite-difference scheme (16) can be written in the form

$$\|S\| \leqslant 1 + c_0\tau,$$

where a non-negative constant c_0 is independent on τ and h_1, h_2. The necessary and sufficient condition to define a matrix norm $\|\cdot\|_*$ such that $\|S\|_* < 1$ is the inequality [1]

$$\rho(S) = \max_{\lambda(S)} |\lambda(S)| < 1,$$

where $\lambda(S)$ is the eigenvalues of S and $\rho(S)$ is the spectral radius of S. If S is a simple-structured matrix, i.e., the number of linearly independent eigenvectors is equal to the order of the matrix, then it is possible to define the norm [12]

$$\|S\|_* = \|P^{-1}SP\|_\infty = \rho(S),$$

which is compatible with the vector norm

$$\|U\|_* = \|P^{-1}U\|_\infty,$$

where columns of the matrix P are linearly independent eigenvectors of S, and the norms

$$\|B\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{m} |b_{ij}|, \quad \|V\|_\infty = \max_{1\leqslant i\leqslant m} |v_i|,$$

m is the order of matrix $B = (b_{ij})_{i,j=1}^m$ and vector $V = (v_1, v_2, \ldots, v_m)^T$. Therefore, we will use the stability condition $\rho(S) < 1$ in the analysis of the stability of the finite-difference scheme (16).

The spectral structure of finite-difference and differential operators with nonlocal conditions are investigated by many authors (see, e.g., [8,9,10] and references therein). The eigenvalue problem for the matrix $(-\widetilde{\Lambda}_1)$ has been investigated in the paper [13]. When $\gamma_1 + \gamma_2 \leqslant 2$, then all the eigenvalues of the matrix $(-\widetilde{\Lambda}_1)$ are non-negative and algebraically simple real numbers: $\lambda_i(-\widetilde{\Lambda}_1) \geqslant 0$, $i = 1, 2, \ldots, N_1 - 1$. If $\gamma_1 + \gamma_2 > 2$, then there exists one and only one negative

eigenvalue of the matrix $(-\widetilde{\Lambda}_1)$. It is well-known [11] that all the eigenvalues of the matrix $(-\widetilde{\Lambda}_2)$ are real, positive and algebraically simple:

$$\lambda_j(-\widetilde{\Lambda}_2) = \frac{4}{h_2^2}\sin^2\frac{j\pi h_2}{2}, \quad j = 1, 2, \ldots, N_2 - 1.$$

Let us denote

$$\Delta_2 = \max_{\lambda(A_2)} \lambda(A_2) = \max_{1\leqslant j\leqslant N_2-1} \lambda_j(-\widetilde{\Lambda}_2) = \lambda_{N_2-1}(-\widetilde{\Lambda}_2) = \frac{4}{h_2^2}\cos^2\frac{\pi h_2}{2}.$$

Since A_1 and A_2 are simple-structured matrices as Kornecker products of two simple-structured matrices, then S is a simple-structured matrix too, and the eigenvalues of the matrix S can be expressed as follows:

$$\lambda(S) = \big(1 - \tau\lambda(A_1)\big) \cdot \big(1 - \tau\lambda(A_2)\big). \tag{17}$$

If all the eigenvalues of the matrix A_1 are real and non-negative numbers, then the finite-difference scheme (16) is stable under condition

$$\tau < \tau^* = 2\min\left\{\frac{1}{\rho(A_1)}, \frac{1}{\Delta_2}\right\}. \tag{18}$$

Indeed, from (17) it follows that

$$|\lambda(S)| = |1 - \tau\lambda(A_1)| \cdot |1 - \tau\lambda(A_2)|.$$

Thus, we conclude that $\rho(S) < 1$, if condition (18) are fulfilled.

The eigenvalues of the matrix A_1 coincide with the eigenvalues of the matrix $(-\widetilde{\Lambda}_1)$ and they are multiple. Thus, the fully-explicit finite-difference scheme is stable if all the eigenvalues of the matrix $(-\widetilde{\Lambda}_1)$ are non-negative, i.e., if $\gamma_1 + \gamma_2 \leqslant 2$. The non-negativity of the eigenvalues of the matrix $(-\widetilde{\Lambda}_1)$ ensures the stability of the finite-difference scheme (16), but it is notable [14] that the scheme can be stable even if there exists a negative eigenvalue of the matrix $(-\widetilde{\Lambda}_1)$.

4 Numerical Results

In order to demonstrate the efficiency of the considered fully-explicit finite-difference scheme and practically justify the stability analysis technique, we solved a particular test problem. In our test problem, functions $f(x, y, t)$, $\mu_1(y, t)$, $\mu_2(y, t)$, $\mu_3(x, t)$, $\mu_4(x, t)$ and $\varphi(x, y)$ were chosen so that the function

$$u(x, y, t) = x^3 + y^3 + t^3$$

would be the solution to the differential problem (1)–(5), i.e.,

$$
\begin{gathered}
f(x,y,t) = -3(2x + 2y - t^2), \\
\mu_1(y,t) = y^3 + t^3 - \gamma_1(0.25 + y^3 + t^3), \\
\mu_2(y,t) = 1 + y^3 + t^3 - \gamma_2(0.25 + y^3 + t^3), \\
\mu_3(x,t) = x^3 + t^3, \\
\mu_4(x,t) = x^3 + 1 + t^3, \\
\varphi(x,y) = x^3 + y^3.
\end{gathered}
$$

The finite-difference scheme was implemented in a stand-alone C application. Numerical experiment with $h_1 = h_2 = 10^{-2}$, $T = 2.0$, $\gamma_1 = -\gamma_2 = 1$ and with different values of τ was performed using the technologies of grid computing. To estimate the accuracy of the numerical solution, we calculated the maximum norm of computational error,

$$
\|\varepsilon\|_{C_h} = \max_{0 \leqslant k \leqslant M} \max_{\substack{0 \leqslant i \leqslant N_1 \\ 0 \leqslant j \leqslant N_2}} |U_{ij}^k - u(x_i, y_j, t^k)|.
$$

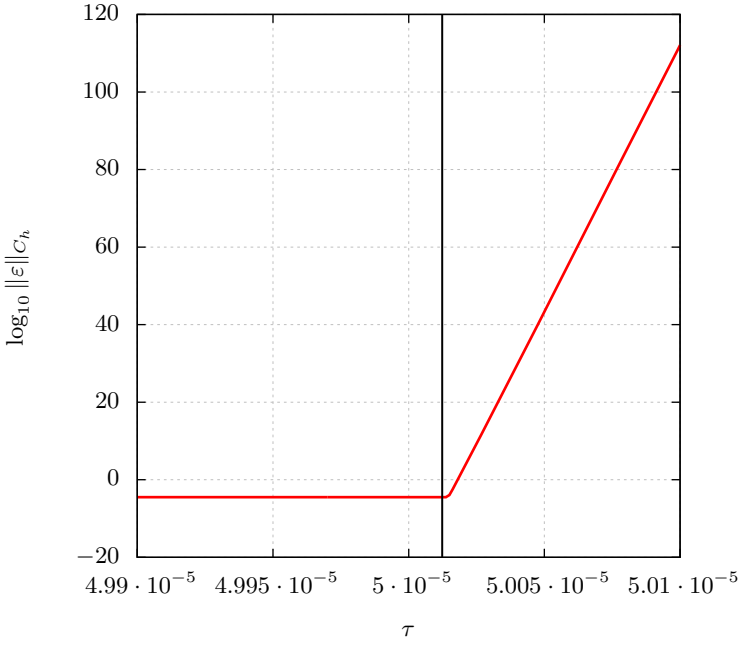


Fig. 1. The dependence of $\log_{10} \|\varepsilon\|_{C_h}$ for various values of τ ($\gamma_1 = -\gamma_2 = 1$). The vertical straight line denotes the line $\tau = \tau^*$.

Note that

$$\min_{0\leqslant t\leqslant T}\min_{\substack{0\leqslant x\leqslant 1\\ 0\leqslant y\leqslant 1}} u(x,y,t) = u(0,0,0) = 0,$$

$$\max_{0\leqslant t\leqslant T}\max_{\substack{0\leqslant x\leqslant 1\\ 0\leqslant y\leqslant 1}} u(x,y,t) = u(1,1,T) = 10.$$

For the numerical analysis of the spectrum of the matrix $(-\widetilde{\Lambda}_1)$, MATLAB (The MathWorks, Inc.) software package [15] was used. The results of the numerical analysis of the spectrum of the matrix $(-\widetilde{\Lambda}_1)$ shown that $\rho(A_1) = 3.999013170072345 \cdot 10^4$. Since $\Delta_2 = 3.999013120731463 \cdot 10^4 < \rho(A_1)$, then $\tau^* = 2/\rho(A_1) = 5.001233841807574 \cdot 10^{-5}$.

The influence of condition (18) for the stability of the finite-difference scheme (16) was investigated. From Fig. 1 we see that the norm $\|\varepsilon\|_{C_h}$ is small enough when $\tau < \tau^*$.

5 Concluding Remarks

We developed the fully-explicit finite-difference scheme for the two-dimensional parabolic equation with two nonlocal integral conditions. Applying quite a simple technique allow us to investigate the stability of this method. The stability analysis technique is based on the analysis of the spectrum of the transition matrix of a finite-difference scheme. We demonstrate that the proposed scheme can be stable or unstable depending on the parameters of nonlocal conditions. The results of numerical experiment with a particular test problem justify theoretical results.

The fully-explicit finite-difference scheme can be generalized for the corresponding two-dimensional differential problem with nonlocal integral conditions

$$u(0,y,t) = \gamma_1 \int_0^1 \alpha(x)u(x,y,t)\mathrm{d}x + \mu_1(y,t),$$

$$u(1,y,t) = \gamma_2 \int_0^1 \beta(x)u(x,y,t)\mathrm{d}x + \mu_2(y,t), \quad 0 < y < 1, \quad 0 < t \leqslant T,$$

where $\alpha(x)$ and $\beta(x)$ are given functions.

References

1. Atkinson, K.E.: An Introduction to Numerical Analysis. John Willey & Sons, New York (1978)
2. Cahlon, B., Kulkarni, D.M., Shi, P.: Stepwise stability for the heat equation with a nonlocal constraint. SIAM J. Numer. Anal. 32(2), 571–593 (1995)
3. Dehghan, M.: Implicit locally one-dimensional methods for two-dimensional diffusion with a non-local boundary condition. Math. Comput. Simulat. 49(4–5), 331–349 (1999)

4. Jesevičiūtė, Ž., Sapagovas, M.: On the stability of finite-difference schemes for parabolic equations subject to integral conditions with applications to thermoelasticity. Comput. Methods Appl. Math. 8(4), 360–373 (2008)
5. Sajavičius, S.: On the stability of alternating direction method for two-dimensional parabolic equation with nonlocal integral conditions. In: Kleiza, V., Rutkauskas, S., Štikonas, A. (eds.) Proceedings of International Conference Differential Equations and their Applications (DETA 2009), pp. 87–90. Panevėžys, Lithuania (2009)
6. Sajavičius, S.: The stability of finite-difference scheme for two-dimensional parabolic equation with nonlocal integral conditions. In: Damkilde, L., Andersen, L., Kristensen, A.S., Lund, E. (eds.) Proceedings of the Twenty Second Nordic Seminar on Computational Mechanics / DCE Technical Memorandum, vol. 11, pp. 87–90. Aalborg, Denmark (2009)
7. Sajavičius, S.: On the stability of locally one-dimensional method for two-dimensional parabolic equation with nonlocal integral conditions. In: Pereira, J.C.F., Sequeira, A., Pereira, J.M.C. (eds.) Proceedings of the V European Conference on Computational Fluid Dynamics (ECCOMAS CFD 2010). CD-ROM, Paper ID 01668, 11 p. Lisbon, Portugal (2010)
8. Sajavičius, S.: On the eigenvalue problems for differential operators with coupled boundary conditions. Nonlinear Anal., Model. Control 15(4), 493–500 (2010)
9. Sajavičius, S.: On the eigenvalue problems for finite-difference operators with coupled boundary conditions. Šiauliai Math. Semin. 5(13), 87–100 (2010)
10. Sajavičius, S., Sapagovas, M.: Numerical analysis of the eigenvalue problem for one-dimensional differential operator with nonlocal integral conditions. Nonlinear Anal., Model. Control 14(1), 115–122 (2009)
11. Samarskii, A.A.: The Theory of Difference Schemes. Marcel Dekker Inc., New York–Basel (2001)
12. Sapagovas, M.: On the stability of a finite-difference scheme for nonlocal parabolic boundary-value problems. Lith. Math. J. 48(3), 339–356 (2008)
13. Sapagovas, M.P.: On stability of finite-difference schemes for one-dimensional parabolic equations subject to integral condition. Obchysl. Prykl. Mat. 92, 77–90 (2005)
14. Sapagovas, M., Kairytė, G., Štikonienė, O., Štikonas, A.: Alternating direction method for a two-dimensional parabolic equation with a nonlocal boundary condition. Math. Model. Anal. 12(1), 131–142 (2007)
15. Yang, W.Y., Cao, W., Chung, T.-S., Morris, J.: Applied Numerical Methods Using MATLAB®. John Willey & Sons, New York (2005)

Numerical Solution of Multi-scale Electromagnetic Boundary Value Problems by Utilizing Transformation-Based Metamaterials

Ozlem Ozgun[1] and Mustafa Kuzuoglu[2]

[1] Dept. of Electrical Engineering, Middle East Technical University, Northern Cyprus Campus, Guzelyurt, Mersin 10, Turkey
[2] Dept. of Electrical Engineering, Middle East Technical University, 06531, Ankara, Turkey
{ozozgun,kuzuoglu}@metu.edu.tr

Abstract. We present numerical solution techniques for efficiently handling multi-scale electromagnetic boundary value problems having fine geometrical details or features, by utilizing spatial coordinate transformations. The principle idea is to modify the computational domain of the finite methods (such as the finite element or finite difference methods) by suitably placing anisotropic metamaterial structures whose material parameters are obtained by coordinate transformations, and hence, to devise easier and efficient numerical simulation tools in computational electromagnetics by allowing uniform and easy-to-generate meshes or by decreasing the number of unknowns. Inside the modified computational domain, Maxwell's equations are still satisfied, but the medium where the coordinate transformation is applied turns into an anisotropic medium whose constitutive parameters are determined by the Jacobian of the coordinate transformation. In other words, by employing the form-invariance property of Maxwell's equations under coordinate transformations, an equivalent model that mimics the original problem is created to get rid of mesh refinement around the small-scale features. Various numerical applications of electromagnetic scattering problems are illustrated via finite element simulations.

Keywords: Anisotropic metamaterials; coordinate transformation; finite element method; multi-scale methods; electromagnetic scattering.

1 Introduction

There have been significant advances in techniques for solving problems of computational electromagnetics (CEM) in recent years, due to the need for efficient simulation tools that can handle cases involving geometrically-complex objects. Finite Element Method (FEM) has been widely used for solving such problems owing to its adaptability to arbitrary geometries and material inhomogeneities. However, FEM or other conventional CEM methods run into difficulties when dealing with multi-scale structures due to the need to capture the small scale effects on the large

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 11–25, 2011.

scale with adequate numerical precision. For instance, if an electrically-small object whose size is only a small fraction of the wavelength inside the domain is located in an inhomogeneous medium along with other large scale objects, the mesh must be refined over the object, leading in turn to a large number of unknowns. Practical applications may include the modeling of a small antenna mounted on a large platform, or simulating thin interconnect structures in integrated circuits. Such problems also challenge the process of mesh generation because it is required to create a fine mesh that accurately captures the smaller-scale variation of the object, and often introduce ill-conditioned matrix systems because of poor mesh quality. This, in turn, makes it difficult to extract an accurate solution of such problems in a numerically efficient manner. There are some alternative approaches to alleviate such difficulties in the literature, some of which may be found in [1-6].

In this paper, we present numerical simulation tools in numerical modeling of electromagnetic boundary value problems via finite methods (such as the finite element or finite difference methods) by placing special transformation media inside their computational domains for the purpose of uniform and easy-to-generate meshes and less number of unknowns. The concept of "transformation media" refers to artificial media whose constitutive parameters are designed by employing the form invariance property of Maxwell's equations under coordinate transformations so as to mimic field behavior in the modified coordinate system. The coordinate transformation technique achieves the duality between geometry and material parameters in the sense that Maxwell's equations preserve their form in the modified coordinate system, but the medium transforms into an anisotropic medium. The design of the invisibility cloak as a transformation medium initiated the widespread use of the coordinate transformation technique as a powerful and intuitive design tool [7]. However, the range of applications of the coordinate transformation technique goes well beyond cloaking, and varies from the design of perfectly matched layers (PMLs) for the purpose of mesh truncation in finite methods [8-9], to the design of electromagnetic reshapers for both objects and waveguides, concentrators, rotators, lenses, etc [10-17]. Except for the PML, all applications so far concern with the material itself, i.e. its constitutive parameters, how it behaves, and how it can be realized physically. However, this paper is interested in utilizing such materials for the sake of efficient numerical modeling in computer-aided simulation tools, and the physical realization of these materials is beyond the scope of this paper.

The organization of this paper is as follows: In Sec. 2, the form-invariance property of Maxwell's equations under a general coordinate transformation is discussed. Each sub-section in Sec. 3 is devoted to a single application in computational electromagnetics, exhibiting multi-scale behavior in numerical modeling. The techniques for the numerical solutions of the scattering problems involving electrically-small objects; an object coated by electrically-thin dielectric layers; an object with electrically-thin overhang; and objects that are separated by electrically-short distances are described and demonstrated via the finite element method in Sec. 3.1, 3.2, 3.3 and 3.4, respectively. Finally, the conclusions are presented in Sec. 4.

2 Form-Invariance Property of Maxwell's Equations under General Coordinate Transformations

A coordinate transformation $\vec{r} \to \tilde{\vec{r}} = \mathrm{T}(\vec{r})$ transforms (maps or moves) each point P in the original space Ω to another point $\tilde{\mathrm{P}}$ in the transformed space, where $\vec{r} = (x, y, z)$ and $\tilde{\vec{r}} = (\tilde{x}, \tilde{y}, \tilde{z})$ are the position vectors of the points P and $\tilde{\mathrm{P}}$ in the original and transformed coordinate systems, respectively. The electromagnetic fields in a source-free medium with constitutive parameters (ε, μ) satisfy Maxwell's equations in transformed coordinates as follows:

$$\tilde{\nabla} \times \vec{E}\left(\tilde{\vec{r}}\right) = -j\omega\mu \vec{H}\left(\tilde{\vec{r}}\right) \tag{1.a}$$

$$\tilde{\nabla} \times \vec{H}\left(\tilde{\vec{r}}\right) = j\omega\varepsilon \vec{E}\left(\tilde{\vec{r}}\right) \tag{1.b}$$

where $\vec{E}\left(\tilde{\vec{r}}\right)$ and $\vec{H}\left(\tilde{\vec{r}}\right)$ are the transformed fields (i.e., mapped versions of $\vec{E}(\vec{r})$ and $\vec{H}(\vec{r})$ to transformed space, respectively), and $\tilde{\nabla} = [\bar{\bar{\mathrm{J}}}^{-1}]^{\mathrm{T}} \cdot \nabla$ is the nabla operator in transformed space. Here, $\bar{\bar{\mathrm{J}}} = \partial(\tilde{x}, \tilde{y}, \tilde{z}) / \partial(x, y, z)$ is the Jacobian tensor of the coordinate transformation.

Maxwell's equations in transformed coordinates are equivalent to Maxwell's equations in original coordinates as follows:

$$\nabla \times \tilde{\vec{E}}(\vec{r}) = -j\omega\bar{\bar{\mu}} \cdot \tilde{\vec{H}}(\vec{r}) \tag{2.a}$$

$$\nabla \times \tilde{\vec{H}}(\vec{r}) = j\omega\bar{\bar{\varepsilon}} \cdot \tilde{\vec{E}}(\vec{r}) \tag{2.b}$$

In other words, the effect of the coordinate transformation is reflected to the electromagnetic fields by replacing the original medium with an anisotropic medium so that transformed fields satisfy original forms of Maxwell's equations. The permittivity and permeability tensors $(\bar{\bar{\varepsilon}}, \bar{\bar{\mu}})$ of the new anisotropic medium can be determined from the Jacobian of the transformation as follows [18]:

$$\bar{\bar{\varepsilon}} = \varepsilon \bar{\bar{\Lambda}} \tag{3.a}$$

$$\bar{\bar{\mu}} = \mu \bar{\bar{\Lambda}} \tag{3.b}$$

where

$$\bar{\bar{\Lambda}} = (\det \bar{\bar{\mathrm{J}}})(\bar{\bar{\mathrm{J}}}^{\mathrm{T}} \cdot \bar{\bar{\mathrm{J}}})^{-1} \tag{4}$$

where *det* denotes the determinant.

If the original medium is an arbitrary anisotropic medium with parameters $(\bar{\bar{\varepsilon}}_a, \bar{\bar{\mu}}_a)$, the parameters of the new material in the transformed space are obtained by

$$\bar{\bar{\varepsilon}} = \frac{1}{\det\left(\bar{\bar{\mathrm{J}}}^{-1}\right)} \left(\bar{\bar{\mathrm{J}}}^{-1}\right)^{\mathrm{T}} \cdot \bar{\bar{\varepsilon}}_a \cdot \left(\bar{\bar{\mathrm{J}}}^{-1}\right) \tag{5.a}$$

$$\bar{\bar{\mu}} = \frac{1}{\det\left(\bar{\bar{\mathrm{J}}}^{-1}\right)}\left(\bar{\bar{\mathrm{J}}}^{-1}\right)^{\mathrm{T}} \cdot \bar{\bar{\mu}}_a \cdot \left(\bar{\bar{\mathrm{J}}}^{-1}\right) \tag{5.b}$$

Original fields in transformed coordinates (i.e., $\vec{E}\left(\tilde{\vec{r}}\right)$) and the transformed fields in original coordinates (i.e., $\tilde{\vec{E}}(\vec{r})$) are interdependent, as follows:

$$\vec{E}(\vec{r}) \to \tilde{\vec{E}}(\vec{r}) = \bar{\bar{\mathrm{J}}}^{\mathrm{T}} \cdot \vec{E}\left(\tilde{\vec{r}}\right) \tag{6.a}$$

$$\vec{H}(\vec{r}) \to \tilde{\vec{H}}(\vec{r}) = \bar{\bar{\mathrm{J}}}^{\mathrm{T}} \cdot \vec{H}\left(\tilde{\vec{r}}\right) \tag{6.b}$$

Hence, the original fields inside the transformed space can be recovered by using the fields inside the anisotropic material through the use of (6).

By manipulating Maxwell's equations, the electric field satisfies the following vector wave equation in transformed and original coordinates, respectively, as follows:

$$\tilde{\nabla} \times \tilde{\nabla} \times \vec{E}\left(\tilde{\vec{r}}\right) - k^2 \vec{E}\left(\tilde{\vec{r}}\right) = 0 \tag{7.a}$$

$$\nabla \times \left\{ \bar{\bar{\Lambda}}^{-1} \cdot \nabla \times \tilde{\tilde{E}}(\vec{r}) \right\} - k^2 \bar{\bar{\Lambda}} \cdot \tilde{\tilde{E}}(\vec{r}) = 0 \tag{7.b}$$

where $k = \omega\sqrt{\mu\varepsilon}$ is the wavenumber of the original medium. In two-dimensions (2D) where $\vec{E}(\vec{r}) = \hat{a}_z E_z(x,y)$, the vector wave equations reduce to scalar Helmholtz equations in transformed and original coordinates, respectively, as follows:

$$\tilde{\nabla}^2 E_z + k^2 E_z = 0 \tag{8.a}$$

$$\nabla \cdot \left(\bar{\bar{\Lambda}}_{\text{2-by-2}} \nabla \tilde{E}_z \right) + k^2 \Lambda_{33} \tilde{E}_z = 0 \tag{8.b}$$

where

$$\bar{\bar{\Lambda}} = \left[\begin{array}{cc|c} & & 0 \\ \multicolumn{2}{c|}{\bar{\bar{\Lambda}}_{\text{2-by-2}}} & 0 \\ \hline 0 & 0 & \Lambda_{33} \end{array}\right] \tag{9}$$

because of z-independence.

It is worthwhile noting that the numerical solution of wave equations can be performed either in original coordinates by replacing the original medium with the anisotropic material whose parameters are determined by (3), or in transformed coordinates by interchanging the original coordinates with the transformed counterparts inside the original medium. This conclusion is interesting in a way such that the numerical simulations can be realized by just interchanging the coordinates without computing the material parameters.

It must also be noted that the coordinate transformation must be continuous. That is, for any given $\xi > 0$, there exists a $\delta > 0$ (δ depends on ε), such that $\left\| \mathrm{T}(\vec{r}) - \mathrm{T}\left(\vec{r}^{\,*}\right) \right\| < \xi$, whenever $\left\| \vec{r} - \vec{r}^{\,*} \right\| < \delta$. In other words, two closely-located points $\vec{r}$ and $\vec{r}^{\,*}$ should be mapped to closely-located points $\tilde{\vec{r}}$ and $\tilde{\vec{r}}^{\,*}$, respectively, inside the

transformed region to avoid considerable deviations in the entries of the permittivity and permeability tensors corresponding to contiguous points. Otherwise, finite methods yield undesirable results because of ill-conditioned matrices.

Finally, it is a remarkable fact that various problems that are related to steady-state oscillations (mechanical, acoustical, or electromagnetic) lead to the 2D Helmholtz equation. Hence, the scope of the Helmholtz equation is broad due to its relationship to the wave equation in sinusoidal steady-state. Although the techniques in this paper are presented in conjuction with electromagnetics, they can be applied to various areas with suitable modifications.

3 Design of Transformation Media in Multi-scale Applications

In this section, the coordinate transformation design procedure of anisotropic metamaterials is introduced to achieve the numerical simulation of multi-scale electromagnetic scattering problems without resorting to mesh refinement around the objects. Each application in each sub-section is illustrated by numerical simulations performed by the finite element method.

3.1 Scattering from Electrically-Small Objects

Numerical solution of the scattering problem involving an electrically-small object (whose size is small compared to wavelength λ) is still a challenging task, because a large number of unknowns must be introduced inside the computational domain that is not occupied by the object (see Fig. 1.a). The reason of this requirement is twofold: (i) The mesh must be refined around the object to accurately estimate field variation with adequate numerical precision; (ii) Mesh truncation boundaries (such as PML or absorbing boundary condition) must be located sufficiently far away from the object to reduce spurious reflections.

In the proposed technique, an equivalent problem is designed by locating anisotropic layer(s) at an arbitrary distance from the object(s) (see Fig. 1.b). The equivalent problem transforms the original problem into a relatively high-frequency problem, and employs a more convenient and uniform mesh generation scheme with less number of unknowns inside the computational domain. The "equivalence" here means that both problems yield identical field values in their common free-space regions, and the fields inside the anisotropic layer are related to the original near-fields at the close vicinity of the object through the field equivalence in (6). An interesting feature of the equivalent problem is that the same mesh can be used for any arbitrarily-shaped object by simply changing the constitutive parameters of the anisotropic layer with respect to the geometry of the object.

In designing the metamaterial layer, each point P inside Ω_{M} is mapped to $\tilde{\mathrm{P}}$ inside the transformed region $\tilde{\Omega} = \Omega \cup \Omega_{\mathrm{M}}$, by using the following coordinate transformation $\mathrm{T}:\Omega_{\mathrm{M}} \to \tilde{\Omega}$

$$\tilde{\vec{r}} = \frac{\left\| \vec{r}_{\mathrm{a}} - \vec{r}_{\mathrm{c}} \right\|}{\left\| \vec{r}_{\mathrm{a}} - \vec{r}_{\mathrm{b}} \right\|} \left(\vec{r} - \vec{r}_{\mathrm{b}} \right) + \vec{r}_{\mathrm{c}} \tag{10}$$

where $\|\cdot\|$ denotes the Euclidean norm. In addition, $\vec{r}_a$, $\vec{r}_b$ and $\vec{r}_c$ are the position vectors of the points P_a, P_b and P_c, through the unit vector $\hat{a}$ originating from a point inside the innermost domain (such as the center-of-mass point) in the direction of the point P inside the metamaterial layer.

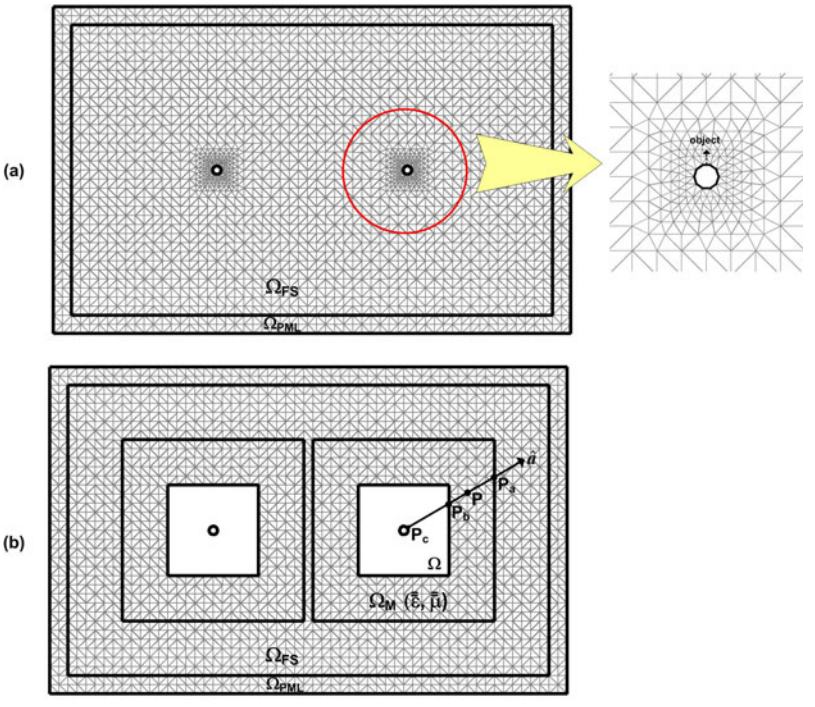


Fig. 1. Scattering from electrically-small circular objects: (a) Original problem with refined mesh, (b) Equivalent problem with metamaterial layer and uniform mesh. [Ω_M : metamaterial layer, Ω_{FS} : free-space, Ω_{PML} : perfectly matched layer]

It is useful to note that this approach transforms the boundary of the small object to the inner boundary of the metamaterial layer. Therefore, the boundary conditions that must be imposed on the boundary of the conducting object (i.e., tangential component of the total field must be zero on the boundary) must be imposed on the inner boundary of the metamaterial layer. This approach can be extended to dielectric objects in a straightforward manner. This can be achieved by making $\vec{r}_c = 0$, and by computing the parameters of the metamaterial layer with respect the dielectric constant of the dielectric object when the transformed point falls into the object.

The technique is simulated by the finite element method, and illustrated in Fig. 2. In this example, a plane wave whose angle of incidence is 90° with respect to the *x*-axis is incident to two circular conducting objects, each of which has a radius of λ/20 (λ=1m). The element size in the equivalent problem is approximately λ/40, whereas

the element size in the original problem is gradually decreased from the boundary of the outermost boundary to the boundary of the object. The magnitude of the total electric field, as well as the radar cross section (RCS) profile, for both original and equivalent problems are plotted in Fig. 2. To measure the performance of the proposed method, a mean-square error $\mathrm{Err} = \sum_{\Omega_{FS}} \left| E^{\text{equivalent}} - E^{\text{original}} \right|^2 \Big/ \sum_{\Omega_{FS}} \left| E^{\text{original}} \right|^2$ is defined, where $E^{\text{equivalent}}$ and E^{original} are the electric fields calculated in the equivalent and original problems, respectively, inside the common free-space region. The error is computed as 0.074% in this simulation. Moreover, the reduction in unknowns in the equivalent problem is found as 35%, compared to the original problem. The computation times on Intel Core 2 Duo 1.6 GHz with 2 GB RAM are tabulated in Table 1, together with the computation times of the other examples below.

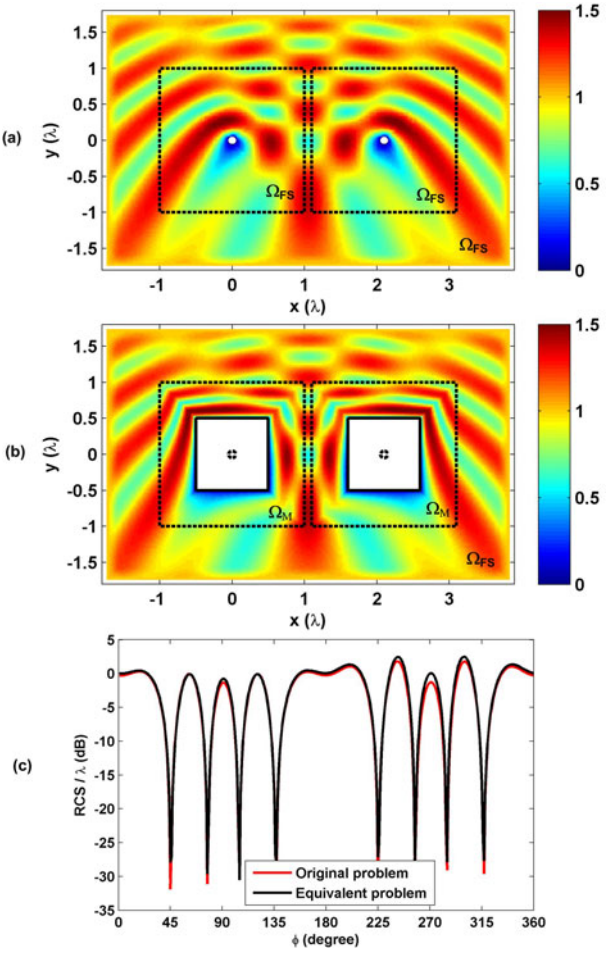


Fig. 2. Finite element simulation of scattering from electrically-small circular objects: (a) Magnitude of total electric field contour in original problem, (b) Magnitude of total electric field contour in equivalent problem, (c) RCS profiles

3.2 Scattering from Objects Coated by Electrically-Thin Dielectric Layers

This section examines the scattering problem involving an object coated by electrically-thin dielectric layers. Accurate computation of fields inside the dielectric layers requires fine mesh, especially if the thickness is small, because the wavelength decreases inside the dielectrics and smaller-sized elements must be chosen for good numerical precision. In the proposed approach, an equivalent problem is designed by locating an anisotropic layer at an arbitrary distance from the object (see Fig. 3).

Inside the metamaterial layer, each point P inside Ω_M is mapped to $\tilde{P}$ inside the transformed region $\tilde{\Omega} = \Omega \cup \Omega_M \cup \Omega_d$, by using the same coordinate transformation in (10). Although the same expression is used in both applications, they differ from each other in terms of the way of implementing the approach. In this problem, when the transformed point $\tilde{P}$ lies inside the dielectric region, the parameters of the metamaterial layer are obtained by using the dielectric constant of the dielectric layer. The salutary feature of this technique is that different problems (such as multiple layers with different thicknesses and dielectric constants) can be simulated by employing a single mesh, and by changing only the material parameters accordingly.

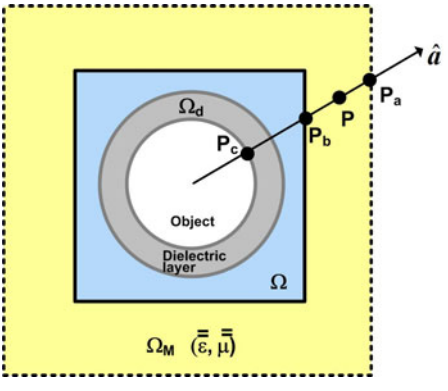


Fig. 3. Metamaterial layer design for scattering from an object coated by a dielectric layer

The finite element simulations are presented in Figs. 4 and 5, where a plane wave whose angle of incidence is 90° illuminates a circular conducting object with diameter of 1λ. The same mesh structure is used in the equivalent problems of both examples. In Fig. 4, the object is coated by a dielectric layer whose dielectric constant is 8 and thickness is λ/10. In the equivalent problem, the element size is approximately λ/30, whereas, in the original problem, the element size inside the dielectric layer is λ/100. In Fig. 5, the object is coated by two dielectric layers. The dielectric constant and thickness of the inner layer is 8 and λ/20, respectively. The dielectric constant and thickness of the outer layer is 4 and λ/20, respectively. In the original problem, the element size inside the inner and outer layers is λ/160 and λ/80, respectively. The mean-square error is computed as 1.068% and 0.205% in Fig. 4 and 5, respectively.

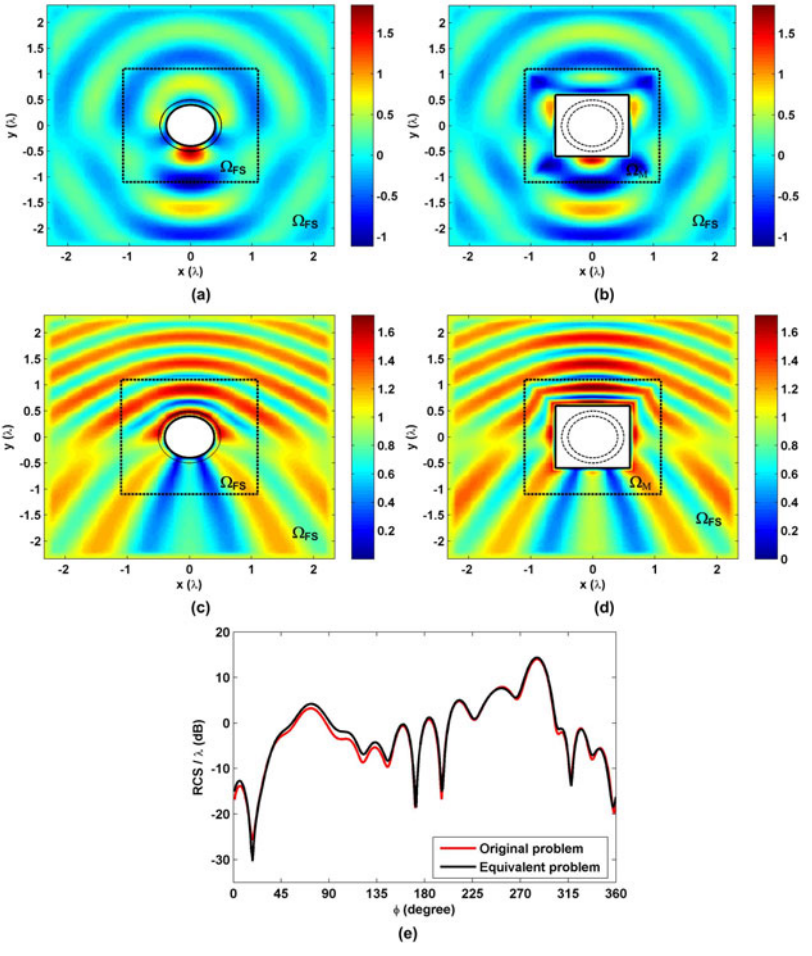


Fig. 4. Finite element simulation of scattering from a circular object coated by a single dielectric layer: (a) Real part of scattered electric field in original problem, (b) Real part of scattered electric field in equivalent problem, (c) Magnitude of total electric field in original problem, (d) Magnitude of total electric field in equivalent problem, (e) RCS profiles.

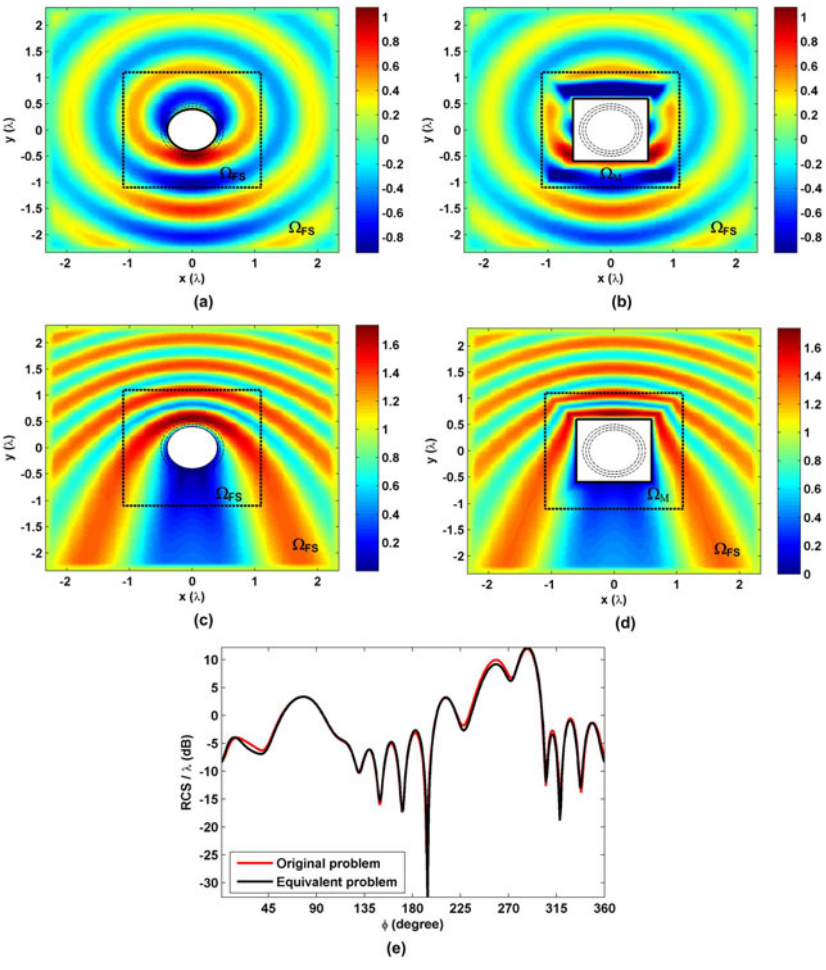


Fig. 5. Finite element simulation of scattering from a circular object coated by double dielectric layers: (a) Real part of scattered electric field in original problem, (b) Real part of scattered electric field in equivalent problem, (c) Magnitude of total electric field in original problem, (d) Magnitude of total electric field in equivalent problem, (e) RCS profiles.

3.3 Scattering from Objects Having Electrically-Thin Features

This section considers the scattering from an object having electrically-thin features or overhang. As shown in Fig. 6.a, if an object has a thin feature compared to other parts of the object, the mesh must be refined around the thin structure in conventional methods. In the proposed technique, an equivalent problem is constructed as shown in this figure. Each point P inside Ω_M is mapped to $\tilde{P}$ inside the transformed region $\tilde{\Omega} = \Omega \cup \Omega_M$, by using the same coordinate transformation in (10). Note that Ω

includes the thin region. By this transformation, the entire rectangular region is compressed to the yellow metamaterial region. In this technique, the unit vector is chosen to be vertical for all points as shown in Fig. 6.a. When the transformed point $\tilde{P}$ lies inside the thin region, the material properties of this region are used. For instance, if the thin structure is conducting, the boundary condition that the tangential component of the total electric field is zero is imposed on the boundary. In this method as well, different geometries (in fact, any geometry with any parameters) inside Ω can be simulated by employing a single mesh, and by changing only the material parameters according to each geometry.

The finite element simulations are presented in Figs. 7 and 8, where a plane wave whose angle of incidence is 45° illuminates an object with thin overhang whose thickness is λ/15. The position of the thin structure is at $x = 0$ and $x = 0.25\lambda$ in Fig. 7 and 8, respectively. The same mesh structure is used in the equivalent problems of both simulations. In the equivalent problem, the element size is approximately λ/30, but non-uniform mesh is generated in original problems. The mean-square error is computed as 0.192% and 0.143% in Fig. 7 and 8, respectively.

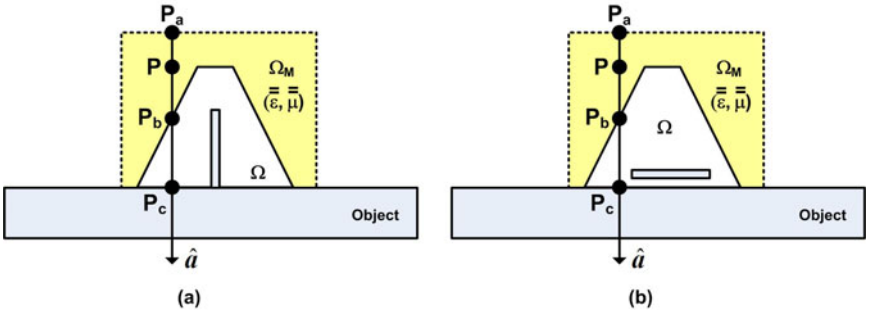


Fig. 6. Design of metamaterial layers: (a) Object with thin overhang, (b) Objects separated by short distances (distance adjuster)

3.4 Scattering from Objects Separated by Electrically-Short Distances

The approach in Sec. 3.3 can be extended to simulate objects that are placed very close to another object (see Fig. 6.b). In order to avoid the mesh refinement between the objects, a metamaterial region can be designed as described in Sec. 3.3. In this technique, it is possible to use the metamaterial layer as a kind of *distance adjuster*. Different distances can be simulated without creating a mesh anew.

The finite element simulations are presented in Fig. 9, where a plane wave whose angle of incidence is -45° is incident to objects that are separated by λ/15. The mean-square error is computed as 1.178%.

Table 1. Computation times of the examples

	Original Problem			Equivalent Problem		
	Mesh generation phase (sec)	FEM solution phase (sec)	Total (sec)	Mesh generation phase (sec)	FEM solution phase (sec)	Total (sec)
Fig. 2	168	1,117	1,285	86	500	586
Fig. 4	71	288	359	55	228	283
Fig. 5	72	292	364	56	240	296
Fig. 7	70	313	383	50	148	198
Fig. 9	51	170	221	40	108	148

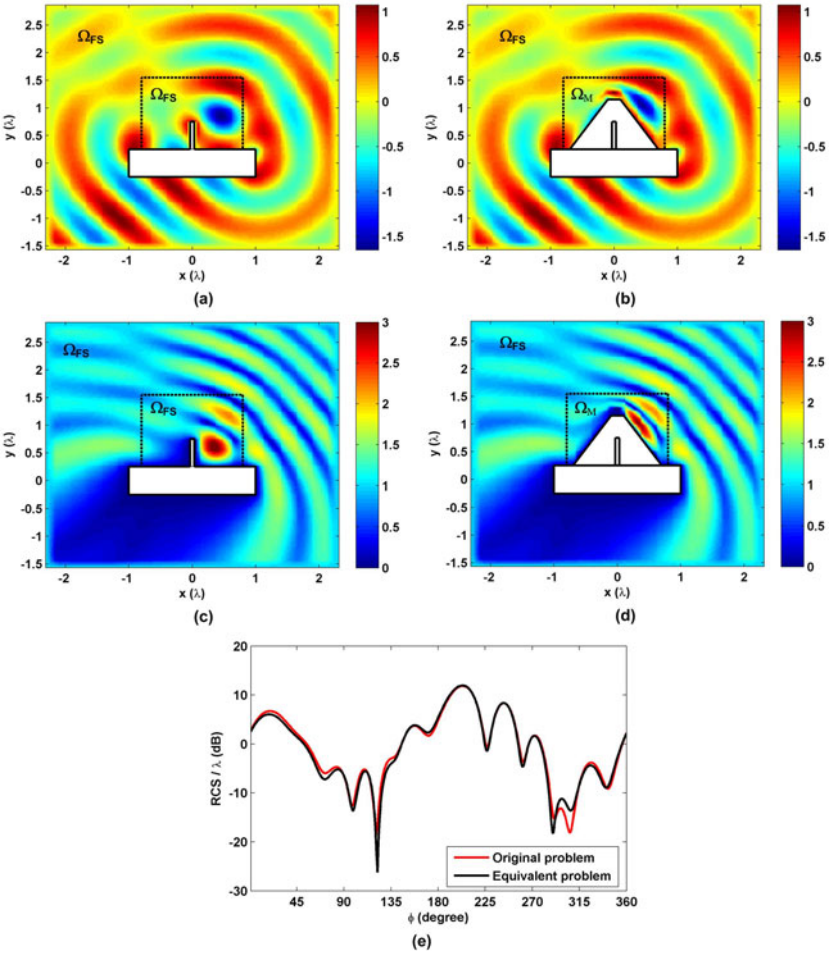


Fig. 7. Finite element simulation of scattering from an object with thin overhang [thickness: λ/15, position: x=0]: (a) Real part of scattered electric field in original problem, (b) Real part of scattered electric field in equivalent problem, (c) Magnitude of total electric field in original problem, (d) Magnitude of total electric field in equivalent problem, (e) RCS profiles.

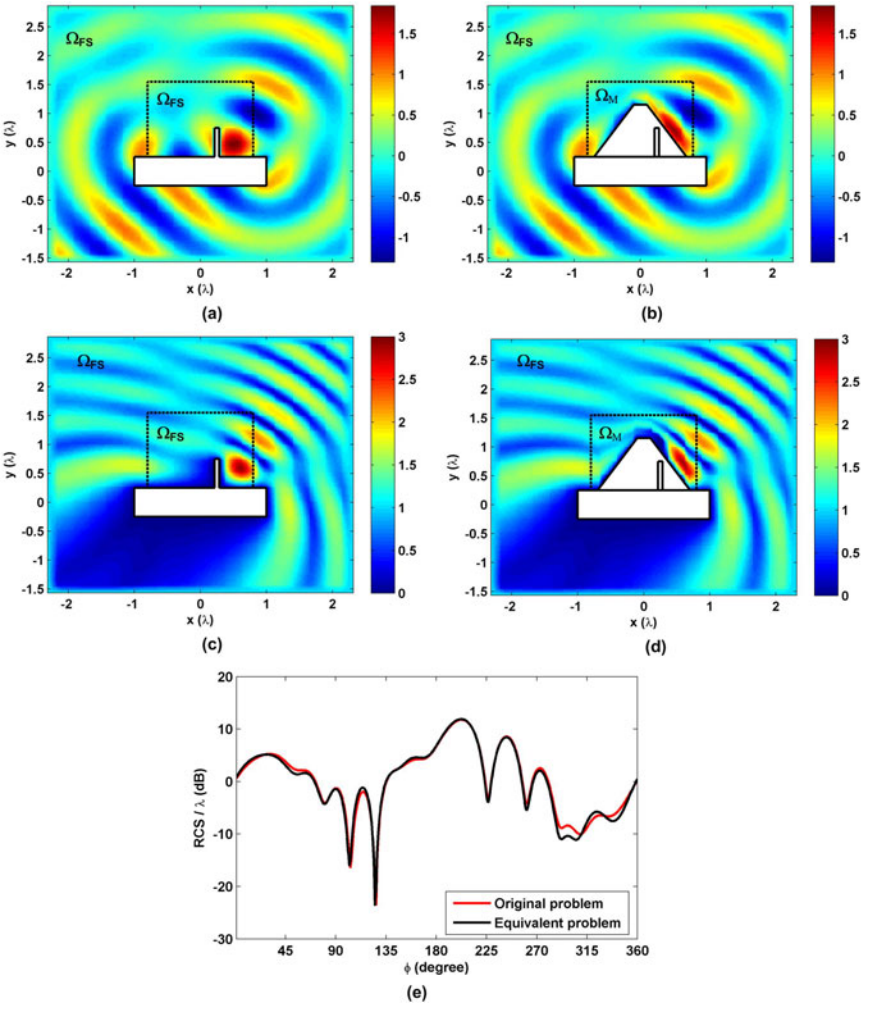


Fig. 8. Finite element simulation of scattering from an object with thin overhang [thickness: λ/15, position: x=0.25λ]: (a) Real part of scattered electric field in original problem, (b) Real part of scattered electric field in equivalent problem, (c) Magnitude of total electric field in original problem, (d) Magnitude of total electric field in equivalent problem, (e) RCS profiles.

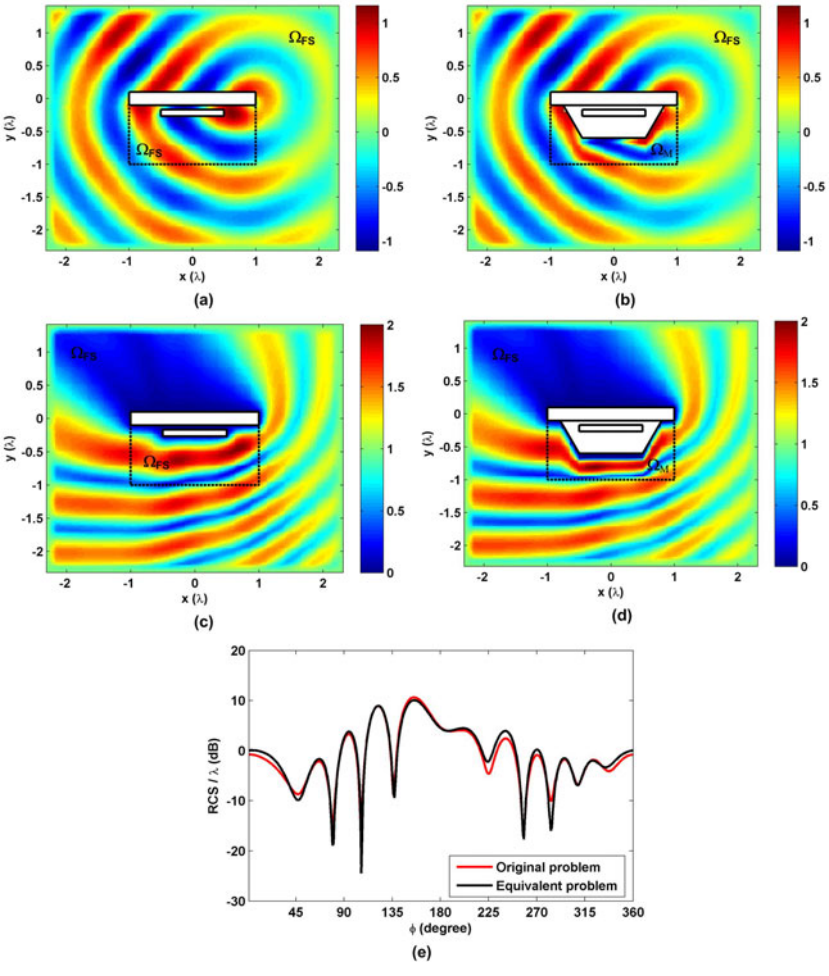


Fig. 9. Finite element simulation of scattering from objects separated by short distances [distance: λ/15]: (a) Real part of scattered electric field in original problem, (b) Real part of scattered electric field in equivalent problem, (c) Magnitude of total electric field in original problem, (d) Magnitude of total electric field in equivalent problem, (e) RCS profiles.

4 Conclusions

We have developed new simulation techniques for efficient solution of multi-scale electromagnetic problems, by placing anisotropic metamaterial structures that are designed by special coordinate transformations in the computational domain in order to overcome the numerical difficulties. We have numerically explored the functionality of all techniques with the aid of finite element simulations. We have observed good agreements between the analytical formulations and numerical simulations.

Acknowledgments. This work has been supported by the Scientific and Technical Research Council of Turkey (TUBITAK) (Project no: 109E169).

References

1. Wu, H., Cangellaris, A.C.: Efficient Finite Element Electromagnetic Modeling of Thin Wires. Microwave Opt. Technol. Lett. 50, 350–354 (2008)
2. Abakar, A., Coulomb, J.L., Meunier, G., Zgainski, F.-X., Guerin, C.: 3-D Modeling of Thin Wire and Thin Plate Using Finite Element Method and Electrical Circuit Equation. IEEE Trans. on Magnetics 37, 3238–3241 (2001)
3. Qian, Z.-G., Chew, W.C.: Fast Full-Wave Surface Integral Equation Solver for Multiscale Structure Modeling. IEEE Trans. Antennas Propagat. 57, 3594–3601 (2009)
4. Ozgun, O., Mittra, R., Kuzuoglu, M.: Finite Element / Dipole Moment Method for Efficient Solution of Multiscale Electromagnetic Problems. In: IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting, Canada, July 11–13 (2010)
5. Ozgun, O., Kuzuoglu, M.: Efficient Finite Element Solution of Low-Frequency Scattering Problems via Anisotropic Metamaterial Layers. Microwave Opt. Technol. Lett. 50, 639–646 (2008)
6. Ozgun, O., Mittra, R., Kuzuoglu, M.: Parallelized Characteristic Basis Finite Element Method (CBFEM-MPI) - A Non-iterative Domain Decomposition Algorithm for Electromagnetic Scattering Problems. J. Comput. Phys. 228, 2225–2238 (2009)
7. Pendry, J.B., Schurig, D., Smith, D.R.: Controlling electromagnetic fields. Science 312, 1780–1782 (2006)
8. Kuzuoglu, M., Mittra, R.: Investigation of nonplanar perfectly matched absorbers for finite element mesh truncation. IEEE Trans. Antennas Propagat. 45, 474–486 (1997)
9. Ozgun, O., Kuzuoglu, M.: Non-Maxwellian Locally-conformal PML Absorbers for Finite Element Mesh Truncation. IEEE Trans. Antennas Propagat. 55, 931–937 (2007)
10. Kong, F., Wu, B.I., Kong, J.A., Huangfu, J., Xi, S., Chen, H.: Planar focusing antenna design by using coordinate transformation technology. Applied Physics Letters 91, 253–509 (2007)
11. Donderici, B., Teixeria, F.L.: Metamaterial blueprints for reflectionless waveguide bends. IEEE Microwave and Wireless Components Letters 18, 233–235 (2008)
12. Vasic, B., Isic, G., Gajic, R., Hingerl, K.: Coordinate transformation based design of confined metamaterial structures. Physical Review B 79, article no. 085103 (2009)
13. Tichit, P.H., Burokur, S.N., Lustrac, A.: Ultradirective antenna via transformation optics. Journal of Applied Physics 105, article no. 104912 (2009)
14. Ozgun, O., Kuzuoglu, M.: Electromagnetic metamorphosis: Reshaping scatterers via conformal anisotropic metamaterial coatings. Microwave Opt. Technol. Lett. 49, 2386–2392 (2007)
15. Ozgun, O., Kuzuoglu, M.: Form-invariance of Maxwell's Equations in Waveguide Cross-section Transformations. Electromagnetics 29, 353–376 (2009)
16. Ozgun, O., Kuzuoglu, M.: Efficient finite element solution of low-frequency scattering problems via anisotropic metamaterial layers. Microwave Opt. Technol. Lett. 50, 639–646 (2008)
17. Ozgun, O., Kuzuoglu, M.: Domain Compression via Anisotropic Metamaterials designed by Coordinate Transformations. J. Comput. Phys. 229, 921–932 (2010)
18. Lindell, I.V.: Methods for Electromagnetic Field Analysis. Oxford University Press, Oxford (1992)

Coupled Finite Element - Scaled Boundary Finite Element Method for Transient Analysis of Dam-Reservoir Interaction

Shangming Li

Institute of Structural Mechanics, China Academy of Engineering Physics,
Mianyang City, Sichuan Province, 621900, China
hustmingsl@126.com

Abstract. The scaled boundary finite element method (SBFEM) was extended to solve dam-reservoir interaction problems in the time domain, where dams were flexible and the fluid in reservoir was semi-infinite and compressible. Transient responses of dam-reservoir systems subjected to horizontal ground motions were analyzed based on the SBFEM and finite element method (FEM) coupling method. A dam was modeled by FEM, while the whole fluid in reservoir was modeled by the SBFEM alone or a combination of FEM and SBFEM. Two benchmark examples were considered to check the accuracy of the SBFEM-FEM coupling method. For a vertical dam-reservoir system, the semi-infinite fluid with a uniform cross section was modeled by the SBFEM alone. For non-vertical dam-reservoir systems, the fluid was divided into a near-field FEM domain and a far-field SBFEM domain. The geometry of near field is arbitrary, and the far field is a semi-infinite prism. Their numerical results obtained through the presented method were compared with those from analytical or substructure methods and good agreements were found.

Keywords: SBFEM; SBFEM-FEM coupling; dam-reservoir system; ground motions; transient analysis.

1 Introduction

A dam-reservoir system subjected to ground motions is often a major concern in the design. To ensure that dams are adequately designed for, the hydrodynamic pressure distribution along the dam-reservoir interface must be determined for assessment of safety. The assessment of safety of dam should include frequency and time domain analyses of dam-reservoir system.

In time-domain analyses, since an analytical solution is not available for dam-reservoir system with general or arbitrary geometry, numerical methods are often adopted. In numerical methods, a semi-infinite reservoir is usually divided into two parts: a near-field and a far-field domain. The near-field domain can be efficiently modeled by finite element method (FEM), while the effect of far-field domain can be represented by some kinds of transmitting boundary conditions (TBC) at the near-far-field interface. The most commonly-used TBC is Sommerfeld radiation condition [1],

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 26–34, 2011.

which is very simply, but precautions have to be taken as the approach may introduce significant errors when the near-field domain is small. Sharan's TBC was developed [2]. Although it produced better results than Sommerfeld's, it does not represent the behaviour of the far-field domain well when the near-to-far-field interface is too near to the dam-reservoir interface. An efficient semi-analytical TBC [3] and later, its application [4] exhibited good results, but it required the full eigen-modes of the near-to-far-field interface. Other types of TBC were also developed [5-7]. Except for the TBC, boundary element method (BEM) was often adopted to model a semi-infinite reservoir [8-10]. BEM can generally yield more accurate results than TBC, but it requires a fundamental solution, which affects its applications.

In this study, the scaled boundary finite element method (SBFEM) was chosen to model the far-field domain. The SBFEM, which not only combines the advantages of FEM and BEM, but also avoids the disadvantages of BEM, was developed in [11] for an infinite soil-structure interaction. The efficiency for applications involving an infinite medium was also validated in [11]. The SBFEM was extended to solve certain infinite fluid-structure interaction problems in [12], where the fluid medium was not layered. For layered fluid medium, some works about frequency analysis based on the SBFEM [13,14] for dam-reservoir systems were done. A hydrodynamic pressure on rigid dams was analyzed based on SBFEM for layered incompressible fluid medium. In the time domain, the SBFEM includes convolution integrals, which greatly affects its analysis efficiency. To improve the SBFEM analysis efficiency, a continued-fraction formulation [16] of SBFEM for time-domain analysis was proposed to avoid evaluating convolution integrals, and a diagonalization procedure [17] was proposed, whose calculation efficiency was found to be high, although it still included convolution integrals. With the improvement of SBFEM evaluation efficiency in the time domain, the SBFEM for time analysis will show gradually its advantages. An alternative method [18] for obtaining the responses in the time domain is to obtain the frequency response first and then transform it into the time response through the inverse Fourier transform. As the inverse Fourier transform was used and the eigenvalue problem of a coefficient matrix must be solved to obtain the frequency response for each excitation frequency, the accuracy and efficiency of the alternative method need further improved. Therefore, based on the diagonaliztion procedure [17], this paper presented a SBFEM-FEM coupling method to solve transient analysis of deformable dam-reservoir systems. Numerical results showed its accuracy.

2 Problem Statement

A dam-reservoir system shown in Fig.1 was considered, where a dam is deformable and the fluid in reservoir is inviscid isentropic with fluid particles undergoing only small displacements. Analytical solutions of dams with arbitrary geometry in the time domain have not been reported in literatures. To circumvent the difficulty, the fluid in reservoir is divided into two parts: a near-field and a far-field domain as shown in Fig.1. The interaction between the near-field domain and the far-field domain occurs at Interface 2, while the interaction between the dam and the fluid occurs at Interface 1. Of note, Interface 2 is chosen to be vertical and the reservoir bottom in the far-field domain is level and rigid. These assumptions are necessary in SBFEM formulations, because the SBFEM requires no energy to be radiated from the infinity towards the dam.

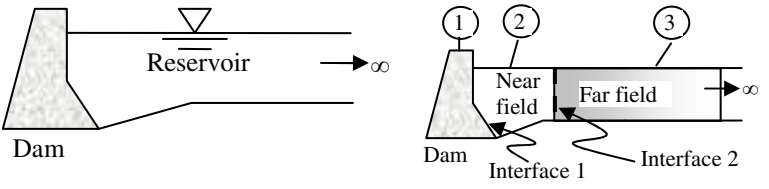


Fig. 1. Dam-reservoir system

3 Motion Equation of Dam

The motion equation for a dam subjected to both ground motions and external forces can be written in the standard finite element form as follows.

$$\mathbf{M}^d\ddot{\mathbf{U}} + \mathbf{C}^d\dot{\mathbf{U}} + \mathbf{K}^d\mathbf{U} = -\mathbf{M}^d\mathbf{R}\ddot{\mathbf{U}}_g + \mathbf{F}_e + \mathbf{F}_f \tag{1}$$

where $\mathbf{M}^d$, $\mathbf{C}^d$ and $\mathbf{K}^d$ denote the global mass, damping and stiffness matrices for the solid dam, respectively; $\ddot{\mathbf{U}}$, $\dot{\mathbf{U}}$ and $\mathbf{U}$ are the vectors of the nodal acceleration, velocity and displacement of dams, respectively; $\ddot{\mathbf{U}}_g$, $\mathbf{F}_e$ and $\mathbf{F}_f$ are the ground acceleration vector, the external force vector and the hydrodynamic force vector, respectively; and $\mathbf{R}$ denotes the acceleration transformation matrix. In Eq.(1), all matrices and vectors except $\mathbf{U}$ and $\mathbf{F}_f$ can be derived in the standard manner using the traditional finite element procedures. It leads to an expression such that the displacement $\mathbf{U}$ is a function of the hydrodynamic force $\mathbf{F}_f$.

4 FE Equation of Near Field

For a dam-reservoir system, the force term $\mathbf{F}_f$ on the right hand side of Eq.(1) is the force derived from the reservoir. It can be expressed as

$$\mathbf{F}_f = \sum_e \int_{\Gamma_1^e} \mathbf{N}^T \mathbf{N}_f \mathbf{p}_1 d\Gamma_1^e \tag{2}$$

where $\mathbf{N}$ and $\mathbf{N}_f$ denote the shape functions of a typical solid element for a dam and of a typical element for the near-field fluid medium, respectively; Γ_1 denotes Interface 1; the pressure $\mathbf{p}_1$ is a nodal pressure column vector, which is obtained from the near-field fluid domain. Note that $\mathbf{N}$ is not the same as $\mathbf{N}_f$.

After partitioning it into sub-matrices corresponding to variables at Interface 1, Interface 2 and other interior locations, the FE equation of Near field is written as

$$\begin{bmatrix} \mathbf{m}_{11}^f & \mathbf{m}_{12}^f & \mathbf{m}_{13}^f \\ \mathbf{m}_{21}^f & \mathbf{m}_{22}^f & \mathbf{m}_{23}^f \\ \mathbf{m}_{31}^f & \mathbf{m}_{32}^f & \mathbf{m}_{33}^f \end{bmatrix} \begin{Bmatrix} \ddot{\mathbf{\Phi}}_1 \\ \ddot{\mathbf{\Phi}}_2 \\ \ddot{\mathbf{\Phi}}_3 \end{Bmatrix} + \begin{bmatrix} \mathbf{k}_{11}^f & \mathbf{k}_{12}^f & \mathbf{k}_{13}^f \\ \mathbf{k}_{21}^f & \mathbf{k}_{22}^f & \mathbf{k}_{23}^f \\ \mathbf{k}_{31}^f & \mathbf{k}_{32}^f & \mathbf{k}_{33}^f \end{bmatrix} \begin{Bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \mathbf{\Phi}_3 \end{Bmatrix} = \begin{Bmatrix} \mathbf{V}_{n1} \\ \mathbf{V}'_{n2} \\ \mathbf{V}_{n3} \end{Bmatrix} \tag{3}$$

where the subscripts 1 and 2 refer to nodal variables at Interfaces 1 and 2, respectively; while the subscript 3 refers to other interior nodal variables in the near-field fluid domain. $\mathbf{m}$, $\mathbf{k}$ are the mass and stiffness matrices of Near field. $\mathbf{\Phi}$, $\mathbf{V}_n$ are the velocity potential and efficient nodal velocity. They can be obtained by traditional finite element procedures. At Interface 2, the near-field FEM-domain couples with the far-field SBFEM-domain.

5 SBFEM-FEM Coupling Formulation for Near Field

For Far field shown in Fig.1, its SBFEM discretization mesh is plotted in Fig.2. Each element at Interface 2 represents a sub-domain (i.e. sub-semi-infinite layered medium), so that the whole far-field domain is represented by an assemblage of elements at Interface 2 and its dynamic characteristics is described by the following SBFEM formulation

$$\mathbf{V}_{n2}(t) = \int_0^t \mathbf{M}^{\infty}(t-\tau)\ddot{\mathbf{\Phi}}_2(\tau)d\tau \tag{4}$$

$\mathbf{M}^{\infty}(t)$ is the dynamic mass matrix of the whole far-field domain. Upon discretization of Eq.(4) with respect to time and assuming all initial conditions equal to zero, one can get the following equation.

$$\mathbf{V}_{n2}^n = \mathbf{M}_1^{\infty}\dot{\mathbf{\Phi}}_2^n + \sum_{j=1}^{n-1}\left(\mathbf{M}_{n-j+1}^{\infty} - \mathbf{M}_{n-j}^{\infty}\right)\dot{\mathbf{\Phi}}_2^j \tag{5}$$

$\mathbf{M}_{n-j+1}^{\infty} = \mathbf{M}^{\infty}\left((n-j+1)\Delta t\right)$, $\mathbf{\Phi}_2^j = \mathbf{\Phi}_2(j\Delta t)$ and $\mathbf{V}_{n2}^n = \mathbf{V}_{n2}(n\Delta t)$ where Δt denotes an increment in time step. Details about $\mathbf{M}^{\infty}(t)$ can be found in [11, 17].

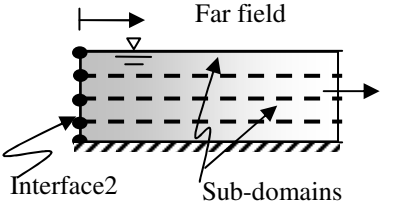


Fig. 2. SBFEM mesh of far field

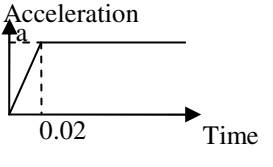


Fig. 3. Horizontal ramped acceleration

According to the kinematic continuity at Interface 2, one has

$$-\mathbf{V}'^{n}_{n2}=\mathbf{V}^{n}_{n2}=\mathbf{M}^{\infty}_{1}\dot{\mathbf{\Phi}}^{n}_{2}+\sum_{j=1}^{n-1}\left(\mathbf{M}^{\infty}_{n-j+1}-\mathbf{M}^{\infty}_{n-j}\right)\dot{\mathbf{\Phi}}^{j}_{2} \tag{6}$$

Substituting Eq.(6) into Eq.(3) and re-arranging leads to

$$\begin{bmatrix}\mathbf{m}^{f}_{11} & \mathbf{m}^{f}_{12} & \mathbf{m}^{f}_{13}\\ \mathbf{m}^{f}_{21} & \mathbf{m}^{f}_{22} & \mathbf{m}^{f}_{23}\\ \mathbf{m}^{f}_{31} & \mathbf{m}^{f}_{32} & \mathbf{m}^{f}_{33}\end{bmatrix}\begin{Bmatrix}\ddot{\mathbf{\Phi}}^{n}_{1}\\ \ddot{\mathbf{\Phi}}^{n}_{2}\\ \ddot{\mathbf{\Phi}}^{n}_{3}\end{Bmatrix}+\begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{M}^{\infty}_{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{Bmatrix}\dot{\mathbf{\Phi}}^{n}_{1}\\ \dot{\mathbf{\Phi}}^{n}_{2}\\ \dot{\mathbf{\Phi}}^{n}_{3}\end{Bmatrix}$$
$$+\begin{bmatrix}\mathbf{k}^{f}_{11} & \mathbf{k}^{f}_{12} & \mathbf{k}^{f}_{13}\\ \mathbf{k}^{f}_{21} & \mathbf{k}^{f}_{22} & \mathbf{k}^{f}_{23}\\ \mathbf{k}^{f}_{31} & \mathbf{k}^{f}_{32} & \mathbf{k}^{f}_{33}\end{bmatrix}\begin{Bmatrix}\mathbf{\Phi}^{n}_{1}\\ \mathbf{\Phi}^{n}_{2}\\ \mathbf{\Phi}^{n}_{3}\end{Bmatrix}=\begin{Bmatrix}\mathbf{V}^{n}_{n1}\\ -\sum_{j=1}^{n-1}\left(\mathbf{M}^{\infty}_{n-j+1}-\mathbf{M}^{\infty}_{n-j}\right)\dot{\mathbf{\Phi}}^{j}_{2}\\ \mathbf{V}^{n}_{n3}\end{Bmatrix} \tag{7}$$

where the superscript n denotes the instant at time $t=n\Delta t$. Note that a damping matrix appears on the left hand side of Eq.(7). It can be regarded as the damping effect derived from the far-field medium and imposed on the dam-reservoir system. Upon determining the velocity potential $\mathbf{\Phi}$ through Eq.(7), the fluid pressure $\mathbf{p}_1$ can be evaluated by

$$\mathbf{p}_1=-\rho\dot{\mathbf{\Phi}}_1 \tag{8}$$

Note that $\mathbf{p}_1$ depends on the normal velocity v_{n1} at Interface 1, while the solution of Eq.(1) including v_{n1} also depends on $\mathbf{p}_1$. Therefore, Eq.(1) and Eq.(7) form a coupling system. Following the numerical procedure with a Newmark scheme, one can solve the coupling system by assuming

$$\mathbf{p}_1(t)=f\left(v^{t}_{n1}\right)\approx f\left(v^{t-\Delta t}_{n1}\right) \tag{9}$$

where f denotes a function. If the time increment Δt is small, accurate results can be obtained. However, if Δt is relatively large, an iteration scheme should be used in each time step, i.e., a term ${}^{t}v^{j-1}_{n1}$ is used instead of $v^{t-\Delta t}_{n1}$ in the global iteration scheme, where j denotes the jth iteration within a time step. Its corresponding formulation can be found in [12].

6 Numerical Examples

Consider transient responses of dam-reservoir systems where dams were subjected to horizontal ground acceleration excitations shown in Fig.3 and Fig.4. In the transient analysis, only the linear behavior was considered, the free surface wave effects and the reservoir bottom absorption were ignored, and the damping of dams was excluded. Newmark's time-integration scheme was used to solve FE equations (Eq.(1) and Eq.(7)). Newmark integration parameters $\alpha=0.25$ and $\delta=0.5$.

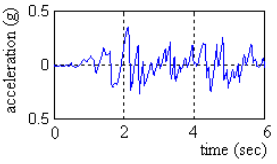


Fig. 4. El Centro N-S horizontal acceleration

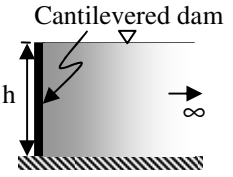
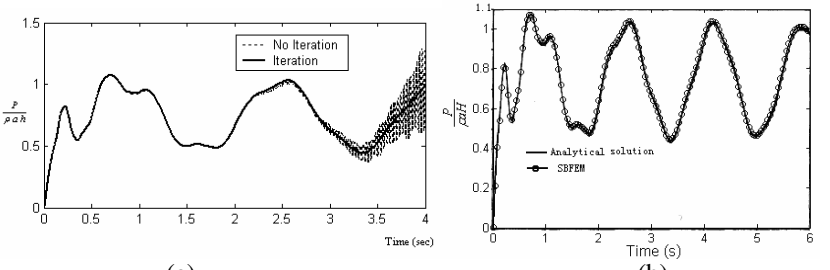


Fig. 5. Vertical dam-reservoir system

6.1 Vertical Dam

As the cross section of the vertical dam-system as shown in Fig.5 is uniform, a near-field fluid domain is not necessary and the whole reservoir can be modeled by a far-field domain alone. Its transient response was solved through coupling Eqs.(1) and (5). Sound speed in the reservoir is 1438.656*m/s* and the fluid density ρ is 1000kg/m^3. The weight per unit length of the cantilevered dam is 36000*kg/m*. The height of the cantilevered dam *h* is equal to 180*m*. The dam was modeled by 20 numbers of simple 2-noded beam elements with rigidity EI (=9.646826x10^{13} Nm^2), while the whole fluid domain was modeled by 10 numbers of SBFEM 3-noded elements, whose nodes matched side by side with nodes of the dam. In this problem, the shear deformation effects were not included in the 2-noded beam elements. Time step increment was 0.005*sec*. Results for the pressure at the heel of dam for ramped acceleration obtained by using a "no-iteration" scheme in Eq.(9) and an iteration scheme were plotted in Fig.6a. Results at late time from no-iteration scheme were divergent, whereas results obtained by both schemes were the same at early time. As such, an iteration scheme was adopted in the following studies. Pressures at the heel of dam obtained from the SBFEM coupling FEM procedure and an analytical method [3] were shown in Fig.6b. The SFEM-FEM solutions were very close to analytical solutions.



(a) (b)

Fig. 6. Pressures at the heel of dam subjected to horizontal ramped acceleration

6.2 Gravity Dam

This example was analyzed to verify the accuracy and efficiency of the SBFEM-FEM coupling formulation (Eq.(1) and Eq.(7)) for a dam-reservoir system having arbitrary slopes at the dam-reservoir interface. The density, Poisson's ratio and Young's

modulus of the dam are 2400 kg/m^3, 0.2 and $3.43x10^{10}$ N/m^2, respectively. The fluid density ρ is 1000 kg/m3, and wave speed in the fluid is 1438.656*m/s*. The height of the dam *H* is 120*m*. A typical gravity-dam-reservoir system and its FEM and SBFEM meshes were shown in Fig.7.The dam and the near-field fluid were discretized by FEM, while the far-field fluid was discretized by the SBFEM. 40 numbers and 20 numbers of 8-noded elements were used to model the dam and the near-field fluid domain, respectively, while 10 numbers of 3-noded SBFEM elements were employed to model the whole far-field fluid domain. Note that the size of the near-field fluid domain can be very small, which was proven in the literature [14]. In this example, the distance between the heel of the dam and the near-far-field interface was 6*m* (=0.05*H*). The pressure at the heel of the gravity dam caused by the ground acceleration shown in Figs.3 and 4 were plotted in Fig.8. The time increment was 0.002*sec*. Results from SBFEM and FEM coupling procedure were very close to solutions from the sub-structures method [4].

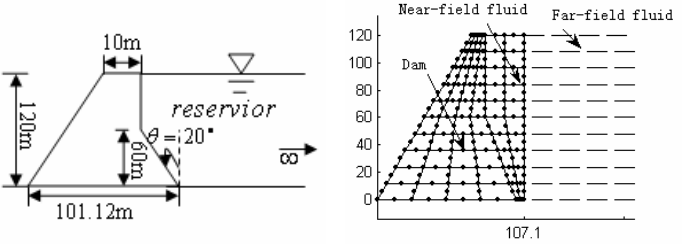
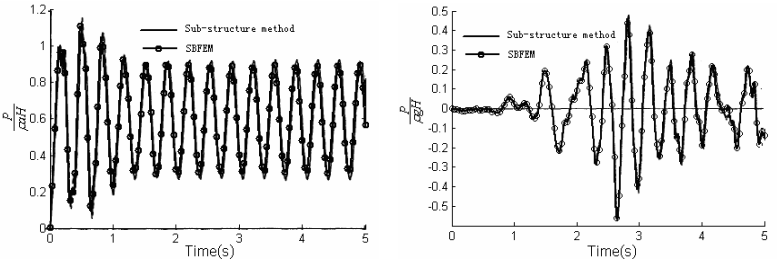


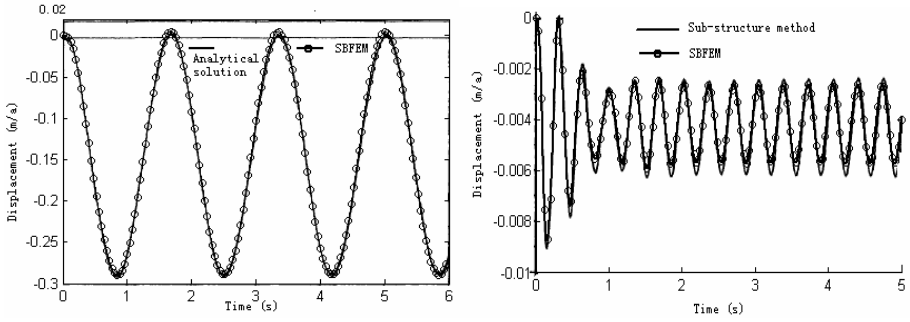
Fig. 7. Gravity dam-reservoir system and its FEM-SBFEM mesh



(a) Ramped acceleration (b) El Centro acceleration

Fig. 8. Pressure at the heel of gravity dam subjected to horizontal acceleration

The displacements at the top of vertical and gravity dams subjected to ramped acceleration were plotted in Fig.9. The displacement solutions of vertical dam from the presented method are the same with analytical solutions. Fig.10 showed the displacement at the top of gravity dam subjected to the El Centro N-S horizontal acceleration. The displacements obtained by the present method agreed well with sub-structure method's results [4], especially at early time.



(a) Vertical dam (b) Gravity dam

Fig. 9. Displacement at top of dam subjected to horizontal ramped acceleration

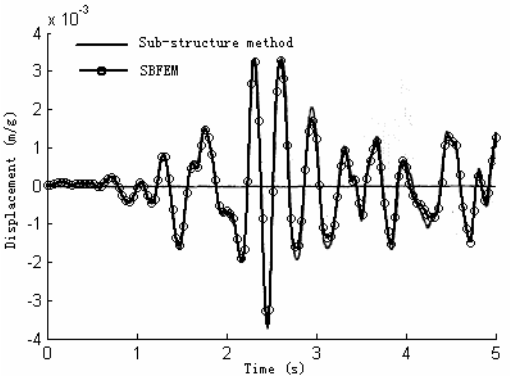


Fig. 10. Displacement at top of gravity dam subjected to horizontal El Centro acceleration

7 Conclusions

The SBFEM was employed in conjunction with FEM to solve the dam-reservoir interaction problems in the time domain. A coupled FE-SBFEM formulation was presented, which included a damping matrix induced by the damping effect of semi-infinite reservoir. The merits of the SBFEM in representing the unbounded fluid medium were illustrated through comparisons against benchmark solutions. Numerical results showed that its accuracy and efficiency of the coupled FE-SBFEM formulation for the transient analysis of dam-reservoir system. Of note, the SBFEM is a semi-analytical method. Its solution in the radial direction is analytical so that only a near field with a small volume is required. Compared to the sub-structure method [4], its formulations are in a simpler mathematical form and can be coupled with FEM easily and seamlessly.

Acknowledgments. This research is supported by the National Natural Science Foundation of China (No. 10902060) and China Postdoctoral Science Foundation (201003123), for which the author is grateful.

References

1. Sommerfeld, A.: Partial differential Equations in physics. Academic Press, New York (1949)
2. Sharan, S.K.: Finite element analysis of unbounded and incompressible fluid domains. International Journal for Numerical Methods in Engineering 21(9), 1659–1669 (1985)
3. Tsai, C.S., Lee, G.C., Ketter, R.L.: A semi-analytical method for time-domain analyses of dam-reservoir interactions. International Journal for Numerical Methods in Engineering 29, 913–933 (1990)
4. Tsai, C.S., Lee, G.C.: Time-domain analyses of dam-reservoir system. II: substructure method. Journal of Engineering Mechanics 117(9), 2007–2026 (1991)
5. Maity, S.K., Bhattacharyya, S.K.: Time-domain analysis of infinite reservoir by finite element method using a novel far-boundary condition. Finite Elements in Analysis and Design 32(2), 85–96 (1999)
6. Kucukarslan, S.: An exact truncation boundary condition for incompressible-unbounded infinite fluid domains. Applied Mathematics and Computation 163(1), 61–69 (2005)
7. Gogoi, I., Maity, D.: A non-reflecting boundary condition for the finite element modeling of infinite reservoir with layered sediment. Advances in Water Resources 29, 1515–1527 (2006)
8. Czygan, O., Von, E.O.: Fluid-structure interaction by coupling BEM and nonlinear FEM. Engineering Analysis with Boundary Elements 26, 773–779 (2002)
9. Dominguez, J., Maeso, O.: Earthquake analysis of arch dams. dam-water-foundation interaction. Journal of Engineering Mechanics-ASCE 119, 513–530 (1993)
10. Dominguez, J., Gallego, R., Japon, B.R.: Effects of porous sediments on seismic response of concrete gravity dams. Journal of Engineering Mechanics-ASCE 123, 302–311 (1997)
11. Wolf, J.P., Song, C.: Finite-Element Modeling of Unbounded Media. Wiley, Chichester (1996)
12. Fan, S.C., Li, S.M., Yu, G.Y.: Dynamic fluid-structure interaction analysis using boundary finite element method/finite element method. Journal of Applied Mechanics 72(4), 591–598 (2005)
13. Fan, S.C., Li, S.M.: Boundary finite-element method coupling finite-element method for steady-state analyses of dam-reservoir systems. Journal of Engineering Mechanics-ASCE 134, 133–142 (2008)
14. Li, S.M., Liang, H., Li, A.M.: A semi-analytical solution for characteristics of a dam-reservoir system with absorptive reservoir bottom. Journal of Hydrodynamics 20(6), 727–734 (2008)
15. Du, J.G., Lin, G., Hu, Z.Q.: Hydrodynamic pressure on dams based on SBFEM. Geotechnical Special Publication, Num. 151, 262–269 (2006)
16. Bazyar, M.H., Song, C.M.: A continued-fraction-based high-order transmitting boundary for wave propagation in unbounded domains of arbitrary geometry. International Journal for Numerical Methods in Engineering 74, 209–237 (2008)
17. Li, S.M.: Diagonalization procedure for scaled boundary finite element method in modelling semi-infinite reservoir with uniform cross section. International Journal for Numerical Methods in Engineering 80(5), 596–608 (2009)
18. Wang, Y., Lin, G., Hu, Z.Q.: A Coupled FE and Scaled Boundary FE-Approach for the Earthquake Response Analysis of Arch Dam-Reservoir-Foundation System. In: IOP Conference Series: Materials Science and Engineering, vol. 10, pp. 1–10 (2010)

A Comparison of Different Advective Solvers in the CHIMERE Air Quality Model

Pedro Molina[1], Luis Gavete[1], Marta García Vivanco[2], Inmaculada Palomino[2], M. Lucía Gavete[3], Francisco Ureña[4], and Juan José Benito[5]

[1] Universidad Politécnica de Madrid, Spain
[2] C.I.E.M.A.T. Madrid, Spain
[3] Universidad Rey Juan Carlos, Madrid, Spain
[4] Universidad Castilla-la Mancha, Ciudad Real, Spain
[5] U.N.E.D., Madrid, Spain
p.molina@upm.es, lu.gavete@upm.es, m.garcia@ciemat.es, inma.palomino@ciemat.es, lucia.gavete@urjc.es francisco.urena@uclm.es, jbenito@ind.uned.es

Abstract. Extensive research has been performed to solve the advection equation and different numerical methods have been proposed. Most of these methods, including semi-lagrangian methods, are not conservative. In this paper we compare in the CHIMERE eulerian chemistry transport model different conservative algorithms for solving the advection equation. The numerical results are compared with a set of observation sites in the area of Spain and some conclusions are obtained.

Keywords: advection equation, conservative scheme, Chimère.

1 Introduction

This paper is focused on air pollution modeling, a challenging scientific problem, specially in many industrialized countries where there is an increasing consciousness of the effect, on health and environment, of the emissions of pollutants into the urban atmospheres. The environmental benefits of reliable air quality forecasts are obvious: populations can be more efficiently protected by means of information or real-time emission abatement strategies.

Air pollution modeling is based on the assumption of no reciprocal effect of the chemical species on flow fields (wind velocity, turbulent diffusivity, temperature). After having pre-processed the flow fields by meteorological computations or parametrizations, the reaction-advection-diffusion PDE (Partial Differential Equation), that is, the mass continuity equation, is solved to estimate the concentrations of chemical species

$$\frac{\partial f}{\partial t} + \nabla\left(uf\right) = \nabla\left(k\nabla f\right) + P - L \tag{1}$$

.

In this equation, characteristic of the Eulerian approach, f is a vector containing the concentrations of all model species for every grid box, u is the three dimensional

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 35–49, 2011.

wind vector, k the tensor of eddy diffusivity and P and L represent production and loss terms due to chemical reactions, emissions and deposition.

A class of conservative schemes for the advection equation has been so far proposed following the pioneering work of Godunov [1]. A Godunov type schemes computes the cell-integrated average values of a prognostic variable by using a conservative finite difference method of flux form, or a finite volume method as preferred by some researchers, and results in an exact conservation for the transported quantity. High order Godunov schemes can be devised by reconstructing high order interpolations within each mesh cell.

Rather than the piecewise constant interpolation in the original Godunov scheme, a linear interpolation function and a parabolic polynomial were used in the MUSCL[2,3] and the piecewise parabolic method (PPM) [4] schemes.

Other conservative schemes including a rational method can also be employed to solve the advection equation [5].

Basically, using a linear or higher order interpolation function for reconstruction tends to bring about oscillations to the numerical solutions. To get rid of this, slope modifications and the adjustments of the cell-interface values prove to be necessary. As a result, the piece-wisely constructed interpolation function is not usually continuous cross the cell interfaces.

Other studies have been carried out with the CHIMERE model but using only the PPM numerical solver[15].

Our main goal of this research is to compare three different algorithms for the transport module included in the European scale Eulerian chemistry transport model CHIMERE. The results of the different methods are compared with a set of observation sites in the area of the Iberian Peninsula in Spain.

Section 2 introduces the conservative methods that have been evaluated. In section 3 we introduce the European-scale chemistry-transport model (CHIMERE). The comparison of observed and modeled data is given in Section 4, and finally some conclusions are given in Section 5.

2 Modeling of Linear Advection 1D

For simplicity of presentation we start with the scalar advection problem in one space dimension. We note f as the concentration of one typical atmospheric pollutant. The advection in one space dimension of this pollutant, during the interval $[0,T]$ is given by the following linear hyperbolic equation, called also transport equation, to which we add an initial concentration; the overall Cauchy problem is consequently the following one.

Given a field of initial concentration $f(x,0) = f_0$, a velocity wind u and a time T > 0, we want to calculate $f(x,t)$ such that

$$\begin{cases} \dfrac{\partial f}{\partial x} + \dfrac{\partial (uf)}{\partial x} = 0 & \forall (x,t) \in \mathbb{R} \times [0,T] \\ f(x,0) = f_0(x) & \forall (x,t) \in \mathbb{R} \end{cases} \tag{2}$$

In the following sections it will be showed the three different finite volume methods to solving it.

The finite volume method is applied to the conservative form of the transport equation in one dimension in space. First we divide the spatial domain in cells called finite or controls volumes, this corresponds in one dimension to a partition of $[0,L] \subset \mathbb{R}$ by intervals, on the other hand one built discrete equations from the integral form of the equation.

Let be $\left(\left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}} \right] \right)_{i \in \mathbb{Z}}$ a partition of $[0,L]$ and $([t^n, t^{n+1}])_{n \in \mathbb{N}}$ a regular partition of $[0,T]$. We define steps of time and space respectively written $\Delta t = t^{n+1} - t^n$ and $\Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, the middle points $x_i = \frac{1}{2}\left(x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}} \right)$ and the control volume $\Omega_i = \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}} \right]$.

By integrating (2) on $\Omega_i \times \left[t^n, t^{n+1} \right]$, we obtain the integral form of the conservation law

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta t}{\Delta x_i}\left(g^n_{i+\frac{1}{2}} - g^n_{i-\frac{1}{2}} \right) . \tag{3}$$

where we define the exact flux by

$$g^n_{i+\frac{1}{2}} = \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} (uf)\left(x_{i+\frac{1}{2}}, t \right) dt \tag{4}$$

and the average values of the exact solution, at the time t^n, on each cell by

$$\rho_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} F(x, t^n) dx \ . \tag{5}$$

where F is the approximation of the function.

This equation is the finite volume formulation of the problem and belongs to the general framework of conservative schemes. In fact the conservative property means that the mass of specie A (with concentration f) is preserved (when time t move from t^n to t^{n+1}) only according to the boundaries conditions.

Three different approximations F_i are used in this paper corresponding to constant, linear and quadratic approximations of F_i over each one of the central cells.

By using time-splitting the method can be easily extended to solve advection equation in two and three dimensions. For example in two dimensions the time splitting is equivalent to do the transport of particles to the direction (Ox) and then according to the other direction (Oy).

2.1 The Upwind Method

The Godunov Method, or Upwind Method use a constant function a_i which is expressed in a generic mesh element with boundaries $x_{i-1/2}$, and $x_{i+1/2}$, and taking into account the velocity, u, the following constant approximation is used

$$\begin{aligned} F_i(x) &= a_i \quad \text{if } \mathrm{u} > 0 \\ F_i(x) &= a_{i-1} \quad \text{if } \mathrm{u} < 0 \end{aligned} \qquad \text{for } x \in \left[x_{i-1/2}, x_{i+1/2} \right] \tag{6}$$

where a_i is the value of the function in the initial moment where this value is known

$$F_i(x_{i-1/2}) = f^n_{i-1/2} = a_i \tag{7}$$

We can calculate the value of the flux for u>0, as follows

$$\begin{aligned} g^n_{i+\frac{1}{2}} &= \frac{1}{\Delta t} \int_{x_{i+1/2}-u^n_{i+1/2}\Delta t}^{x_{i+1/2}} F(x)dx = \frac{1}{\Delta t} \int_{x_{i+1/2}-u^n_{i+1/2}\Delta t}^{x_{i+1/2}} a_i dx = \frac{a_i}{\Delta t}(x)\Big|_{x_{i+1/2}-u^n_{i+1/2}\Delta t}^{x_{i+1/2}} = \\ &= \frac{a_i}{\Delta t}(x_{i+1/2} - (x_{i+1/2} - u^n_{i+1/2}\Delta t)) = \frac{a_i}{\Delta t} u^n_{i+1/2}\Delta t = u^n_{i+1/2} a_i \end{aligned} \tag{8}$$

where the superindex corresponds to the time step.

Then, we can completely explicit the flux function for the original Godunov method with respect to the value of the cell average integral ρ^n_i as follows

$$g_{i+\frac{1}{2}} = \begin{cases} \mathrm{u}^n_{i+\frac{1}{2}} \rho^n_i, & \text{if } \mathrm{u}^n_{i+\frac{1}{2}} > 0 \\ \mathrm{u}^n_{i+\frac{1}{2}} \rho^n_{i+\frac{1}{2}}, & \text{if } \mathrm{u}^n_{i+\frac{1}{2}} < 0 \end{cases} \tag{9}$$

2.2 VanLeer Method

The VanLeer method better known as MUSCLE or Monotocnic Upstream-Centered Scheme for Conservation Laws, using a minmod technique, use a linear function that is expressed in a generic mesh element with boundaries $x_{i-1/2}$, and $x_{i+1/2}$, and considering the velocity u < 0 as

$$F_i(x) = a_i + b_i(x - x_{i-1/2}) \quad \text{for } x \in \left[x_{i-1/2}, x_{i+1/2} \right] \tag{10}$$

where a_i, and b_i are the coefficients of the interpolation function.

a_i is calculated by the initial condition of each cell by

$$\begin{aligned} F_i(x_{i-1/2}) &= f^n_i \\ a_i &= f^n_i \end{aligned} \quad \text{if } u^n_{i+1/2} \geq 0$$

$$\begin{aligned} F_i(x_{i-1/2}) &= f^n_{i+1} \\ a_i &= f^n_{i+1} \end{aligned} \quad \text{if } u^n_{i+1/2} < 0 \tag{11}$$

To calculate the slope term of the interpolation function, b_i, we need to define a slope-limiter to assume a good relation between the interpolation function with the original function as the behavior of its monotonicity. A good choice of slope is given by minmod [4], as follows

$$\min \mathrm{mod} = b_i^n = \begin{cases} f_i - f_{i-1} & \text{if } (\delta f_{i-1/2} - \delta f_{i-3/2}) > 0 \\ f_{i-1} - f_{i-2} & \text{if } (\delta f_{i-1/2} - \delta f_{i-3/2}) > 0 \end{cases} . \tag{12}$$

were $\delta f_i = (f_i - f_{i-1})$.

Using (5) we can calculate the flux formulation, as follows

Monotonic case

$$g^n_{i+1/2} = \begin{cases} u^n_{i+1/2}\rho^n_i + (1 - CFL)u^n_{i+1/2}\Delta x_i b^n_i & \text{if } u^n_{i+1/2} \geq 0 \\ u^n_{i+1/2}\rho^n_{i+1} + (1 - CFL)u^n_{i+1/2}\Delta x_i b^n_i & \text{if } u^n_{i+1/2} < 0 \end{cases}$$

otherwise

$$g^n_{i+1/2} = \rho^n_i \quad . \tag{13}$$

where CFL is the Courant number, needed for the stability of the method, and it is calculated by

$$CFL = \left| \frac{u^n_i \Delta t}{\Delta x_i} \right| \tag{14}$$

We need three values $(\rho_{i-2}, \rho_{i-1}, \rho_i)$ to know the behavior of function's monotonicity.

2.3 Piecewise Parabolic Method (PPM)

The Piecewise Parabolic Method use a parabolic function, is expressed in a generic mesh element with boundaries $x_{i-1/2}$, and $x_{i+1/2}$, and considering the velocity u<0 as

$$F_i(x) = a_i + b_i \overline{X}_i + c_i \overline{X}_i (1 - \overline{X}_i) \quad \text{for } x \in \left[x_{i-1/2}, x_{i+1/2} \right] . \tag{15}$$

where

$$\overline{X}_i = \frac{1}{\Delta x_i}(x - x_{i-1/2}) .$$

a_i, bi and c_i are the coefficients of the interpolation function which are obtained by using the constraint conditions, where the constraint conditions are given in each cell by

$$
\begin{aligned}
&F_i(x_{i-1/2}) = f^n_{i-1/2} \\
&F_i(x_{i+1/2}) = f^n_{i+1/2} \\
&\frac{1}{\Delta x_i}\int_{x_{i-1/2}}^{x_{i+1/2}} F_i(x)dx = \rho^n_i
\end{aligned}
\quad . \tag{16}
$$

where $\Delta x_i = x_{i+1/2} - x_{i-1/2}$.

Then the coefficients a_i, b_i and c_i are given by

$$
\begin{aligned}
&a_i = F_i(x_{i-1/2}) \\
&b_i = F_i(x_{i+1/2}) - F_i(x_{i-1/2}) \\
&c_i = 6\left(f_i - \frac{1}{2}(F_i(x_{i-1/2}) + F_i(x_{i+1/2}))\right)
\end{aligned}
\quad . \tag{17}
$$

It is well-known that any high order interpolation tends to create spurious oscillations in numerical solutions. As a remedy for this, slope modifications were introduced in the PPM schemes. In this paper, we adopted the method of Colella and Woodward [4] for computing the interface values as,

$$
f^n_{i+1/2} = \frac{1}{2}(\rho^n_i + \rho^n_{i+1}) - \frac{1}{6}(\overline{\delta} f^n_i - \overline{\delta} f^n_{i-1}) \quad . \tag{18}
$$

with $\overline{\delta} f^n_i$ being the average slope in cell $\left[x_{i-1/2}, x_{i+1/2}\right]$ as follows

$$
\overline{\delta} f_i = \begin{cases} 2\cdot\min(\delta f_i, \alpha_1(\rho_{i+1} - \rho_i), \alpha_2(\rho_i + \rho_{i-1})), & \text{if } (\rho_{i+1} - \rho_i) > 0 \\ 2\cdot\max(\delta f_i, \alpha_1(\rho_{i+1} - \rho_i), \alpha_2(\rho_i + \rho_{i-1})), & \text{if } (\rho_{i+1} - \rho_i) < 0 \end{cases} \quad . \tag{19}
$$

were $\delta f_i = (\rho_{i+1} - \rho_{i-1})/4$. The positives α_1 and α_2 are parameters that control the average slope and affect the dispersion errors of the numerical solutions. For the piecewise parabolic method used in Chimère $\alpha_1 = \alpha_2 = 1$.

To insurance good properties for the reconstruction we must to insure the parabolic monotonicity built on each cell. Then a second corrector algorithm is used to assume the monotonicity.

The flux calculation is given by

$$
g^n_{i+1/2} = \begin{cases} u^n_{i+1/2}\left(F_i(x_{i+1/2}) - \frac{1}{2}CFL\left(\Delta a_i - \left(1 - \frac{2}{3}\frac{u^n_{i+1/2}\Delta t}{\Delta x_i}\right)c_i\right)\right), & if\ u^n_{i+1/2} \geq 0 \\ u^n_{i+1/2}\left(F_{i+1}(x_{i+1/2}) - \frac{1}{2}CFL\left(\Delta a_{i+1} + \left(1 + \frac{2}{3}\frac{u^n_{i+1/2}\Delta t}{\Delta x_{i+1}}\right)c_{i+1}\right)\right), & if\ u^n_{i+1/2} < 0 \end{cases} . \tag{20}
$$

where $\mathrm{CFL} = \left| \frac{u^n_{i+1/2}\Delta t}{\Delta x_i} \right|$

3 Chemical Model Description

The numerical method for the temporal solution of the stiff system of partial differential equations which defines the chemical reactions in the atmosphere is adapted from the second-order TWO-STEP algorithm originally proposed by [6] for gas phase chemistry only. It is based on the application of a Gauss-Seidel iteration scheme to the 2-step implicit backward differentiation (BDF2) formula

$$f^{n+1} = \frac{4}{3} f^{n} - \frac{1}{3} f^{n-1} + \frac{2}{3} \Delta t R(f^{n+1}) \ . \tag{21}$$

with f^n being the vector of chemical concentrations at time t^n, Δt the time step leading from time t^n to t^{n+1} and $R_{(f)} = P_{(f)} - L_{(f)}$ the temporal evolution of the concentrations due to chemical production and emissions (P) and chemical loss and deposition (L). Note that L is a diagonal matrix here. After rearranging and introducing the production and loss terms this equation reads

$$f^{n+1} = \left(I + \frac{2}{3} \Delta t L(f^{n+1}) \right)^{-1} \left(\frac{4}{3} f^{n} - \frac{1}{3} f^{n-1} + \frac{2}{3} \Delta t P(f^{n+1}) \right) . \tag{22}$$

The implicit nonlinear system obtained in this scheme can be solved pertinently with a Gauss-Seidel method [6]. The obtained values of L and P due to the chemical reactions are added in equation (1). A more complete description and evaluation of the CHIMERE model for seasonal simulations and real time forecasts without the use of super-computers, can be found in [7].

4 Numerical Results

Simulations were carried out using the regional V2008 version of the CHIMERE model for 2007. This version calculates the concentration of 44 gaseous species and both inorganic and organic aerosols of primary and secondary origin, including primary particulate matter, mineral dust, sulfate, nitrate, ammonium, secondary organic species and water. The effect of the different numerical resolution scheme on model estimates was analyzed for a domain centred on the Iberian Peninsula in Spain (SP in Figure 2). This simulation scheme was to that used in [9]; A finer domain at a horizontal resolution of 0.2 degree and , covering the Iberian Peninsula was nested to a coarser European scale domain (EUR in Figure 2), ranging from 10.5W to 22.5E and from 35N to 57.5 N and a 0.5 degree horizontal resolution. A one-way nesting procedure was used: coarse-grid simulations forced the fine-grid ones at the boundaries without feedback.

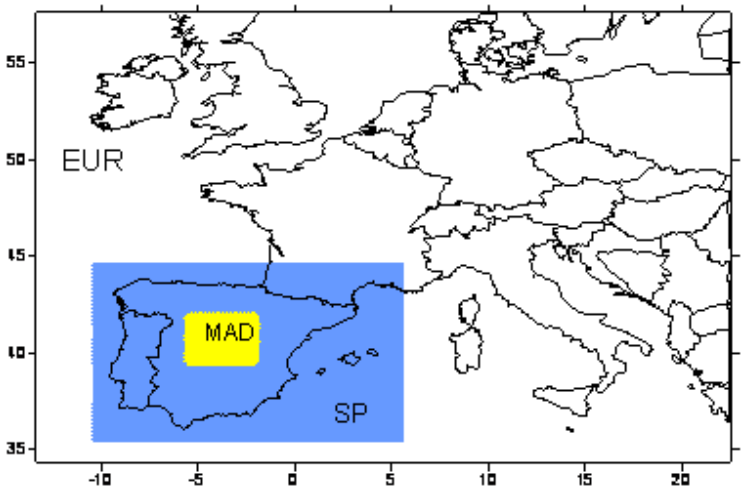


Fig. 1. Location of the two domains simulated with the CHIMERE model

Boundary conditions for the coarsest domain were provided from monthly 2006 climatology from LMDz-INCA model [10] for gases concentrations and from monthly 2004 GOCART model [11] for particulate species.

Accurate emissions data in each time step are not available, then emissions for all the simulations were derived from the annual totals of the EMEP database for 2007 [12]. Original EMEP emissions were spatially disaggregated taking into account land use information (Global Land Cover Facility, GLCF, http://change.gsfc.nasa.gov/create.html) in order to get higher resolution emission data. For each SNAP activity sector, the total NMVOC emission was split into emissions of 227 real individual NMVOC, according to the AEAT speciation [13]. These species were then aggregated into the CHIMERE model ones.

The MM5 model was used to obtain the meteorological input fields. The simulations were carried out also for two domains, with respective horizontal resolutions of 36 Km and 19 Km. Both MM5 simulations were forced by the National Centres for Environmental Prediction model (GFS) analyses.

The CHIMERE model was used with a time step of 3.3 minutes for the three methods.

The quality of model predictions obtained with the three algorithms for the transport module was analyzed by comparing model results to observations at the monitoring sites. Figure 3 shows the location of the NO_2, NO_x SO_2, O_3, and PM_{10} monitoring stations located inside the domain. Figures 5, 6, 7 and 8 present NO_x, O_3, SO_2, and PM_{10} time series showing the concentration obtained with the three methods and the values registered at one of the monitoring sites (01016001 station, Figure 4) for the period between June, 30th and August, 3th in 2007, as an example of the model performance.

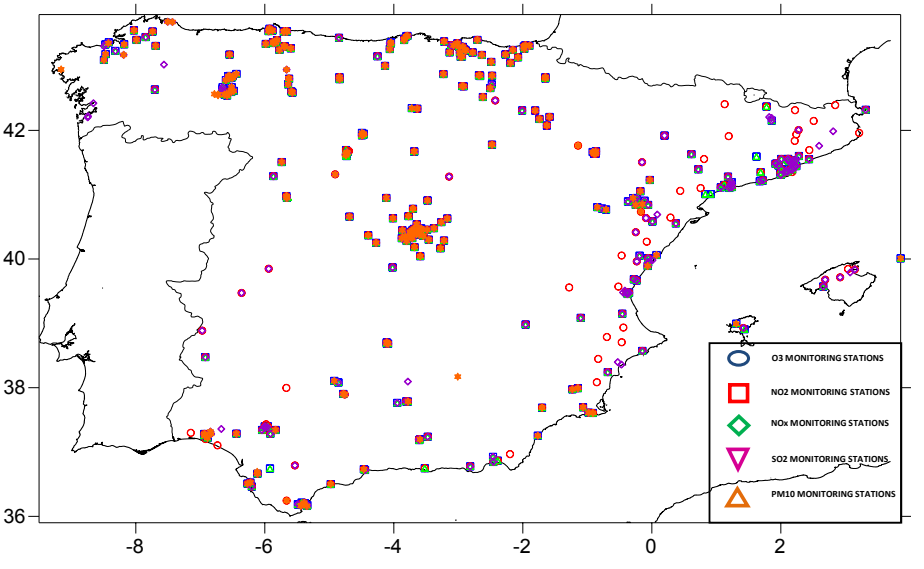


Fig. 2. Location of the monitoring sites recording NO_2, NO_x, SO_2,PM_{10} and O_3 in Spain in 2007

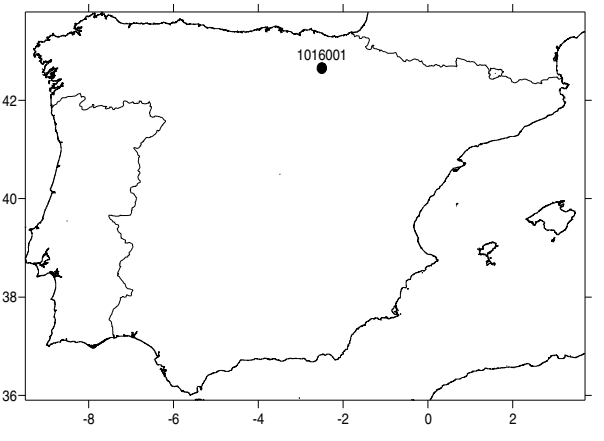


Fig. 3. Location of the 01016001 monitoring station

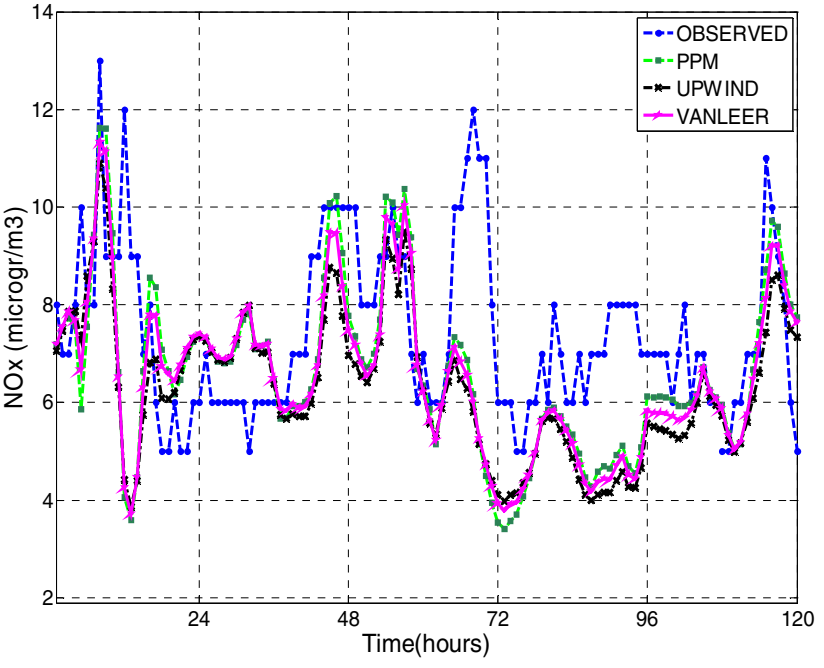


Fig. 4. Observed and simulated NO_x concentration at 01016001 station

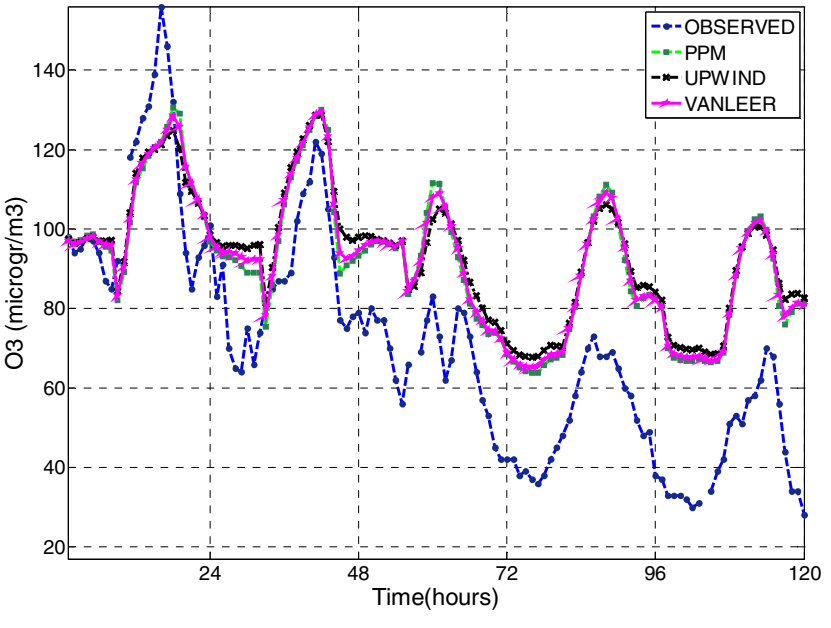


Fig. 5. Observed and simulated O_3 concentration at 01016001 station

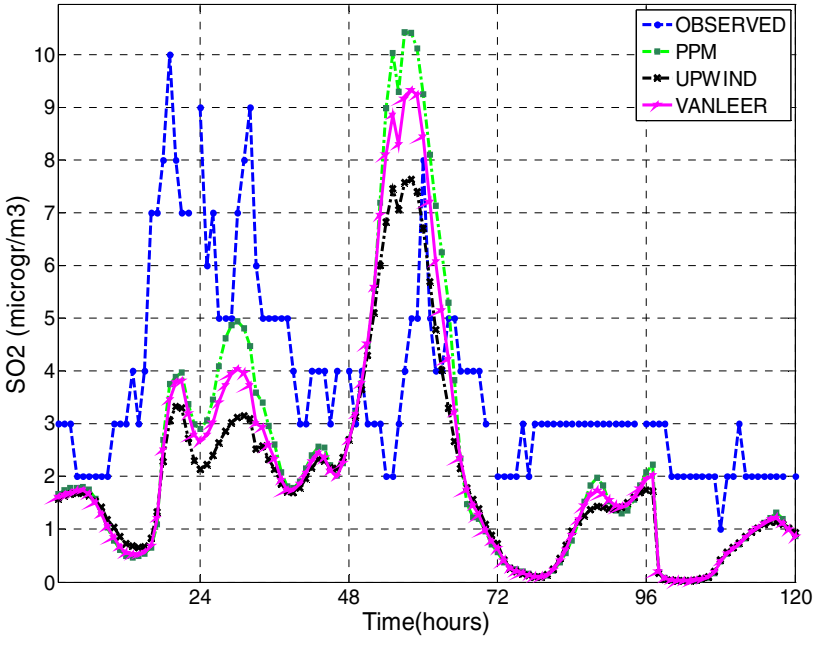


Fig. 6. Observed and simulated SO_2 concentration at 01016001 station

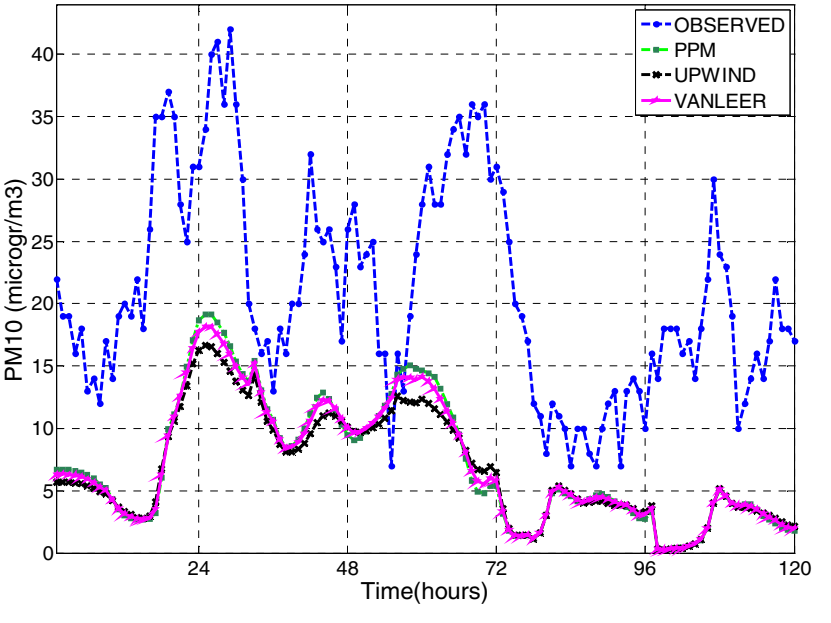


Fig. 7. Observed and simulated PM_{10} concentration at 01016001 station

Table 1. Definition of the metrics used in the evaluation of the CHIMERE model performance

Mean bias	$B_{MB} = \frac{1}{N}\sum(M_i - O_i) = \overline{M} - \overline{O}$
Mean normalized bias	$B_{MNB} = \frac{1}{N}\sum\left(\frac{M_i - O_i}{O_i}\right) = \left(\frac{1}{N}\sum\frac{M_i}{O_i} - 1\right)$
Mean normalized absolute error	$E_{MNAE} = \frac{1}{N}\sum\left(\frac{\lvert M_i - O_i\rvert}{O_i}\right)$
Root mean square error	$E_{RMSE} = \left[\frac{1}{N}\sum(M_i - O_i)^2\right]^{\frac{1}{2}}$
Root mean normalized square error	$E_{RMNSE} = \left[\frac{1}{N}\sum(\frac{M_i - O_i}{O_i})^2\right]^{\frac{1}{2}}$

N: pairs of modeled and observed concentrations *Mi* and *Oi.* The index *i* is over time series and over all the locations in the domain. * $\overline{M} = \frac{1}{N}\sum M_i \overline{O} = \frac{1}{N}\sum O_i$.

Table 2. Mean Concentration for observed and simulated pollutant with the three different conservative methods

	NO_2				SO_2				O_3			
Station	Observed	Upwind	Vanleer	PPM	Observed	Upwind	Vanleer	PPM	Observed	Upwind	Vanleer	PPM
1016001	4.7866	6.2246	6.3935	6.489	3.1098	1.1537	1.2398	1.31	73.5554	88.5497	87.6326	87.0642
1055001	9.4117	6.5169	6.6793	6.774	3.6507	1.371	1.4688	1.5237	69.6142	84.6495	83.8801	83.3886
6016999	3.3103	6.6538	6.7025	6.737	1.4735	0.5893	0.5912	0.6007	74.0728	88.1028	87.8342	87.7665
12032001	10.9689	11.1228	11.3921	11.65	3.8059	2.4019	2.4747	2.5252	31.9679	99.0982	98.4951	97.8441
12099001	8.7443	8.0439	8.249	8.381	4.9698	2.5303	2.899	3.1748	67.9787	93.1618	92.5146	91.9271
12127002	6.8648	5.5356	5.6136	5.673	91.2196	98.3157	98.1126	97.9295	7.8644	3.6566	3.9096	4.0000
15070006	2.5303	5.3069	5.3401	5.368	3.1528	5.1946	4.877	4.5892	36.1126	68.8811	67.7442	67.3601
27033001	2.7271	5.0501	5.0266	5.007	13.0682	17.282	16.7511	16.3985	55.675	69.4079	68.559	68.3416
27058999	3.2231	5.3446	5.3904	5.402	4.0529	4.5851	4.5566	4.4867	56.9156	72.872	72.0308	71.692
27065004	3.1986	5.3884	5.3868	5.383	11.3763	15.5587	15.6326	15.6811	44.8347	70.2854	69.3949	69.0793
31107001	3.01	6.0139	6.198	6.291	3.9642	0.6581	0.7146	0.7553	85.9682	88.8763	87.7539	87.1223
31232002	8.4504	6.2111	6.4245	6.555	3.2836	0.5689	0.6258	0.6717	69.4824	87.1073	86.0155	85.3741
33031029	20.3268	10.3993	10.5258	10.65	10.4492	8.2066	8.347	8.4698	32.6792	73.6432	72.5261	71.7362
43005002	5.4196	10.0196	10.0342	10.06	8.3955	2.4806	2.3846	2.3124	67.7855	92.1313	91.7707	91.4043
44051002	1.2951	6.5892	6.6021	6.635	1.9179	7.482	7.4566	7.5252	83.2927	92.8247	92.2461	91.9131
44118001	4.7056	6.339	6.3015	6.292	8.9103	6.4856	6.6358	6.721	79.383	90.3541	89.9555	89.701

In order to evaluate the performance of the CHIMERE model estimates using the three different models some statistics were calculated. Table 2 presents the metrics used and their definition. Parameters such as mean bias (BMB), mean normalized bias (BMNB), mean normalized absolute error (EMNAE), root mean square error (ERMSE) and root mean normalized square error (ERMNSE) were estimated for NO_2, NO_x, SO_2, O_3 and PM_{10}. Regarding ozone, only statistics for moderate-to-high

ozone concentration cases (more important for human health protection) were considered by selecting predicted-observed value pairs when hourly observations were equal to or greater than the cutoff of 80 μgm^{-3}. For NO_2, NO_x, SO_2 and PM_{10} a cutoff value of 10μgm^{-3} was used. It was taken into account 73, 76, 72, 86 and 54 air quality sites to estimate the statistics of O_3, NO_x, SO_2 and PM_{10} respectively.

In table 3 we show the mean observed and simulated concentration of some pollutants at 16 monitoring stations. Statistical results for all the pollutants are presented in Table 4. Mean normalized bias and mean normalized absolute error for ozone present values inside the range proposed by Tesche et al.[14] to decide on the suitability of a model.

Table 3. Statistics for all the pollutants evaluation for July 2007

O_3	**BMB**	**BMNB**	**EMNAE**	**ERMSE**	**ERMNSE**
Uwind	-1.7035	-0.0052	12.90%	16.9665	16.54%
Van Leer	-1.8389	-0.0069	12.99%	17.0798	16.66%
PPM	-2.0593	-0.0093	13.06%	17.187	16.74%
NO_2					
Uwind	-10.0749	-0.426	61.10%	17.5791	74.83%
Van Leer	-10.0465	-0.4237	61.35%	17.6013	75.44%
PPM	-10.0409	-0.4228	61.53%	17.596	75.79%
NO_x					
Uwind	-14.7716	-0.4768	66.81%	27.7608	81.24%
Van Leer	-14.739	-0.4747	67.04%	27.8095	82.14%
PPM	-14.7318	-0.4741	67.19%	27.828	82.66%
SO_2					
Uwind	-27.5862	-0.6279	76.51%	57.4427	85.92%
Van Leer	-27.5404	-0.6243	77.07%	57.5217	87.10%
PPM	-27.5213	-0.6225	77.68%	57.6022	88.33%
PM_{10}					
Uwind	-21.1558	-0.7469	74.88%	26.9477	76.88%
Van Leer	-21.1623	-0.748	75.03%	26.9582	77.11%
PPM	-21.1913	-0.7498	75.24%	26.9839	77.37%

5 Conclusions

In this paper we have compared the concentration of some pollutants predicted by the CHIMERE Eulerian chemistry transport model using three different conservative methods to solve the advection equation.

The advantage of these methods is that the cell-integrated average is predicted via a flux formulation, thus the mass is exactly conserved. The simulated concentrations for all the pollutants have been compared with a set of observation registered at some monitoring sites in Spain. There are some EPA guidelines to evaluate the accuracy of ozone model predictions (Tesche et al.[14]). The mean normalized absolute error, included in these guidelines, present, for ozone, values inside the ranges proposed inside the suggested EPA range (30-35%), (see Table 4), in order to consider an

acceptable model performance. For the other pollutants, errors present higher values, as it commonly found when evaluating air quality model performance with EMEP database. The disagreement between model and observations for these pollutants is more related to accuracy of the input information, such as emissions, meteorology or land use data. For the time increment used the three advection solvers give similar results, using the 19 km resolution.

If the three advection solvers give similar results, it can be due to the very smooth functions calculated in the emission temporal disaggregation. As conclusion it appears that the type of algorithm used for the advection problem is not so determninant. As it is important to decrease the execution time then, it is sufficient to use the upwind method which is the faster one.

Acknowledgements

This study was supported by Ministerio de Ciencia y Tecnología of Spain under the project CGL2008-1757/CLI.

References

1. Godunov, S.K.: A difference scheme for numerical computation of discontinuous solutions of hydrodynamics equations. Math. Sbornik 47, 271–306 (1959) (in Russian)
2. van Leer, B.: Toward the ultimate conservative difference scheme. Part IV: A new approach to numerical convection. J. Comput. Phys. 23, 276–299 (1997)
3. van Leer, B.: Toward the ultimate conservative difference scheme. Part V: A second order sequel to Godunov's method. J. Comput. Phys. 23, 276–299 (1997)
4. Collela, P., Woodward, P.R.: The piecewise parabolic method (PPM) for gas-dynamical simulations. J. Comput. Phys. 54, 174–201 (1984)
5. Gavete, L., García Vivanco, M., Molina, P., Gavete, M.L., Ureña, F., Benito, J.J.: Implementation in Chimère of a conservative solver for the advection equation. In: Proceedings of the 10th International Conference on Mathematical Methods in Science and Engineering, vol. IV, pp. 1094–1105. CMMSE, Almería (2010), ISBN 13: 978-84-613-5510-5
6. Le Veque, R.J.: Numerical Methods for Conservation Laws. Birkuser Verlag (1990)
7. Verwer, J.G.: A Gauss-Seidel iteration for stiff ODEs from chemical kinetics. SIAM J. Scientific Computing 15, 1243–1250 (1994)
8. Schmidt, H., Derognat, C., Vautard, R., Beekmann, M.: A comparision of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. Atmospheric Environment 35, 6277–6297 (2001)
9. Vivanco, M.G., Palomino, I., Vautard, R., Bessagnet, B., Martín, F., Menut, L., Jiménez, S.: Multi-year assessment of photochemical air quality simulation over SPAIN. Environmental Modelling & Software 24, 63–73 (2009)
10. Hauglustaine, D.A., Hourdin, F., Jourdain, L., Filiberti, M.A., Walters, S., Lamarque, J.F., Holland, E.A.: Interactive chemistry in the Laboratoire de Meteorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation. J. Geophys. Res. 109 (2004), doi:10.1029/2003JD003957

11. Chin, M., Ginoux, P., Kinne, S., Holben, B.N., Duncan, B.N., Martin, R.V., Logan, J.A., Higurashi, A., Nakajima, T.: Tropospheric aerosol optical thickness fromt he GOCART model and comparisons with satellite and sunphotometer measurements. J. Atmos. Sci. 59, 461–483 (2002)
12. Vestreng, V., Breivik, K., Adams, M., Wagener, A., Goodwin, J., Rozovskkaya, O., Pacyna, J.M.: Inventory Review 2005, Emission Data reported to LRTAP Convention and NEC Directive, Initial review of HMs and POPs. Technical report MSC-W 1/(2005), ISSN 0804-2446
13. Passant, N.R.: Speciation of UK Emissions of Non-methane Volatile Organic Compounds. AEAT/ENV/R/0545 (1) (2000)
14. Tesche, T.W., Georgopoulos, P., Seinfeld, J.H., Cass, G., Lurmann, F.W., Roth, P.M.: Improvement of Procedures for Evaluating Photochemical Models. Draft Final Report Prepared for California Air Resources Board. Radian, Sacramento (1990)
15. Vivanco, M.G., Correa, M., Azula, O., Palomino, I., Martín, F.: Influence of Model Resolution on Ozone Predictions over Madrid Area (Spain). In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2008, Part I. LNCS, vol. 5072, pp. 165–178. Springer, Heidelberg (2008)

Chinese Chess Recognition Based on Log-Polar Transform and FFT

Shi Lei[1], Pan Hailang[2], Cao Guo[3], and Li Chengrong[1]

[1] Institute of Automation of Chinese Academy of Sciences, Beijing, 100190 China
[2] School of Electronic and Optical Engineering of Nangjing University of Science And Technology, Nangjing, 210094, China
[3] School of Computer Science and Technology of Nangjing University of Science And Technology, Nangjing, 210094, China

Abstract. In this paper, we present a new method for recognizing Chinese chess. Firstly, histogram equalization and Hough gradient transform is used to determining the location of chess center; Secondly, Log-Polar Transform is used to achieve rotation-invariant and scale-invariant after preprocessing; Thirdly, Fast Fourier Transform (FFT) can achieve translation-invariant. A simple measurement is used to assess the similarity after horizontal and vertical projection. Experiment results have shown that the algorithm can effectively recognize all chesses.

Keywords: Chess Recognition;Log-Polar Transform; Fast Fourier Transform.

1 Introduction

Chinese chess (Xiang Qi or elephant chess) is one of the most popular board games worldwide, being played by approximately one billion people in China, Thailand, Singapore, Vietnam and other Asian countries. At present, many researchers are already attracted by Chinese chess robot. In this project, Chinese character recognition is the key to success. In other words, we need to recognize Chinese character in a chess piece from the image acquired by camera in natural lighting environment.

In 2007, Du[1,2] solved the binarization problem aided by the difference threshold of neighbor pixels; then applied circle template to locating chess. In [3,4], prior information and adaptive thresholding are used to acquire binary image, and then Hough transform can determine the location of chess. In these papers, chessboard and all chesses must be in the right location, such as non-rotation and non-skewness. Under variable illumination, it is difficultly to acquire ideal binary image in this case. About Chinese character recognition, Du[2] presented connected areas statistic, annual ring statistic and template matching. Zhang[3] applied pixels numbers statistic and BP neural network to chess recognition; Zhao[4] added SVM to Zhang[3]' method. All these methods will be unstable under light reflection and unsuccessful autofocus.

In our project, we applied Hough gradient transform [5,6] to determine the location of chess after Histogram Equalization. The next step, Log-Polar and a fast Fourier

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 50–58, 2011.

transform (FFT) is used to achieve rotation-invariant, scale-invariant and translation-invariant. Finally, template matching based on the normalized projection can recognize Chinese chess characters.

2 Preprocessing

After histogram equalization, we applied Hough gradient transform [5,6] to determine the chess location(x,y) and chess radius r as presented in Fig.1. According to H. K. Yuen etc.[7], the Hough gradient transform (also called as 21HT method) is more robustness than the Fast Hough Transform (FHT)[8], the Standard Hough Transform (SHT)[9] and Gerig and Klein Hough Transform (GKHT)[10].

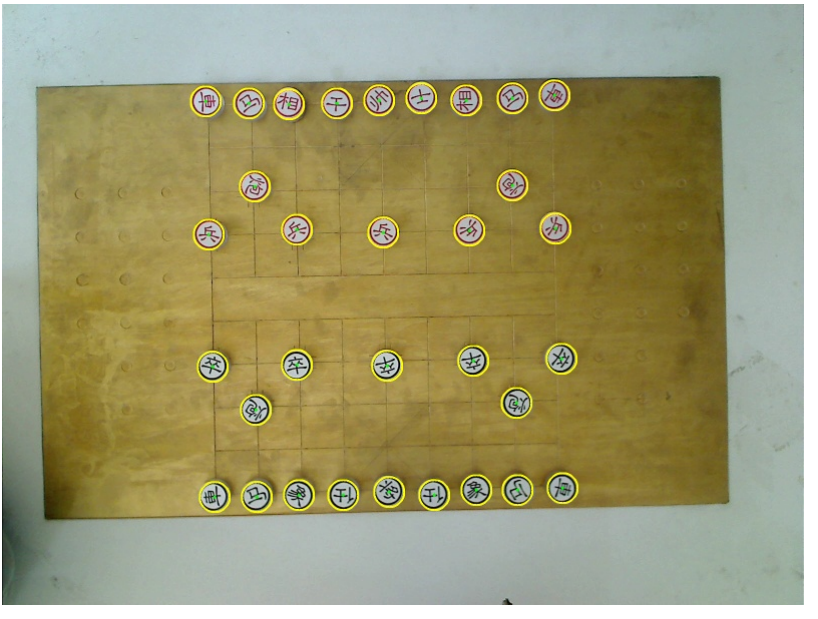


Fig. 1. The results of Hough gradient transform

Three simple steps that can assist us in preprocessing course can be found here. First step: we set up a new（2r+1）×（2r+1） image that its center is (x,y), and pixel intensity will be (255,255,255) if its distance between itself and (x,y) is over r. Second step: we acquire binary image by Otsu'Thresholding[11], then remove the circle by Random Sample Consensus(RANSAC)[12]. Third step: we compute the chess centroid (xo,yo) and maximal radius rmax; then set up a new circular image, its radius is the maximal radius rmax and its center is centroid (xo,yo). At last, the circular image is used as Log-Polar transform input. Fig.2 illustrates the process.

Fig. 2. Preprocessing

3 Feature Extraction

Chess feature extraction consists of three steps. First, Log-Polar Transform is used to provide rotation-invariant and scale- invariant. Second, FFT can provide translation-invariant. Third, feature vector is created by horizontal and vertical projection .

3.1 Log-Polar Transform (LPT)

For image processing applications, LPT is a nonlinear and nonuniform sampling method used to convert image from the Cartesian coordinates I(x, y) to the Log-Polar coordinates ILP(ρ, θ). The mathematical expression of the mapping procedure is shown below

$$\rho = \log \sqrt{(x - x_C)^2 + (y - y_C)^2} \tag{1}$$

$$\theta = \tan^{-1} \frac{y - y_C}{x - x_C} \tag{2}$$

Where (x_c, y_c) is the center pixel of the transformation in the Cartesian coordinates. (x, y) denotes the sampling pixel in the Cartesian coordinates and (ρ, θ) denotes the log-radius and the angular position in the log-polar coordinates. For the sake of simplicity, we assume the natural logarithmic is used in this paper. Sampling is achieved by mapping image pixels in the Cartesian to the log-polar coordinates according to Equation. 1 and 2.

Assume the $I(x, y)$ is magnified in scale k based on the center (x_c, y_c), we have

$$\begin{aligned} \rho' &= \log k \sqrt{(x - x_C)^2 + (y - y_C)^2} \\ &= \log k + \log \sqrt{(x - x_C)^2 + (y - y_C)^2} = \log k + \rho \end{aligned} \tag{3}$$

Assume the $I(x, y)$ is rotated in angle θ_L based on the center (x_c, y_c), we have

$$\theta' = \theta + \theta_L \tag{4}$$

According to Equation.1 and Equation.2, scale-variation and rotation-variation will be transformed to translation-variation.

3.2 Fast Fourier Transform (FFT)

If $I_2(x, y)$ is translated and rotated replica of $I_1(x, y)$ with translation (x_o, y_o) and rotation θo , then

$$I_2(x, y) = I_1(x\cos\theta_O + y\sin\theta_O - x_O, -x\sin\theta_O + y\cos\theta_O - y_O) \tag{5}$$

According to the Fourier translation property and the Fourier rotation property, transforms of I_1 and I_2 are related by

$$\begin{aligned} F_2(\xi,\eta) = e^{-j2\pi(\xi x_O + \eta y_O)} \\ \times F_1(\xi\cos\theta_O + \eta\sin\theta_O, -\xi\sin\theta_O + \eta\cos\theta_O); \end{aligned} \tag{6}$$

Let M_1 and M_2 be the magnitudes of $F1$ and F_2. Therefore, from Equation.6 we have

$$M_2(\xi,\eta) = M_1(\xi\cos\theta_O + \eta\sin\theta_O, -\xi\sin\theta_O + \eta\cos\theta_O) \tag{7}$$

If we consider the magnitudes of F_1 and F_2, then from Equation.7, it is easy to see that the magnitudes of both the spectra are the same, but one is a rotated replica of the other.

3.3 Projection Matching Method

In the Log-Polar coordinates, the ξ -axis is the horizontal axis and The ψ -axis is the vertical axis; the $P(\xi)$ is the projection of ξ-axis and The $P(\psi)$ is the projection of ψ-axis. $I(\xi,\psi)$ denotes a Log-Polar image, then:

$$\begin{aligned} P(\xi) = \sum_{\psi=1}^{M} I(\xi,\psi) \qquad \xi = 1,2...[k\cdot\log\rho_{\max}] \\ M = [k\cdot 2\pi] \end{aligned} \tag{8}$$

$$\begin{aligned} P(\psi) = \sum_{\xi=1}^{N} I(\xi,\psi) \qquad \xi = 1,2...[k\cdot 2\pi] \\ N = [k\cdot\log\rho_{\max}] \end{aligned} \tag{9}$$

For the sake of illumination compensation, $P(\xi)$ and $P(\psi)$ need be normalized, as followed :

$$P'(\xi) = \frac{P(\xi)}{S(\xi)} - \overline{P(\xi)} \quad \xi = 1,2...[k \cdot \log \rho_{\max}]$$
$$M = [k \cdot 2\pi] \tag{10}$$
$$\overline{P(\xi)} = P(\xi) \Big/ \sum_{\xi=1}^{M} S(\xi)$$

$$P'(\psi) = \frac{P(\psi)}{S(\psi)} - \overline{P(\psi)} \quad \psi = 1,2...[k \cdot 2\pi]$$
$$N = [k \cdot \log \rho_{\max}] \tag{11}$$
$$\overline{P(\psi)} = P(\psi) \Big/ \sum_{\psi=1}^{N} S(\psi)$$

Where $S(\xi)$ and $S(\psi)$ denote the numbers of pixels in ξ-axis and ψ-axis; $P'(\xi)$ and $P'(\psi)$ denote normalized projection.

4 Chinese Chess Recognition

We recognize Chinese chess aided by weighted correlation. In fact, we compute inclination cosines of two kinds of projections, then apply weighted value to acquired similarity.

Consider the following definition of the correlation of two images I_1 and I_2, which is the cosine of the angle θ, between them:

$$\Gamma(v_1, v_2) = \cos\theta = \frac{I_1 \cdot I_2}{\|I_1\| \cdot \|I_2\|}$$
$$= \frac{\sum_{\xi=1}^{N1} (P_1'(\xi) \cdot P_2'(\xi)) + \sum_{\psi=1}^{N2} (P_1'(\psi) \cdot P_2'(\psi))}{\sqrt{(\sum_{\xi=1}^{N1} P_1'^2(\xi) + \sum_{\psi=1}^{N2} P_1'^2(\psi)) \cdot (\sum_{\xi=1}^{N1} P_2'^2(\xi) + \sum_{\psi=1}^{N2} P_2'^2(\psi))}} \tag{12}$$

Where N1 is klogρmax, and N2 is k2π.
Γ has two interesting properties:

1) $0 \le |\Gamma(I_1, I_2)| \le 1$;
2) Invariance to affine changes in illumination.

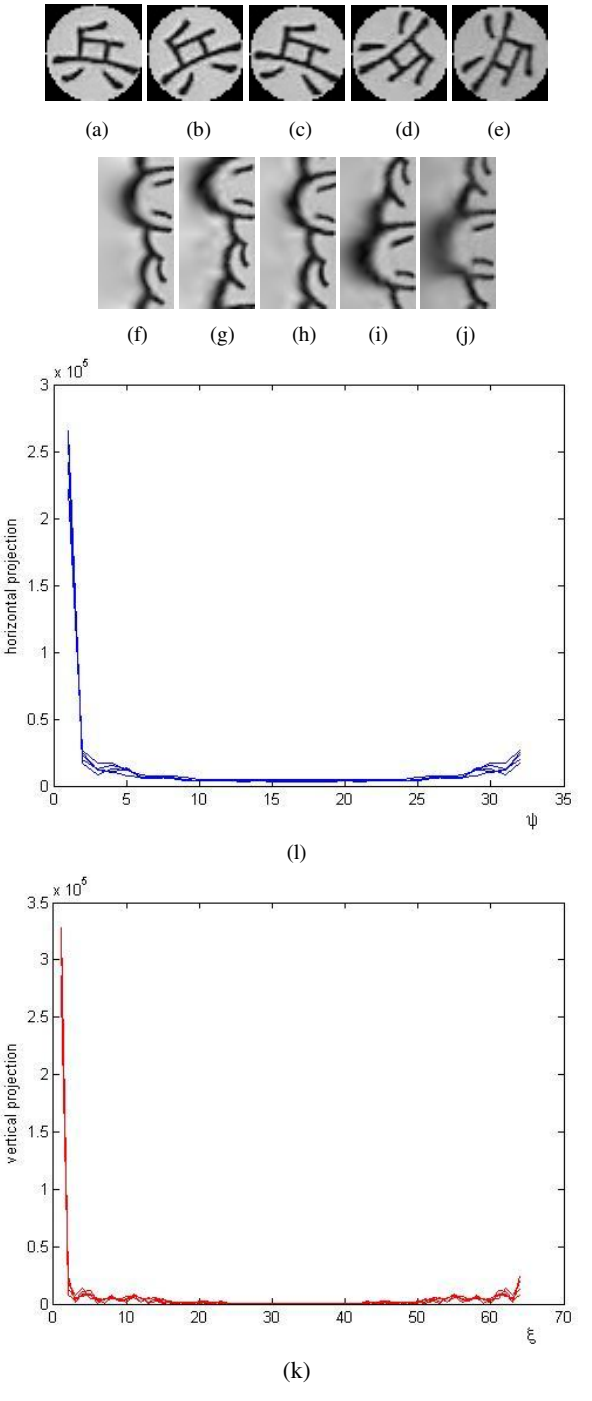


Fig. 3. The feature vector extraction.(a)-(e) denote five "bing" pattern;(f)-(j) denote the Log-Polar transform of five "bing" pattern; (1) and (k) denote horizontal and vertical projection.

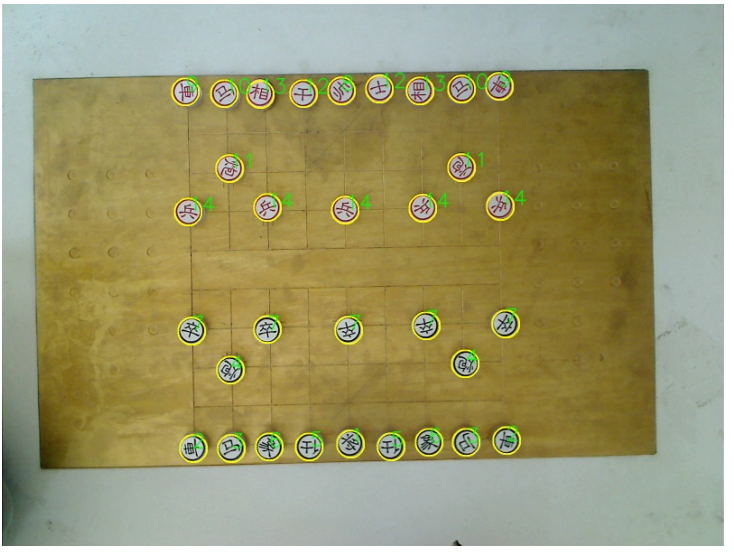


Fig. 4. The final recognition result

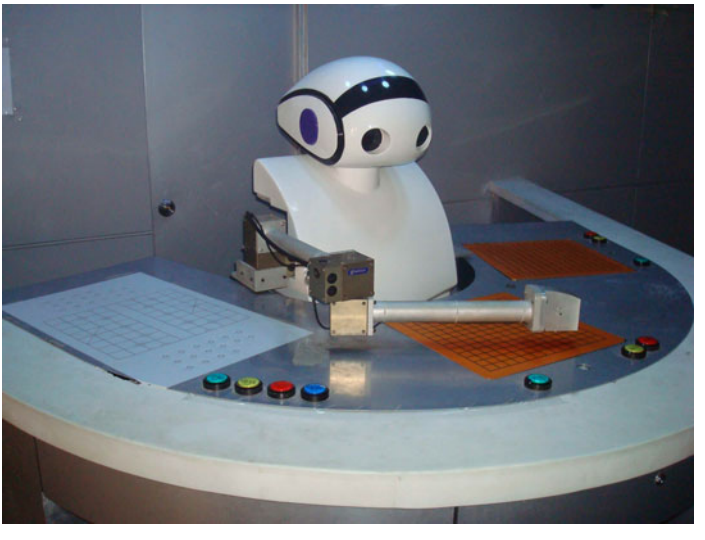


Fig. 5. The Chinese chess robot

5 Some Experimental Results

Some details can be seen in Fig.3. Where (a) - (e) denotes input chess images after preprocessing, respectively; (f) - (j) denotes Log-Polar images of (a) - (c) respectively; (l) - (k) denotes normalized projection after FFT. The size of Image (f)- (j) are 64×32 pixels. As presented in Fig.3, we can acquire the similar results. If we create a database

by the presented method, the database can be used to recognize chess characters. The recognize result can be seen in Fig .4. Several science & technology museums ordered the Chinese chess robot. It is very glad for us that our method is very robust. The Chinese chess robot can be seen in Fig .5.

Of course, there are two drawbacks to the presented method. First, it is assumed that that we have a good illumination; second, the camera must be orthogonal to the board. We will continue to develop the presented method.

6 Conclusion

The Chinese characters recognition in chess pieces is a very key step for Chinese chess robot. The proposed approach addressed in this paper tries to solve illumination invariant, rotation-invariant, scale-invariant and translation-invariant characters recognition. The results of the proposed solution were outstanding. The chess recognition algorithm cycles conducted in VC++, On a PC with Pentium 3.0GHz Processor, and 512MB RAM memory, finishes in less than a second, which may be characterized as a real-time chess playing system.

Acknowledge

The paper is supported by National Natural Science Foundation of China (Grant No. 61075031) and National Natural Science Funds for Distinguished Young Scholar (Grant No.61003108).

References

1. Du, J.-l., Huang, X.-h.: Design of Chinese chess robot vision system. Application Of Electronic Technique 33(9), 133–138 (2007)
2. Du, J.-l., Zhang, J.-f., Huang, X.-h.: Chess-board recognition based on Vision. Computer Engineering And Applications 43(34), 220–224 (2007)
3. Zhiwei, Z., Fanrang, K., Jiwen, Z., Qingbo, H., Zengrong, W.: Image Processing and Recognition in the Vision of Chinese Chess Playing Robot. Computer Applications And Software 25(2), 215–219 (2008)
4. Czajkowski, Zhao, J.-w., Zhang, Z.-w., Xie, F., Liu, Y.-b., Cheng, P.: Image Processing in Vision of Chinese Chess Playing Robot Based on SVM. Journal Of System Simulation 19(18), 4235–4260 (2007)
5. Davies, E.R.: A modified Hough scheme for general circle location. Pattern Recognition Letters 7(1), 37–44 (1988)
6. Illingworth, J., Kittler, J.: The adaptive Hough Transform. IEEE Trans. Pattern Analysis & Machine Intelligence 9(5), 690–697 (1987)
7. Yuen, H.K., Princen, J., Illingworth, J., Kittler, J.: Comparative study of Hough transform methods for circle finding. Image and Vision Computing 8(1), 71–77 (1990)
8. Li, H., Lavin, M.A., LeMaster, R.J.: Fast Hough Transform: A hierarchical approach. In: CVGIP, vol. 36, pp. 139–161 (1986)

9. Smereka, M., Duleba, I.: Circular object detection using a modified Hough transform. Int. J. Appl. Math. Comput. Sci. 18(1), 85–91 (2008)
10. Gerig, G., Klein, F.: Fast contour identification through efficient HoughTransform and simplified interpretation strategy. In: 8th IJCPR, Paris, pp. 498–500 (1986)
11. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. System Man and Cybernetics 9(1), 62–66 (1979)
12. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24, 381–395 (1981)

Adaptive Discontinuous Galerkin B-Splines on Parametric Geometries

Maharavo Randrianarivony

Institute for Numerical Simulation,
Wegelerstrasse 6, Bonn 53115, Germany
randrian@ins.uni-bonn.de,
maharavo@informatik.tu-chemnitz.de

Abstract. We treat numerical simulations on exact geometries by using a discontinuous Galerkin scheme with B-spline bases. That will be realized by the interior penalty form. The initial meshless geometry is constituted of large parametric blocks. We propose an a-posteriori error indicator based on the control points of the B-spline approximations. That enables an efficient adaptive local refinement. Discrete B-splines are used for cascading conjugate gradient iterations. In addition, we provide results on single patches in two and three dimensions.

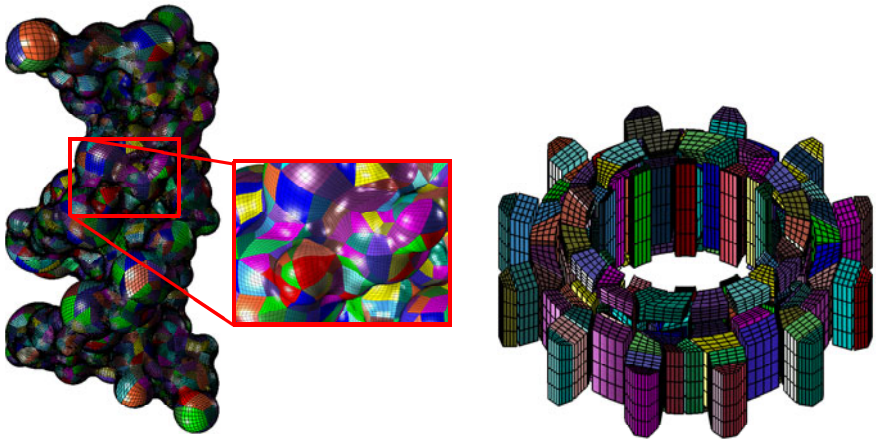
Keywords: B-spline, discontinuous Galerkin, CAD, exact geometry.

1 Introduction

Traditional simulation methods employ fine meshes to represent geometries. Such a representation suffers from a geometric flaw: an unnecessarily large degree of freedom is often required to capture the geometric accuracy. In this paper, we propose a method that can be categorized as a meshless one because we do not consider a fine mesh to represent the initial geometry but rather a set of very coarse parametric spline blocks.

Before presenting our method, a survey of some pertaining works is in order. The initial purpose of Bézier and B-splines entities [3,10] was to design curves and surfaces especially for car bodies and CAD components. But later they found their use in different disciplines such as molecular modelings and statistical data processing. The desire to apply simulations on curved models is not new. In fact, Höllig and Reif have [16] used the WEB spline to approximate the solution to a PDE. Curved elements motivate equally the isogeometric analysis using NURBS (Non-Uniform Rational B-Spline) described in [17]. As for wavelets, the Wavelet-Galerkin method [14,15] is able to produce a good accuracy with low computational cost by means of adaptivity. Harbrecht and Randrianarivony [14,15] have successfully applied Wavelet methods on CAD and molecular models similar to those in Fig. 1(a). As inputs, they accept a CAD file in an IGES format [22] or a molecular model in PDB format. In the domain of CAD preprocessings, some former works are as follows. For transfinite interpolations [8,12], Coons patches usually serve as tools to generate the mappings [10] from parametric

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 59–74, 2011.



(a) (b)

Fig. 1. Parametric decompositions: (a) DNA molecule, (b)Hexahedral decomposition

domains. Brunnett and Randrianarivony have proposed [19] a splitting method for CAD surfaces. They have also invested a lot to implement their methods by using the IGES format [22]. But they did not treat the global continuity of the resulting patches. For molecular surfaces, global continuity can be obtained exactly [14,15,20] because all boundary curves are circular arcs which can be easily parametrized. That is not the case for other CAD curves which need more careful treatments. The main task in [21] is the correlation between the Coons patch which resides in an individual patch and the global continuity.

This paper is organized as follows. We start by formulating the problem in Section 2 which contains also the space of approximation and the discontinuous Galerkin (DG) scheme using B-splines. We formulate there the treatment of the problem on the parameter domain instead of the physical domain. We will see in Section 3 the establishment of an a-posteriori error estimator [1]. That will be deduced from the de Boor-Fix functional. In Section 4, we describe some outcomes of practical adaptive simulation using our DG method.

2 Discontinuous Galerkin Using B-Splines

2.1 Parametric Setting

Our purpose in this document is to solve a Poisson problem having a Dirichlet boundary condition on a multi-dimensional domain $\boldsymbol{\Omega} \subset \mathbb{R}^d$ where $d = 2, 3$. More precisely, we would like to solve the following problem

$$\Delta u(\mathbf{x}) = f(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \boldsymbol{\Omega}, \tag{1}$$

$$u(\mathbf{x}) = g(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \partial\boldsymbol{\Omega} \tag{2}$$

where the domain $\boldsymbol{\Omega}$ is supposed to be constituted of a set of very coarse parametric patches as in Fig. 1(b). That is, we suppose that there are mappings $\mathcal{M}_i$ such that

$$\boldsymbol{\Omega} = \bigcup_{i=1}^{N} \mathcal{M}_i\Big(\prod_{j=1}^{d}[a_{i,j}, b_{i,j}]\Big). \tag{3}$$

For the sake of notational convenience, we suppose that we have only a single patch $\mathcal{M} = \mathcal{M}_1$ as illustrated in Fig. 2(b). Later on, the domain $\boldsymbol{\Omega}$ will be termed *physical domain* while $\mathbf{P} := \prod_{i=1}^{d}[a_i, b_i]$ *parameter domain*. We suppose that $\mathcal{M}$ is invertible, differentiable and that it admits a regularity condition meaning that the Jacobian matrix $D\mathcal{M}$ has nonvanishing determinant everywhere. In addition, we suppose that there is some positive constant μ such that the linear operators $D\mathcal{M}$ and $D\mathcal{M}^{-1}$ have the next bounds

$$\|D\mathcal{M}\|_* := \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\|D\mathcal{M}\mathbf{x}\|}{\|\mathbf{x}\|} \leq \mu, \qquad \text{and} \qquad \|D\mathcal{M}^{-1}\|_* \leq \mu. \tag{4}$$

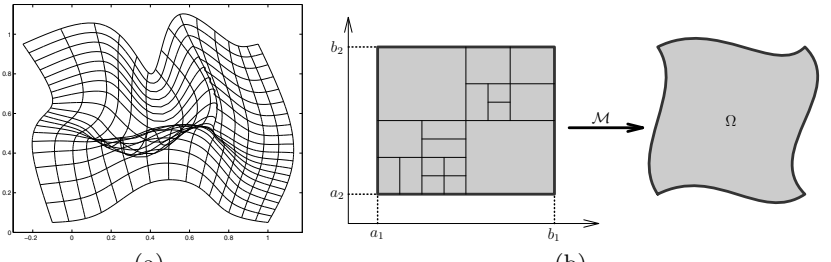
Finally, we suppose that $\mathcal{M}$ is of smoothness $\mathcal{C}^m$ where m is sufficiently large. The mapping $\mathcal{M}$ is in general supposed to be a B-spline or NURBS patch [10]. A counter example of a mapping violating the above condition is observed in Fig. 2(a). An efficient preparation of such mappings is discussed in [18,19,21]. For two integers $k \geq 1$ and $n \geq k-1$, the definition of B-spline basis functions with respect to the knot sequence $\boldsymbol{\zeta} = (\zeta_i)_{i=0}^{n+k}$ uses the divided difference of the truncated power functions $(\cdot - t)_+^k$ given by

$$(x-t)_+^k := \begin{cases} (x-t)^k & \text{if} \quad x \geq t, \\ 0 & \text{if} \quad x < t. \end{cases} \tag{5}$$

More precisely, one has the definition and support property

$$N_i^k(t) = N_i^{n,k,\boldsymbol{\zeta}}(t) := (\zeta_{i+k} - \zeta_i)[\zeta_i, ..., \zeta_{i+k}](\cdot - t)_+^{k-1}, \tag{6}$$

$$\mathrm{Supp}(N_i^k) = [\zeta_i, \zeta_{i+k}]. \tag{7}$$



(a) (b)

Fig. 2. (a)Counter example of a regular mapping $\mathcal{M}$, (b) A discretization $\mathcal{T}_h$ on the parameter domain and a mapping onto the physical domain $\boldsymbol{\Omega}$

To ensure that the B-spline functions are open, we assume that the knot sequence $\boldsymbol{\zeta}$ is *clamped*. That is, the sequence $\zeta_0, ..., \zeta_{n+k}$ is provided as: $\zeta_0 = \cdots = \zeta_{k-1}$ and $\zeta_{n+1} = \cdots = \zeta_{n+k}$. For the multi-variate case ($d = 2, 3$), let us consider d spline properties $(n_1, k_1, \boldsymbol{\zeta}^1)$,..., $(n_d, k_d, \boldsymbol{\zeta}^d)$ and let us define

$$\left(N_{i_1}^{n_1,k_1,\boldsymbol{\zeta}^1} \otimes \cdots \otimes N_{i_d}^{n_d,k_d,\boldsymbol{\zeta}^d}\right)(t_1, ..., t_d) := N_{i_1}^{n_1,k_1,\boldsymbol{\zeta}^1}(t_1) \cdots N_{i_d}^{n_d,k_d,\boldsymbol{\zeta}^d}(t_d). \quad (8)$$

We denote the space of d-dimensional splines by

$$\mathbb{S}[n_1, k_1, \boldsymbol{\zeta}^1; \cdots ; n_d, k_d, \boldsymbol{\zeta}^d] := \operatorname{span}\left\{ N_{i_1}^{n_1,k_1,\boldsymbol{\zeta}^1} \otimes \cdots \otimes N_{i_d}^{n_d,k_d,\boldsymbol{\zeta}^d} \right\}. \quad (9)$$

The efficient de Boor algorithm [3,10] serves as a method to evaluate a B-spline function without using the above divided differences.

In our present method, we do not use a mesh on the physical domain $\boldsymbol{\Omega}$. Instead, we have a non-conforming discretization $\mathcal{T}_h$ on the parameter domain as depicted in Fig. 2(b). More precisely, the discretization $\mathcal{T}_h$ is composed of elements Q_1,...,Q_N satisfying the following conditions.

(C1) Each element is a closed domain $Q_i = \prod_{\nu=1}^{d} [r_{i,\nu}, s_{i,\nu}] \subset \mathbf{P}$.
(C2) For two different indices i, j, we have $\mathring{Q}_i \cap \mathring{Q}_j = \emptyset$ where $\mathring{Q}$ designates the topologic interior of Q.
(C3) We have $\mathbf{P} = \cup_{i=1}^{N} Q_i$.

According to those assumptions, the discretization $\mathcal{T}_h$ is allowed to be non-conforming. For an element $Q := \prod_{\nu=1}^{d} [r_\nu, s_\nu]$, we denote $h_\nu(Q) := s_\nu - r_\nu$ so that the element measure is $h(Q) := \prod_{\nu=1}^{d} h_\nu(Q)$. In addition, we assume the shape regularity condition throughout this document. That is, there is a positive constant ρ independent of Q such that

$$r(Q)/R(Q) \leq \rho \qquad \forall\, Q \in \mathcal{T}_h, \quad (10)$$

where $R(Q)$ is the largest circle for $d = 2$ (resp. sphere for $d = 3$) contained in Q while $r(Q)$ is the smallest circle (resp. sphere) including Q. An internal edge (resp. face) is defined to be a nonempty intersection e of two different elements $Q_i \in \mathcal{T}_h$ and $Q_j \in \mathcal{T}_h$ such that e is not a point. Similarly, a boundary edge (resp. face) is a nonempty intersection of an element $Q_i \in \mathcal{T}_h$ and the boundary $\partial\mathbf{P}$ which is not a point. The sets of internal and boundary edges (resp. faces) are denoted by $\mathcal{J}_h$ and $\mathcal{B}_h$ respectively. We will denote by $\mathcal{E}_h = \mathcal{J}_h \cup \mathcal{B}_h$ the set of all edges (resp. faces). Finally, the length (resp. area) of an edge (resp. face) $e \in \mathcal{E}_h$ is denoted by $h(e)$.

For the mesh $\mathcal{T}_h$ on the parameter domain, the approximating space will be

$$\mathbb{V}_h := \left\{ s \in \mathbb{L}_2(\mathbf{P}) : \quad s_{|Q} \in \mathbb{S}\left[n_1, k_1, \boldsymbol{\zeta}^1; \cdots ; n_d, k_d, \boldsymbol{\zeta}^d\right] \qquad \forall\, Q \in \mathcal{T}_h \right\} \quad (11)$$

where the spline properties depend on Q. That is, $n_i = n_i(Q)$, $k_i = k_i(Q)$ and $\boldsymbol{\zeta}^i = \boldsymbol{\zeta}^i(Q)$ for $i = 1, ..., d$.

2.2 Parametric Discontinuous Galerkin

The values of a function from $\mathbb{V}_h$ generally do not admit continuities at element interfaces. As a consequence, let us introduce the *jump* value $[\![v]\!]$ and *average* value $\{\{v\}\}$. For an internal edge $e \in \mathcal{J}_h$ such that $e = \partial Q_1 \cap \partial Q_2$, let $\mathbf{n}_{Q_1}(\mathbf{x})$ and $\mathbf{n}_{Q_2}(\mathbf{x})$ designate the outward normals at $\mathbf{x} \in e$ with respect to $Q_1 \in \mathcal{T}_h$ and $Q_2 \in \mathcal{T}_h$ respectively. For a scalar valued function v, the jump is defined to be the vector

$$[\![v]\!](\mathbf{x}) := u_{Q_1}(\mathbf{x})\mathbf{n}_{Q_1}(\mathbf{x}) + u_{Q_2}(\mathbf{x})\mathbf{n}_{Q_2}(\mathbf{x}) \qquad \forall\, \mathbf{x} \in e. \tag{12}$$

Additionally, the average is defined as

$$\{\{v\}\}(\mathbf{x}) := 0.5\big(v_{Q_1}(\mathbf{x}) + v_{Q_2}(\mathbf{x})\big) \qquad \forall\, \mathbf{x} \in e. \tag{13}$$

The jump and average for a boundary edge $e \in \mathcal{E}_h$ are defined similarly where the exterior value is assumed to be zero. Later, the unknown function is approximated by a function in $\mathbb{V}_h$ where the jump values are constrained to be zero with the help of some penalty terms.
The broken Sobolev space with respect to the non-conforming discretization $\mathcal{T}_h$ is denoted by

$$\mathbb{H}^k(\mathcal{T}_h) := \big\{w \in \mathbb{L}^2(\Omega) : \quad w_{|_Q} \in \mathbb{H}^k(Q) \quad \forall\, Q \in \mathcal{T}_h\big\}. \tag{14}$$

We will need also

$$|\!|\!|w|\!|\!| := \Big[\sum_{Q\in\mathcal{T}_h} |w|^2_{1,Q} + h(Q)^2|w|^2_{2,Q} + \sum_{e\in\mathcal{E}_h} \frac{1}{h(e)}\Big\|[\![w]\!]\Big\|^2_{0,e}\Big]^{1/2}. \tag{15}$$

By using DG variational formulations [2,4], the initial problem in (1) reduces to seek $u_h \in \mathbb{V}_h$ such that

$$\mathcal{B}(u_h, v_h) = \mathcal{L}(v_h) \qquad \forall\, v_h \in \mathbb{V}_h \tag{16}$$

where $\mathcal{B}$ and $\mathcal{L}$ are as follows when expressed in terms of the parameter domain

$$\begin{aligned}
\mathcal{B}(u,v) := &\sum_{Q\in\mathcal{T}_h} \int_Q (D\mathcal{M}^T)^{-1}\nabla_{\mathbf{t}} u \cdot (D\mathcal{M}^T)^{-1}\nabla_{\mathbf{t}} v \,\det D\mathcal{M}\, d\mathbf{t} \\
&- \sum_{e\in\mathcal{E}_h} \int_e \{\{(D\mathcal{M}^{-1})(\nabla_{\mathbf{t}} u)\}\} \cdot [\![v]\!] \,\det D\mathcal{M}\, d\mathbf{t} \\
- \sum_{e\in\mathcal{E}_h} &\int_e [\![u]\!] \cdot \{\{(D\mathcal{M}^{-1})(\nabla_{\mathbf{t}} v)\}\} \,\det D\mathcal{M}\, d\mathbf{t} + \frac{1}{\eta}\sum_{e\in\mathcal{E}_h}\int_e [\![u]\!]\cdot[\![v]\!] \,\det D\mathcal{M}\, d\mathbf{t}, \\
\mathcal{L}(v) := &\sum_{Q\in\mathcal{T}_h} \int_Q f\circ\mathcal{M} \det D\mathcal{M} d\mathbf{t} - \sum_{e\in\mathcal{B}_h}\int_e g\circ\mathcal{M}\Big[\nabla v\cdot\mathbf{n} + \frac{1}{\eta}v\Big] \det D\mathcal{M} d\mathbf{t}
\end{aligned}$$

in which η is a certain large positive number. The above expressions are similar to the usual DG variational formulation [2,4,5] but we use the Jacobians to

transform them onto the parameter domain. We will need the operator $\mathcal{A}$ which is defined to be such that

$$\mathcal{B}(u,\phi) = \langle \mathcal{A}u, \phi \rangle \qquad \forall\, \phi \in \mathbb{H}^k(\mathcal{T}_h). \tag{17}$$

Note that although it is possible to reformulate the above form $\mathcal{B}$ on the physical domain $\boldsymbol{\Omega}$ by using curved elements, we want to avoid that because it is difficult to apply geometric operations such as refinements to curved entities.

Lemma 1. *Under the above conditions on the mapping $\mathcal{M}$, we have*

$$\frac{1}{\mu} \le \|D\mathcal{M}\|_* \le \mu, \qquad \text{and} \qquad \frac{1}{\mu} \le \|D\mathcal{M}^{-1}\|_* \le \mu. \tag{18}$$

Thus, the determinant verifies

$$0 < C_1 \le \det(D\mathcal{M}) \le C_2. \tag{19}$$

Hence, with respect to $\|\!|\cdot|\!\|$, the bilinear form $\mathcal{B}$ admits coercivity

$$\|\!|f|\!\|^2 \le C_{\mu,\boldsymbol{\Omega}}\, \mathcal{B}(f,f), \tag{20}$$

and boundedness

$$\big|\mathcal{B}(f,g)\big| \le C_{\mu,\boldsymbol{\Omega}} \|\!|f|\!\| \|\!|g|\!\| \tag{21}$$

Proof. Use the property of $\mathcal{M}$, and proceed as in [2,4,5]. □

3 Adaptive Simulation

3.1 Adaptive Refinements and Spline Operators

In an adaptive simulation, we deduce a finer discretization $\mathcal{T}_{h+1}$ from a coarse one $\mathcal{T}_h$ by refining some elements $Q \in \mathcal{T}_h$. We consider two kinds of 2D subdivisions. The first one consists in bisecting the rectangle $Q \in \mathcal{T}_h$ by inserting a vertical cut resulting in two sub-rectangles of the same size as shown in Fig. 3(a). The second one does the same but with a horizontal cut as in Fig. 3(b). Note that those two subdivisions could deteriorate the shape regularity (10). As a consequence, we choose the subdivision such that the aspect ratios of the resulting rectangles do not exceed the threshold ρ. In fact, it is possible to subdivide in both directions but it is a bit more difficult to apply space hierarchies (see Section 3.3) for that. The generalization to 3D is done in a straightforward manner as illustrated in Fig. 3(c),(d),(e).

In the next description, we show a method to establish if an element $Q \in \mathcal{T}_h$ should be split or not. It is anyhow beyond the scope of this paper to determine which kind of subdivision is optimal. In order to know the elements $Q \in \mathcal{T}_h$ which ought to be subdivided, we need an a-posteriori error indicator. For that purpose, let us first introduce some interesting definitions [3,13] about spline

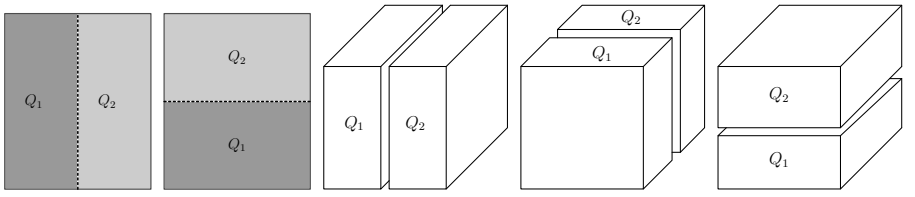


Fig. 3. (a)Vertical 2D-subdivision (b)Horizontal 2D-subdivision (c)3D-subdivision along x-axis (d)3D-subdivision along y-axis (e)3D-subdivision along z-axis

dual functionals where we consider the spline property $(n, k, \boldsymbol{\zeta})$. We define for $i = 0, ..., m$

$$\vartheta_i(t) := \frac{1}{(k-1)!} \prod_{p=1}^{k-1} (t - \zeta_{i+p}). \tag{22}$$

We introduce also the function z_i for $i = 0, ..., m$ as

$$\begin{aligned} z_i(t) &:= 0 \qquad \text{for} \qquad t \le \zeta_i, \\ z_i(t) &:= \vartheta_i(t) \quad \text{for} \qquad t \ge \zeta_{i+k}. \end{aligned} \tag{23}$$

Inside the interval $[\zeta_i, \zeta_{i+k}]$, the function z_i is defined to be a smooth function having $\mathcal{C}^{(k-1)}$-joints at ζ_i and ζ_{i+k}:

$$z_i^{(m)}(\zeta_i) = 0, \qquad z_i^{(m)}(\zeta_{i+k}) = \vartheta_i^{(m)}(\zeta_{i+k}) \qquad \forall\, m = 0, ..., k-1. \tag{24}$$

For that, use for example a higher order Hermite interpolations [11]. For a univariate square integrable function φ in the interval $[\zeta_0, \zeta_{n+k}]$, we define for $i = 0, ..., n$

$$\lambda_i(\varphi) := \int_{\zeta_i}^{\zeta_{i+k}} \varphi(t)\, z_i^{(k)}(t) dt. \tag{25}$$

Under some mild assumption, it can be proved by simple partial integrations that λ_i coincides with the usual de Boor-Fix functional [3,13]

$$\lambda_i(\varphi) = \sum_{m=1}^{k} (-1)^{k-m} \vartheta_i^{(m-1)}(\tau) \varphi^{(k-m)}(\tau) \tag{26}$$

for some $\tau \in [\zeta_i, \zeta_{i+k}]$. In the present context, (25) is more suitable than relation (26) that requires point evaluations of φ which are not appropriate if φ is only square integrable. As for the multivariate case $(d = 2, 3)$, let us fix some $\mathbf{n} = (n_1, ..., n_d)$, $\mathbf{k} = (k_1, ..., k_d)$ and $Q = [\zeta_0^1, \zeta_{n_1+k_1}^1] \times \cdots \times [\zeta_0^d, \zeta_{n_d+k_d}^d]$. For $\mathbf{i} = (i_1, ..., i_d)$ where each i_ν belongs to $\{0, 1, ..., n_\nu\}$ in which $\nu = 1, ..., d$, we define for $\varphi \in L^2(Q)$

$$\boldsymbol{\lambda}_\mathbf{i}(\varphi) := \int_{\zeta_{i_1}}^{\zeta_{i_1+k_1}} \cdots \int_{\zeta_{i_d}}^{\zeta_{i_d+k_d}} \varphi(t_1, ..., t_d)\, z_{i_1}^{(k_1)}(t_1) \cdots z_{i_d}^{(k_d)}(t_d)\, dt_1 \cdots dt_d. \tag{27}$$

Note that for a tensor product function $\varphi(t_1,...,t_d) = \prod_{\nu=1}^{d} \varphi_\nu(t_\nu)$, we have $\boldsymbol{\lambda}_{\mathbf{i}}(\varphi) = \prod_{\nu=1}^{d} \lambda_{i_\nu}(\varphi_\nu)$. The above functional admits B-spline duality: for a tensor product B-spline basis function $N_{\mathbf{i}}^{\mathbf{k}} = N_{i_1}^{k_1} \otimes \cdots \otimes N_{i_d}^{k_d}$ we have $\boldsymbol{\lambda}_{\mathbf{i}}(N_{\mathbf{j}}^{\mathbf{k}}) = \delta_{\mathbf{i},\mathbf{j}}$. Finally, for a square integrable function φ in Q, we define the B-spline quasi-interpolant

$$P(\varphi) := \sum_{\mathbf{i}\in I} \boldsymbol{\lambda}_{\mathbf{i}}(\varphi) N_{i_1}^{k_1} \otimes \cdots \otimes N_{i_d}^{k_d} \tag{28}$$

where the sum is over $I := \{0,...,n_1\} \times \cdots \times \{0,...,n_d\}$.

3.2 A-posteriori Error Indicator

In this section, we discuss about a-posteriori errors where we suppose that the solution u_h of (16) with respect to the current discretization $\mathcal{T}_h$ is available.

Theorem 1. *For each element $Q \in \mathcal{T}_h$ in which we have the B-spline properties $k_i = k_i(Q)$, $n_i = n_i(Q)$, for $i = 1,...,d$, we consider the next error indicator*

$$\varepsilon(Q) := \|f \circ \mathcal{M} - \mathcal{A}u_h\|_{0,Q} \left[\prod_{i=1}^{d} \frac{\sqrt{n_i - k_i + 2}}{\sqrt{h_i(Q)}}\right] \left[\sum_{i=1}^{d} \left(\frac{k_i h_i(Q)}{n_i - k_i + 2}\right)^2\right]^{1+(d/4)}.$$

Under the quasi-uniformity condition

$$\max_{i=k-1,...,n} |\zeta_i - \zeta_{i+1}| \Big/ \min_{i=k-1,...,n} |\zeta_i - \zeta_{i+1}| \le \theta < \infty, \tag{29}$$

we have the following reliability relation

$$|\!|\!| u - u_h |\!|\!| \le c(\theta)\varepsilon(\mathcal{T}_h) = c(\theta)\Big[\sum_{Q\in\mathcal{T}_h} \varepsilon(Q)^2\Big]^{1/2}.$$

Proof. Denote by $P_{\mathbb{V}_h}$ the projection to $\mathbb{V}_h$ such that inside each $Q \in \mathcal{T}_h$, $P_{\mathbb{V}_h}$ is the quasi-interpolant P_Q as defined in (28) with respect to the spline properties $\big(n_1(Q), k_1(Q), \boldsymbol{\zeta}^1(Q)\big)$, ...,$\big(n_d(Q), k_d(Q), \boldsymbol{\zeta}^d(Q)\big)$. From the boundedness (21), the coercivity (20), the operator $\mathcal{A}$ of relation (17) and an orthogonality relation, one obtains

$$\begin{aligned}
|\!|\!| u - u_h |\!|\!| &\le \mathcal{B}\Big(u - u_h, \frac{u - u_h}{|\!|\!| u - u_h |\!|\!|}\Big) \le \sup_{|\!|\!|\phi|\!|\!|=1} \mathcal{B}(u - u_h, \phi) = \\
&= \sup_{|\!|\!|\phi|\!|\!|=1} \mathcal{B}\big(u - u_h, \phi - P_{\mathbb{V}_h}(\phi)\big) \\
&= \sup_{|\!|\!|\phi|\!|\!|=1} \langle \mathcal{A}u - \mathcal{A}u_h, \phi - P_{\mathbb{V}_h}(\phi)\rangle = \sup_{|\!|\!|\phi|\!|\!|=1} \big\langle f \circ \mathcal{M} - \mathcal{A}u_h, \phi - P_{\mathbb{V}_h}(\phi)\big\rangle \\
&= \sup_{|\!|\!|\phi|\!|\!|=1} \sum_{Q\in\mathcal{T}_h} \langle f \circ \mathcal{M} - \mathcal{A}u_h, \phi - P_Q(\phi)\rangle_Q \\
&\le \sum_{Q\in\mathcal{T}_h} \|f \circ \mathcal{M} - \mathcal{A}u_h\|_{0,Q} \sup_{|\!|\!|\phi|\!|\!|=1} \|\phi - P_Q(\phi)\|_{0,Q}.
\end{aligned}$$

Consider now the last supremium within an element $Q \in \mathcal{T}_h$. Consider one spline segment $\Delta(\mathbf{j}) := \prod_{\nu=1}^{d}[\zeta^{\nu}_{j_\nu}, \zeta^{\nu}_{j_\nu+1}] \subset Q$ where $\mathbf{j} = (j_1, ..., j_d)$.Consider also an extension of the spline segment as $\widetilde{\Delta}(\mathbf{j}) := \prod_{\nu=1}^{d}[\zeta^{\nu}_{j_\nu-(k_\nu-1)}, \zeta^{\nu}_{j_\nu+k_\nu}]$. Due to the support property (7), for each $\mathbf{t} = (t_1, ..., t_d) \in \Delta(\mathbf{j})$, the basis functions with nonzero coefficients of $P_Q(\phi)(\mathbf{t})$ from (28) correspond to $\mathbf{i} = (i_1, ..., i_d)$ such that $i_\nu \in \{j_\nu - (k_\nu - 1), ..., j_\nu + k_\nu\}$ for $\nu = 1, ..., d$.

Since the de Boor-Fix quasi-interpolant P_Q keep polynomials unchanged [3], we obtain for every $p \in \mathcal{P}(\widetilde{\Delta}(\mathbf{j}))$ (polynomials on $\widetilde{\Delta}(\mathbf{j})$)

$$\|\phi - P_Q(\phi)\|_{0,\Delta(\mathbf{j})} \leq \sqrt{\mu[\Delta(\mathbf{j})]}\big\|\phi - P_Q(\phi)\big\|_{\infty,\Delta(\mathbf{j})} \tag{30}$$

$$= \sqrt{\mu[\Delta(\mathbf{j})]}\big\|(\phi - p) - P_Q(\phi - p)\big\|_{\infty,\Delta(\mathbf{j})} \tag{31}$$

$$\leq \sqrt{\mu[\Delta(\mathbf{j})]}(1 + \|P_Q\|)\|\phi - p\|_{\infty,\widetilde{\Delta}(\mathbf{j})} \tag{32}$$

($\mu(X)$ standing for the Lebesgue measure of X). Hence,

$$\|\phi - P_Q(\phi)\|_{0,\Delta(\mathbf{j})} \leq \sqrt{\mu[\Delta(\mathbf{j})]} \inf_{p\in\mathcal{P}(\widetilde{\Delta}(\mathbf{j}))} \|\phi - p\|_{\infty,\widetilde{\Delta}(\mathbf{j})}. \tag{33}$$

In virtue of polynomial approximation of [9], we have

$$\|\phi - p\|_{\infty,\widetilde{\Delta}(\mathbf{j})} \leq C\Big[\mathrm{diam}(\widetilde{\Delta}(\mathbf{j}))\Big]^{2+(d/2)}|\phi|_{2,\widetilde{\Delta}(\mathbf{j})}. \tag{34}$$

As a consequence, one obtains

$$\|\phi - P_Q(\phi)\|_{0,\Delta(\mathbf{j})} \leq \sqrt{\mu[\Delta(\mathbf{j})]}\Big[\mathrm{diam}(\widetilde{\Delta}(\mathbf{j}))\Big]^{2+(d/2)}|\phi|_{2,Q}. \tag{35}$$

Since $\|\phi\| = 1$, we obtain from the definition (15) that $h(Q)^2|\phi|^2_{2,Q} \leq 1$ or equivalently $|\phi|_{2,Q} \leq 1/h(Q)$. As a consequence, we deduce

$$\|\phi - P_Q(\phi)\|_{0,\Delta(\mathbf{j})} \leq \sqrt{\mu[\Delta(\mathbf{j})]}\Big[\mathrm{diam}(\widetilde{\Delta}(\mathbf{j}))\Big]^{2+(d/2)}h(Q)^{-1}. \tag{36}$$

On the other hand, due to the knot quasi-uniformity (29), we have

$$\sqrt{\mu[\Delta(\mathbf{j})]} = C(\theta)\prod_{\nu=1}^{d}\left[\frac{h_\nu(Q)}{n_\nu - k_\nu + 2}\right]^{1/2}. \tag{37}$$

We use now the Pythagorean rule and the quasi-uniformity (29) again to obtain

$$\mathrm{diam}[\widetilde{\Delta}(\mathbf{j})] = \left[\sum_{\nu=1}^{d}|\zeta^{\nu}_{j_\nu+k_\nu} - \zeta^{\nu}_{j_\nu-k_\nu+1}|^2\right]^{1/2} \tag{38}$$

$$= C(\theta)\left[\sum_{\nu=1}^{d}\Big(\frac{k_\nu(Q)h_\nu(Q)}{n_\nu(Q) - k_\nu(Q) + 2}\Big)^2\right]^{1/2}. \tag{39}$$

Finally, in order to deduce the theorem, take the sum over $\mathbf{j}$ by noting that there are $\prod_{\nu=1}^{d}\big(n_\nu(Q) - k_\nu(Q) + 2\big)$ spline segments $\Delta(\mathbf{j})$ within each $Q \in \mathcal{T}_h$. □

3.3 Cascading from Coarse to Fine Discretizations

The linear system obtained from (16) using the discretization $\mathcal{T}_h$ is solved by a CG (conjugate gradient) with a simple diagonal preconditioner. For the initial guess of the CG, we use *cascading*. That is, the solution from the previous $\mathcal{T}_{h-1}$ is used as a starting value for the CG on $\mathcal{T}_h$. As opposed to the usual Finite Element bases, applying cascading using B-spline basis is not very straightforward. The next description consists of the expression of a function $f \in \mathbb{V}_h$ in terms of the bases of $\mathbb{V}_{h+1} \supset \mathbb{V}_h$ by using *discrete B-splines*.

Consider two knot sequences $\boldsymbol{\zeta} = (\zeta_0, ..., \zeta_{n+k})$ and $\widetilde{\boldsymbol{\zeta}} = (\widetilde{\zeta}_0, ..., \widetilde{\zeta}_{m+k})$ such that $\boldsymbol{\zeta} \subset \widetilde{\boldsymbol{\zeta}}$. We recall [6,7] the discrete B-splines which enable the expression of a coarse basis $N_i^{k,\boldsymbol{\zeta}}$ as a linear combination of fine bases $N_p^{k,\widetilde{\boldsymbol{\zeta}}}$. Choose $a_i \in [\widetilde{\zeta}_j, \widetilde{\zeta}_{j+k})$ and define

$$\phi_i^k(y) := (y - a_i)_+^0 \Psi_i^k(y), \tag{40}$$

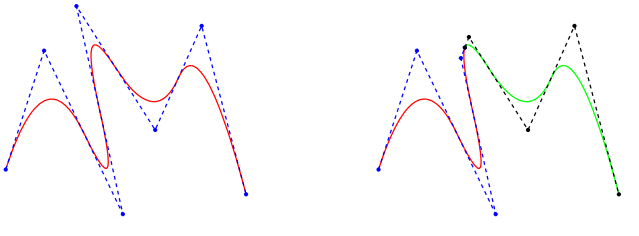
$$(y - a_i)_+^0 := 1 \quad \text{if} \quad y > a_i, \tag{41}$$

$$(y - a_i)_+^0 := 0 \quad \text{if} \quad y \le a_i. \tag{42}$$

where $\Psi_i^k(t) := (t - \widetilde{\zeta}_{i+1}) \cdots (t - \widetilde{\zeta}_{i+k-1})$. One has

$$N_i^{k,\boldsymbol{\zeta}}(x) = \sum_{p=0}^{m} \alpha_i^k(p) N_p^{k,\widetilde{\boldsymbol{\zeta}}}(x) \tag{43}$$

where $\alpha_i^k(p) := (\zeta_{i+k} - \zeta_i)[\zeta_i, \cdots, \zeta_{i+k}]\phi_p^k$.
A fact [6,7] is that a splitting into two B-splines is equivalent to applying knot insertions several times as shown in Fig. 4. Suppose that a 2D element $Q \in \mathcal{T}_h$ has been bisected vertically as in Fig. 3(a) into Q_1 and Q_2. The spline properties on Q are $(n_1, k_1, \boldsymbol{\zeta}^1(Q))$ and $(n_2, k_2, \boldsymbol{\zeta}^2(Q))$. Let $\widetilde{\boldsymbol{\zeta}}^1(Q)$ be defined by inserting the midpoint $\mu := 0.5(\zeta_0^1 + \zeta_{n_1+k_1}^1)$ in the knot sequence $\boldsymbol{\zeta}^1(Q)$ by k_1 times. We



(a) (b)

Fig. 4. (a) Original B-spline (b) Several knot insertions generate two B-splines which have the same parametrization as the original B-spline

have

$$N_i^{k_1,\boldsymbol{\zeta}^1(Q)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)} = \sum_{p=0}^{m} \alpha_i^k(p) N_p^{k_1,\widetilde{\boldsymbol{\zeta}}^1(Q)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)}$$

$$= \sum_{p=0}^{m_1} \alpha_i^k(p) N_p^{k_1,\widetilde{\boldsymbol{\zeta}}^1(Q)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)} + \sum_{p=m_1}^{n} \alpha_i^k(p) N_p^{k_1,\widetilde{\boldsymbol{\zeta}}^1(Q)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)}$$

$$= \sum_{p=0}^{m_1} \alpha_i^k(p) N_p^{k_1,\boldsymbol{\zeta}^1(Q_1)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)} + \sum_{p=0}^{n-m_1} \alpha_i^k(p+m_1) N_p^{k_1,\boldsymbol{\zeta}^1(Q_2)} \otimes N_j^{k_2,\boldsymbol{\zeta}^2(Q)}$$

In the last equality, $\boldsymbol{\zeta}^1(Q_1)$ is the subsequence of $\widetilde{\boldsymbol{\zeta}}^1(Q)$ from the begining until μ while $\boldsymbol{\zeta}^1(Q_2)$ is the remaining subsequence. A similar deduction can be done for the 2D horizontal bisection and for the 3D case.

In practice, the discrete B-splines $\alpha_i^k(p)$ are evaluated by using the recurrence

$$\alpha_i^1(j) = 1 \qquad \text{if} \quad \widetilde{\zeta}_j \in [\zeta_i, \zeta_{i+j}) \tag{44}$$

$$\alpha_i^1(j) = 0 \qquad \text{if} \quad \widetilde{\zeta}_j \notin [\zeta_i, \zeta_{i+j}) \tag{45}$$

$$\alpha_i^k(j) = (\widetilde{\zeta}_{j+k-1} - \zeta_i)\beta_i^{k-1}(j) + (\zeta_{i+k} - \widetilde{\zeta}_{j+k-1})\beta_{i+1}^{k-1}(j) \tag{46}$$

in which

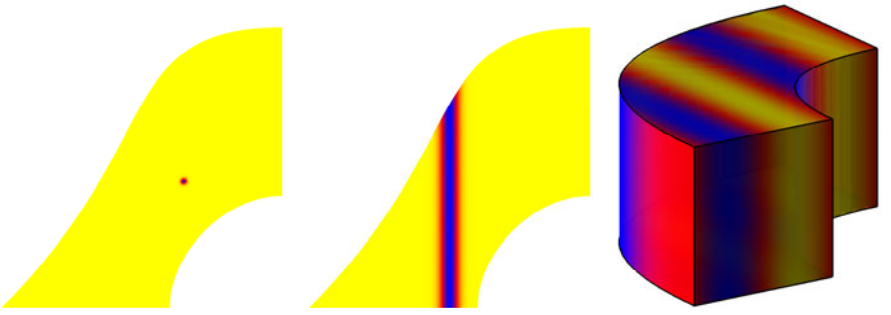
$$\beta_i^k(j) := \begin{cases} \alpha_i^k(j)/(\zeta_{i+k} - \zeta_i) & \text{for} \quad \zeta_{i+k} > \zeta_i \\ 0 & \text{otherwise.} \end{cases} \tag{47}$$

4 Practical Results

In this section, we present some practical results to supplement the previous theory. Let us consider first a 2D adaptive simulation where we consider the exact solution $u(x,y) = \exp\left[-\frac{1}{\alpha}\left((x-a)^2 + (y-b)^2\right)\right]$ on the NURBS physical domain shown in Fig. 5(a) . That function takes unit value at $\mathbf{x} = (a,b)$ which is $(-0.35, 0.75)$ in our test. The large dot in mixed blue and red color inside the physical domain indicates the position where the function nearly takes unit

Table 1. Ratio of exact and estimated errors

Refin. step	Nb. elements	Ratio estim.	Refin. step	Nb. elements	Ratio estim.
0	9	2.595145	9	34	1.889834
2	12	2.177448	10	37	2.129213
3	16	1.616603	13	46	2.199019
5	23	1.708272	18	61	2.299195
6	25	1.796925	20	67	2.649732



(a) (b) (c)

Fig. 5. NURBS physical domains and the exact solutions

values. Note that away from the dot, the function u decays exponentially to zero. The speed of that decay becomes quicker as the value of the parameter α approaches zero. In the depicted case, we have used $\alpha = 1.0e-04$. In Fig. 6, we collect some adaptive history of a few refinement grids. We start from a very coarse discretization $\mathcal{T}_0$ which is a uniform 3×3 tensor product grid in Fig. 6(a). Then, we refine adaptively according to the a-posteriori error estimator described in Section 3. This example illustrates very clearly the situation where the grid refinement takes place strictly within the domain and therefore there is no need to use a fine mesh at the boundary although we deal with a curved physical domain. A mesh-based approach would necessitate a fine mesh at the boundary because the bounding curves are not straight. Approximating the curved portions of the boundary by PL-curves would require a significant number of points which would substantially increase the degree of freedom.

For the next test, we consider the same NURBS for the physical domain but the exact solution is chosen to be $u(x,y) = \exp\left[-\frac{1}{\omega}(x+0.5)^2\right]$. This corresponds to an internal layer whose width is specified by ω as shown in Fig. 5(b). As the parameter ω becomes smaller, the layer gets thinner. In our experiment, we chose the parameter value $\omega = 0.001$. As in the former test, we start again from a very coarse tensor product mesh having 3×3 uniform elements. We apply the previously described theory during the adaptive refinements. The results of that process is depicted in Fig. 7 where the a-posteriori error estimator can efficiently detect the position of the internal layer. It is plainly observed that the elements which are far from the layer are very coarse. That is for example the case of the top right element which is intact from beginning till the finest discretization shown in Fig. 6(c). In addition, we gather in Table 1 the averages of the ratio between the exact error and the a-posteriori error indicator.

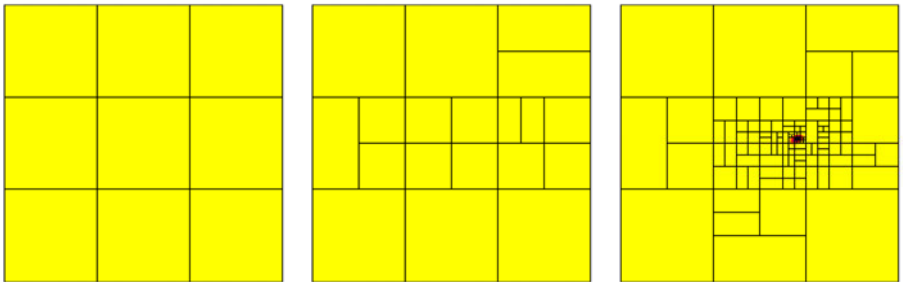


Fig. 6. Adaptive refinement for an internal accumulation

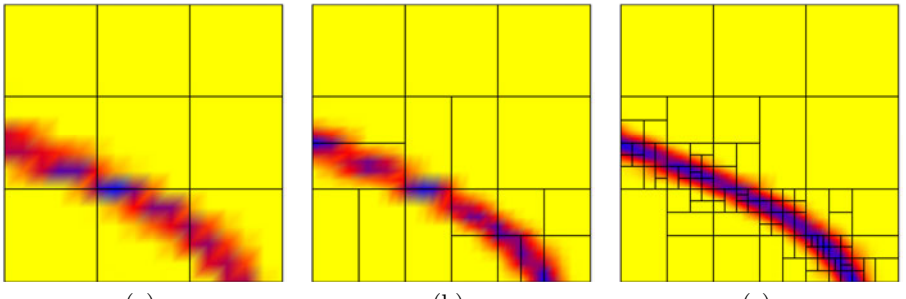


Fig. 7. Local spline adaptivity for a thin internal layer

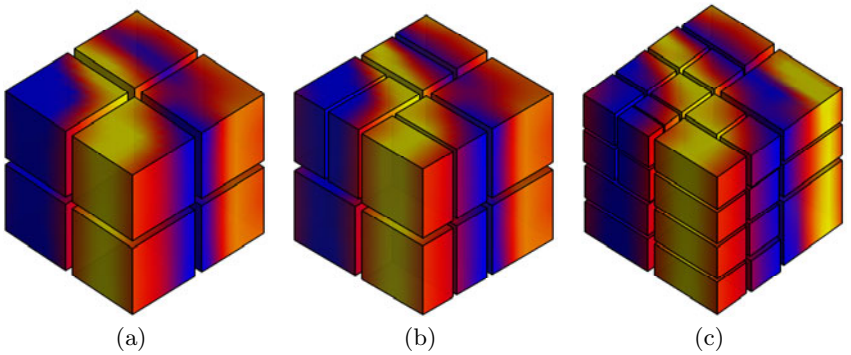


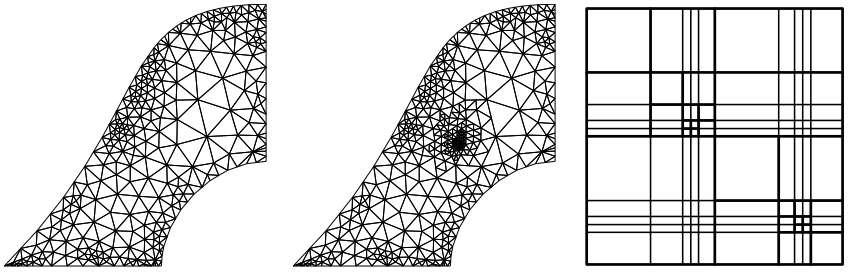
Fig. 8. Adaptive refinement in 3D: the elements are slightly shrunken in order to see the refinements

In addition to the planar situation, we perform also some practical tests in the 3D case. To that end, we consider the exact solution $u(x, y, z) = \sin(\pi x)$. The 3D NURBS for the physical domain is depicted in Fig. 5(c). The history of some adaptive refinements for this case is summarized in Fig. 8.

We would like now to address an example to compare the current method with the traditional discontinuous Galerkin method. For that, we consider the model in Fig. 5(a) again but with respect to a mesh-based DG approach. Since the traditional DG method uses a mesh, we have to approximate the boundary of the model by piecewise linear curves. That approximation is independent of the solution to the PDE. In fact, 530 triangular elements are required to obtain the approximation in Fig. 5(a) which corresponds to a geometric domain accuracy of $1.2216e-03$. Much more elements are required for a more accurate geometric approximation. In contrast, the B-spline DG method uses an exact representation of the geometry with only 9 elements as shown in Fig. 6(a) and Fig. 5(a) because B-splines are used on the parameter domain. It is that significant gain in terms of the required number of elements which mainly makes our proposed method more interesting than the traditional one. Now, we would like to study the the average local errors in each element for the adaptive case. The current approach produces 12 B-spline pieces which have each 4×4 spline segments yielding $12 \times 16 = 192$ elements. The corresponding local error amounts to $1.1286e-02$. In the traditional DG-method, 710 non-conforming elements as shown in Fig. 9(b) produce an average local error of $4.0052e-01$. Another factor of the superiority of our approach over the traditional DG is the ability of using discrete B-splines for cascading as explained in Section 3.3. Now, we want to make a comparison of our approach with the traditional B-spline case. Since a B-spline is nothing else but a piecewise polynomial on a tensor product grid, it is worth noting that the main weakness of the usual B-splines is related to *refinements* and *adaptivity*. It mainly enables global refinements instead of local ones as in Fig. 9(c). For global refinements, an insertion of a new knot in one direction spreads along the complete range of the other directions. For the sake of contrast of the two methods, we use again the problem in Fig. 7 and we gather in Tab. 2 some data which describe the ratio of the required numbers of elements for the usual B-spline method and the DG B-splines. We display there too the regularity condition in (10) which shows the quality of our method compared to the usual B-splines.

Table 2. Comparison of DG B-splines and usual B-splines

DG B-spline		Usual B-spline		RATIO
Nb. elements	Shape regularity	Nb. elements	Shape regularity	
11	1.363636	16	1.500000	68.750 %
49	1.530612	221	1.846154	22.172 %
67	1.432836	300	1.640000	22.333 %
88	1.431818	399	1.691729	22.055 %
133	1.488722	972	1.879630	13.683 %



(a) (b) (c)

Fig. 9. (a)Mesh for the traditional DG-method (b)Nonconforming DG-mesh (c) B-spline with global refinement

5 Conclusion and Discussion

We have presented a method based on B-spline bases using discontinuous Galerkin scheme. Our main emphasis is that instead of working on the physical domain, all computations are carried over onto the parametric domains. That has two advantages: ability of keeping exact geometries, using low degree of freedom and hence reducing the numerical costs. The method is fully adaptive and it works on two and three dimensions. A difficulty about the method is the representation of the whole geometry into tensor product structure.

References

1. Ainsworth, M.: Review of Adaptive Finite Element Methods for Differential Equations by W. Bangerth and R. Rannacher. SIAM Rev. 46, 354–356 (2004)
2. Arnold, D., Brezzi, F., Cockburn, B., Marini, L.: Unified Analysis of Discontinuous Galerkin Methods for Elliptic Problems. SIAM J. Numer. Anal. 39(5), 1749–1779 (2002)
3. De Boor, C., Fix, G.: Spline Approximation by Quasiinterpolants. J. Approx. Theory 8, 19–45 (1973)
4. Castillo, P., Cockburn, B., Perguis, I., Schötzau, D.: An A-Priori Error Analysis of the Local Discontinuous Galerkin Method for Elliptic Problems. SIAM J. Numer. Anal. 38(5), 1676–1706 (2000)
5. Cockburn, B., Kanschatr, G., Perguis, I., Schötzau, D.: Superconvergence of the DG Method for Elliptic Problems on Cartesian Grids. SIAM J. Numer. Anal. 39(1), 267–285 (2001)
6. Cohen, E., Lyche, T., Riesenfeld, R.: Discrete B-Splines and Subdivision Techniques in Computer Aided Geometric Design and Computer Graphics. Comput. Graphics and Image Process. 14, 87–111 (1980)
7. Cohen, E., Lyche, T., Schumaker, L.: Degree Raising for Splines. J. Approx. Theory 46, 170–181 (1986)
8. Coons, S.: Surfaces for Computer Aided Design of Space Forms. Technical report, Project MAC, Department of Mechanical Engineering in MIT, Revised to MAC-TR-41 (1967)

9. Dupont, T., Scott, R.: Polynomial Approximation of Functions in Sobolev Spaces. Math. Comput. 34(150), 441–463 (1980)
10. Farin, G., Hansford, D.: Discrete Coons Patches. Comput. Aided Geom. Des. 16(7), 691–700 (1999)
11. Floater, M.: An $O(h^{2n})$ Hermite Approximation for Conic Sections. Comput. Aided Geom. Design 14, 135–151 (1997)
12. Forrest, A.: On Coons and Other Methods for the Representation of Curved Surfaces. Comput. Graph. Img. Process. 1, 341–359 (1972)
13. Gormaz, R.: A Class of Multivariate de Boor-Fix Formulae. Comput. Aided Geom. Des. 15, 829–842 (1998)
14. Harbrecht, H., Randrianarivony, M.: From Computer Aided Design to Wavelet BEM. Comput. Vis. Sci. 13(2), 69–82 (2010)
15. Harbrecht, H., Randrianarivony, M.: Wavelet BEM on Molecular Surfaces: Parametrization and Implementation. Computing 86, 1–22 (2009)
16. Höllig, K., Reif, U.: Nonuniform WEB-Spline. Comput. Aided Geom. Des. 20(5), 277–294 (2003)
17. Hughes, T., Cottrell, J., Bazilevs, Y.: Isogeometric Analysis: CAD, Finite Elements, NURBS, Exact Geometry, and Mesh Refinement. Comput. Methods in Appl. Mech. Eng. 194, 4135–4195 (2005)
18. Randrianarivony, M., Brunnett, G.: Preparation of CAD and Molecular Surfaces for Meshfree Solvers. Lect. Notes Comput. Sci. Eng. 65, 231–245 (2008)
19. Randrianarivony, M.: Geometric Processing of CAD Data and Meshes as Input of Integral Equation Solvers. PhD thesis, Technische Universität Chemnitz, Germany (2006)
20. Randrianarivony, M., Brunnett, G.: Molecular Surface Decomposition Using Geometric Techniques. In: Conf. Bildverarbeitung für die Medizine, Berlin, pp. 197–201 (2008)
21. Randrianarivony, M.: On Global Continuity of Coons Mappings in Patching CAD Surfaces. Comput.-Aided Design 41(11), 782–791 (2009)
22. U. S. Product Data Association: Initial Graphics Exchange Specification. IGES 5.3. Trident Research Center, http://ts.nist.gov/standards/iges

Development of a Didactic Model of the Hydrologic Cycle Using the TerraME Graphical Interface for Modeling and Simulation

Tiago Lima[1], Sergio Faria[2], and Tiago Carneiro[1]

[1] TerraLAB, Computer Science Department, Federal University of Ouro Preto
Ouro Preto, Minas Gerais, Brazil
[2] Department of Cartography, Federal University of Minas Gerais
Belo Horizonte, Minas Gerais, Brazil
tiagofml@gmail.com, sergiofaria@ufmg.br, tiago@iceb.ufop.br

Abstract. Many modeling and simulation platforms provide general programming languages as interfaces for model construction. Some offers high-level modeling languages with conceptual basis and services to represent data structures and rules that will determine the model behavior. However, the direct use of a computational language is still a limiting factor to the broad usage of these platforms. Modelers often have different scientific backgrounds, presenting a lack of background on algorithms and programming techniques. Furthermore, there is no established methodology for model development. These problems confuse the modelers forcing them to deviate their attention from the problem being solved. We argue that a visual integrated development environment (IDE) can solve these problems, making easy the understanding and communication of the model conception and design. An IDE can also enforce the use of a common model development methodology. In this paper we describe a methodology for modeling Earth system phenomena using the TerraME GIMS tool, which is a visual IDE for the TerraME modeling and simulation platform. It enables users to build environmental models through visual metaphors that graphically describe models structure. We demonstrate the use of TerraME GIMS and present our methodology for the development of a didactic model for the hydrologic cycle. Future works include the development of diagrams to better describe the model behavior, including agent synchronization and communication.

Keywords: Environmental Modeling; Modeling Methodology; Visual Programming; TerraME GIMS; Integrated Development Environment.

1 Introduction

The Earth system comprises the interaction between socio-economic systems (of anthropic origins as the land use system) and biophysical systems (of natural origin as the ecological and atmospheric systems). In general, Earth system phenomena have a complex nature. This complexity requires the use of modeling and simulation techniques and tools to study, to understand and to represent systems behavior.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 75–90, 2011.

Furthermore, to deal with these problems it is necessary a multidisciplinary team of specialists from different fields of knowledge. One of the major challenges is to make explicit the differences that each team member has about phenomena behavior and about the model conception and design. Sophisticated user-friendly computational platforms are required to deal with this latter problem.

Modeling involves the construction of a simplified representation of the reality. It enables one to clearly define problems and concepts. It provides means of analyzing the observed behavior representing it in a synthetic environment and reporting simulation outcomes [1]. Modeling and computer simulation have been used in scientific researches to address problems of complex nature. When the solution has a high cost or cannot be obtained through experimentation [2]. Therefore, these methods and tools are essential to study terrestrial systems behavior. Typically, represented as spatial dynamic models, which describe spatial patterns of changes evolving over time [3].

The TerraME - Terra Modeling Environment [4] is a software platform for modeling and simulation of environmental phenomena which allows building dynamic spatial models integrated to Geographic Information Systems (GIS). Its main users are specialists involved in the development of environmental models and whose major knowledge is about the application domain, like geographers, ecologists, biologists, anthropologists, sociologists and economists. Although TerraME offers a high level programming language for model description called TerraML (TerraME Modeling Language), it still demands some programming skills from the user. During TerraME courses it is common the user to lose focus on modeling and TerraME concepts to pay attention on programming language syntax, style and on programming techniques. Therefore, a visual Integrated Development Environment (IDE) that allows model specification through graphical metaphors will be useful for keeping the focus of TerraME's users on solving problems in the application domain.

The TerraME GIMS (TerraME Graphical Interface for Modeling and Simulation) [5] is the graphical IDE of the TerraME platform. The TerraME GIMS is designed to make model development a more intuitive and effective task, increasing the productivity of the TerraME current users and decreasing the learning effort of new users. It enables users to visually describe models through interactions with graphical components such as menus, tool bars, model file trees, model structure specification diagrams, finite state machine diagrams, etc. The IDE automatically generates the TerraML source code corresponding to the model structure and behavior.

However, TerraME GIMS and TerraML modelers must understand the same basic concepts about the Nested-CA model of computation [4], which is the kind of automaton implemented by TerraME. This fact implies in the use of a common methodology for model development, which this paper briefly describes. To evaluate the use of a visual IDE for supporting this methodology, we have developed a didactic model to the hydrological cycle using both tools, the TerraML language and the TerraME GIMS interface. The hydrologic cycle model has been chosen due to the intuitive behavior of the processes it represents. Afterward, the manually generated and the automatically generated source code are compared to analyze their differences. This analysis demonstrates that the TerraME GIMS implements all

TerraME abstractions and can make easier and clearer the modeling process. Finally, the advantages and limitations on the use of TerraME GIMS to describe realistic environmental models for the TerraME platform are discussed.

2 TerraME and Nested-CA Basic Concepts

TerraME[1] is a software platform for building and simulating spatial dynamic computational models of geographical phenomena. It implements the Nested Cellular Automata (Nested-CA) model of computation, which provides essential features to study and represent Earth systems [4]. A realistic model construction is enabled through the TerraME integration with GIS, allowing to feed models with temporal series of detailed maps and sensors data.

TerraME users with good programming expertise can implement models directly using the TerraME C++ core. On the other hand, users with limited knowledge on programming can alternatively use the modeling language TerraML to build their models. This high level programming language is an extension of the LUA programming language [6] and adds special types and services for environmental modeling.

According to the Nested-CA conceptual design, which is in conformity with the Scale concept of Gibson et al. [7], the representation of a phenomenon in TerraME is performed from the description of the phenomenon behavior in time and space. The behavioral, spatial and temporal aspects of a phenomenon are viewed to form an indivisible model building block (or module) called Scale. Scales can be also seen as micro-worlds, i. e., as a synthetic (virtual) environment in which analytical entities (rules) change spatial properties over time. Scales can be nested to allow model decomposition, so that complex models (worlds) can be described from the composition of simpler ones. Each Scale can be used to represent different processes of a system, to model a different aspect of the same process, or to describe the same process at the different temporal or spatial resolution. Scales are implemented as Nested-CAs. This way, the Nested-CA design enables multiple scale model development.

In detail, the Nested-CA (or scale) is a container of two kinds of autonomous behavioral models, which may change spatial properties over time. The Automaton model is useful to represent continuous fields evolving over time, for example, temperature variation in a certain region. Its structure and functioning are based on the Cellular Automata Theory [8]. Meanwhile, the Agent model is based on the Agent Theory [9] and is useful to represent discrete entities or individuals, like institutions, persons, animals or vehicles. The internal state and the behavior of both models are determined by a Hybrid Automaton [10], whose structure and semantic allows the simulation of discrete, continuous or hybrid systems. Both models can communicate through message exchange. They also have access to a topological structure which defines their neighbors. The neighborhood structure is very flexible and is implemented as a weighted directed graph. The Nested-CA is also a container of cellular space [11] models, which are functions of indexes (like coordinates) into a set

[1] http://www.terrame.org/

of irregular cells used to represent the local aspects of the geographic space. Several discrete event schedulers can be embedded into a Nested-CA to describe the phenomenon dynamics based on the DEVS (Discrete Event System Specification) formalism proposed by Bernard Zeigler [12].

The TerraME software has been used in real case studies. Andrade et al. [13] developed models for spatial games and showed results demonstrating how mobility affects the Nash's equilibrium. A model for the simulation of fire propagation patterns in the Emas National Park, GO - Brazil was developed and presented by Almeida et al. [14]. Moreira et al. [15] used TerraME to understand how to represent spatial relation across scales in computational models for the land use system. TerraME has also been used by Pimenta et al. [16] in a case study in which it was analyzed how the existence of different rules of land use affects the landscape dynamics at regional level.

3 The TerraME Model Development Methodology

Modeling can be performed using different methods and processes [1,4,17]. However, we assume that models should be developed in an iterative and incremental way, such that at the end of each development cycle an improved model version is produced and the phenomenon becomes better understood. Each development cycle includes the following steps: (1) definition of the problem or scientific question, (2) development of the conceptual model, (3) choosing the appropriated modeling approach or paradigm, (4) development of the model by mathematical equations and relationships, (5) computational implementation including tests and code documentation, (6) calibration of model parameters, (7) evaluation of the model through comparison with empirical observations, (8) use of the model to explore alternative simulated scenarios, to make predictions and to support the decision making process.

During the first step, workshops, field works and literature review are essential activities. As the phenomenon behavior can be influenced by different driving forces at different scales [18], no single scale model can describe complex Earth System phenomena. Therefore, in the second step, the modeler should identify the major proximate factors and underlying driving forces affecting the phenomena behavior [19]. The modeler also needs to define the scales in which the phenomenon will be modeled. For this step, she has to choose the extents and resolutions those will be used to represent the phenomena in each scale dimension: behavior, space and time. The term extent stands for the magnitude of the measures used to represent the phenomenon. Resolution refers to the granularity of the measures [7]. Therefore, modelers have to decide about the actors and processes those will be taken into account. They should also choose the boundary of the region under study and the longest time period considered in the study. Afterward, the modelers should identify the resolutions in time and space in which changes in the observed behavior occurs. At this point, the modeler will have chosen the temporal, spatial and behavioral extents and resolutions in which the phenomenon will be modeled. Each chosen combination of extent and resolution for the dimensions space, time and behavior will give rise to one scale.

Nested-CA brings together the most useful abstractions to model environmental systems. The social aspects from Agent Theory, the spatial structures from the Cellular Automata Theory and the flexible specification of the dynamics allowed by the DEVS formalism are combined into a simple model of computation. The Hybrid Automaton used to implement Nested-CA Agents and Automata are capable of going through discrete changes expressed as internal transition rules, and of going through continuous changes described as differential equations. This way, in the third step the user can feel free to choose any of these paradigms to model different aspects of a phenomenon.

However, as Nested-CA Agents and Automata are autonomous machines, exogenous forces cannot ever change their internal state. Only internal transition rules and differential equations can make it. Hence, agents cannot give orders to other agents. But an agent can ask for instructions. The automaton in a central cell cannot directly change the internal state of automata within the neighbor cells. Nevertheless, it can change its internal state based on the information it gathered from its neighbors.

In the hierarchy of nested scales, scales in the higher levels must provide overall control to their internal scales. Each internal scale should be seen as a black box, i. e., as a module. For this reason, the information flow between scales should respect the hierarchy. Internal scales may change its state based on information coming from the parent scale. The parent scale must intermediate communication between internal scales. Elements in a scale should never directly change the internal state elements in other scales. Autonomy is the essential feature that legitimates the patterns that emerge in higher levels of organization from the local interactions between pair of Nested-CA components.

The modeler can use the TerraML or the TerraME GIMS tool in addition to the TerraView[2] GIS to accomplish the fourth and fifth steps. Finally, the TerraME VCToolkit can be used for model validation and calibration [20]. This way, modeling is a cyclic and incremental process, in which a model is built, reviewed and evaluated. In each cycle the understanding of the observed reality is expanded and improved.

4 TerraME Graphical Interface for Modeling and Simulation

As an extension of the Lua programming language, the TerraML syntax is too flexible to prevent modelers to produce source codes in which agents and automata directly change the internal state of each other. TerraME GIMS is a visual IDE for the TerraME platform. It enables the development of spatial dynamic models through the use of graphical representation of model structure and behavior. TerraME GIMS is implemented as a set of plug-ins that runs under the Eclipse[3] platform, adding capabilities to the visual specification and building of spatial dynamic models, automatically generating the model source code. TerraME GIMS has a layered software architecture, as illustrated in Fig. 1. It constitutes a new layer between end-users and the TerraME platform. The Eclipse platform appears as an intermediate layer between the TerraME modeling and simulation engine and the TerraME GIMS plug-in.

[2] http://www.dpi.inpe.br/terraview/
[3] http://www.eclipse.org/

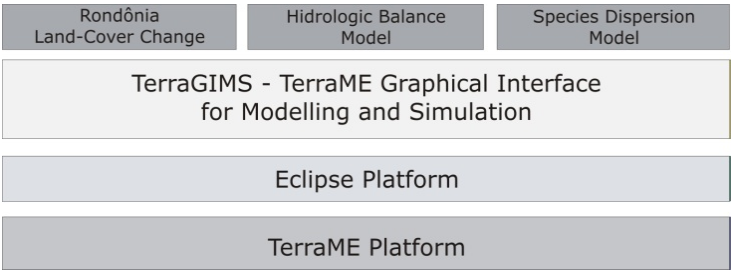


Fig. 1. TerraME GIMS layered software architecture

The TerraME GIMS graphical user interface is composed by a main editor, called Graphical Editor, and the views Project Explorer, Outline and Properties, as illustrated in Fig. 2. Through Project Explorer View, users can access the files that are part of a project and navigate in model hierarchy. The hierarchy structure is shown as a tree of model components. The Graphical Editor shows the graphical representation of the model component selected in the Project Explorer. It is always possible to edit the component data structure and rules by changing its graphical representation. In Properties View, the values of model component properties can be visualized and edited through widgets as text fields, list boxes and check buttons. The Outline view provides a general overview of the model structure. This way, TerraME GIMS allows for the visual development of spatial dynamic models for Earth System phenomena.

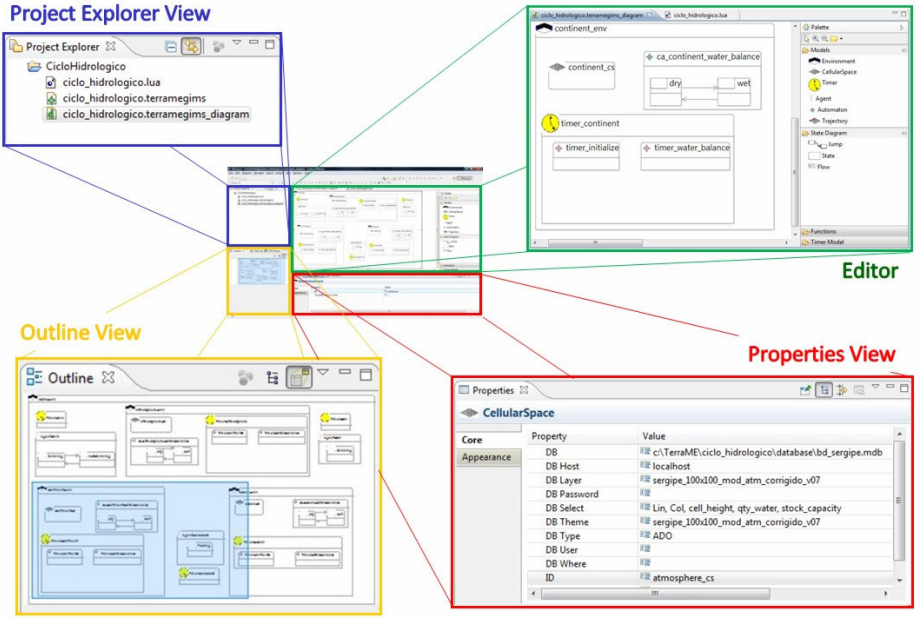


Fig. 2. TerraME GIMS graphical user interface overview

5 Case Study: A Spatial Dynamic Model for the Hydrologic Cycle

A didactic spatial dynamic model for the Hydrologic Cycle was developed in order to evaluate the TerraME model development methodology and to assess the TerraME GIMS utility and correctness. The Hydrologic Cycle was chosen due to the intuitive behavior of its processes, making easier the interpretation of simulation outcomes. As this model requires the simulation of three different environments, i.e., land, ocean and atmosphere, it can be decomposed into three different nested scales. It also requires model integration with geographic database and the modeling of environmental phenomena from three different points of view: Spatial, temporal and behavioral. Furthermore, phenomenon behavior is simulated in the discrete and continuous way. Thus, this model is representative regarding the usage of data structures and services available to TerraME users. And it is also representative regarding the main challenges encountered in Earth System modeling.

In the following sections, the steps proposed in the TerraME model development methodology will be performed in order to build the hydrologic model: 1) problem definition, 2) conceptual model development, 3) computational and mathematical model definition, 4) geographical database and model implementation, 5) model calibration and validation, 6) simulation outcome analysis. A coastal region in the Brazilian state named Sergipe was chosen as study area.

5.1 Problem Definition: the Hydrologic Cycle

The hydrologic cycle (or water cycle) is the global phenomenon of water movement between the land surface and atmosphere. It can be considered a closed system at a global level [21, 22]. The hydrologic cycle occurs in the Earth's surface, in the

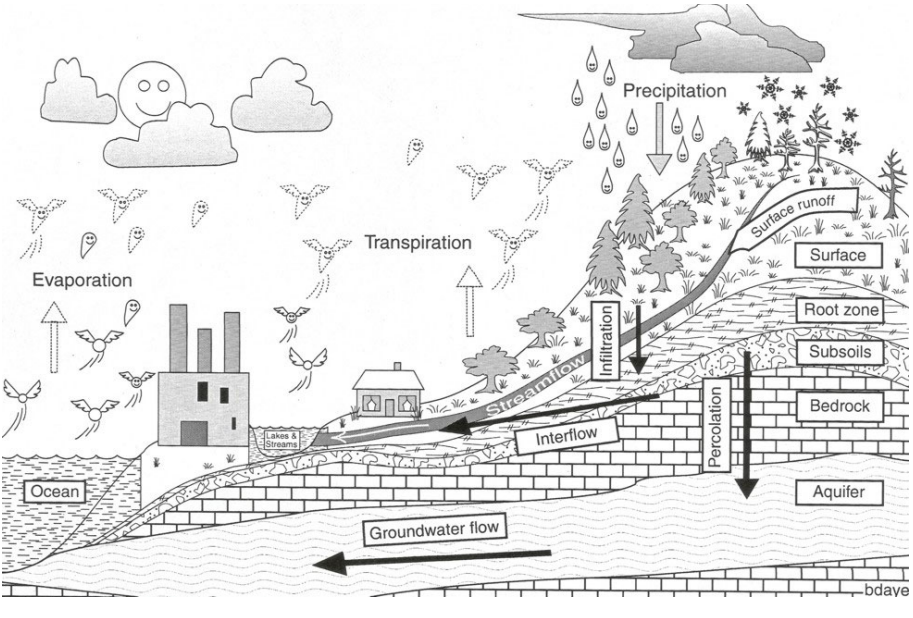


Fig. 3. Hydrologic cycle [21]

atmosphere, and in the interaction between them, as illustrated in Fig. 3. One phase of the cycle happens in land surface, which covers continents and oceans, where water circulation takes place inside and on the surface of soils and rocks, on oceans and on living beings. The water that circulates in the atmosphere comprises another phase. Finally, closing the cycle, there is an exchange between the water circulating on land surface and on atmosphere, which occurs in two ways: 1) towards of Earth's surface to atmosphere, where the transfer of water occurs primarily in vapor form, due to evaporation and transpiration phenomena; 2) towards of atmosphere to Earth's surface, where the transfer of water can happen in all the three physical states, being precipitation of rain and snow the most significant on a global scale [21].

5.2 Conceptual Modeling

In the hydrologic cycle, the interactions between atmosphere and Earth's surface (formed by the continents and oceans) happens from the water flow between them, which are represented by the evaporation (water flow from surface to atmosphere), precipitation (water flow from atmosphere to surface) and drainage (water flow from mainland to ocean) processes. For didactical reasons, the Earth's surface is separately modeled as two different micro-worlds or scales: Continent and ocean. Moreover, we assume that evaporation occurs only form the ocean to the atmosphere and that the precipitation occurs only from the atmosphere to the continent. In addition, the water flow also takes place inside the atmosphere, ocean and continent environments. The processes of convection, surface runoff and infiltration are also modeled. The water flow is mainly controlled by the region topography. On the continent surface and on the ocean bottom, the water flows from higher places to lower ones due to the surface runoff process. In the atmosphere, the water vapor flows from lower places to higher ones due to the convection process. The conceptual model is illustrated in Fig. 4.

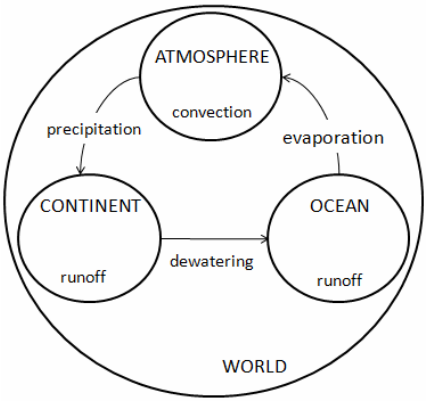


Fig. 4. Conceptual model of water flow in the hydrologic cycle

5.3 Computational and Mathematical Model Definition

In order to simplify the model development, the flows of water within the continent, ocean and atmosphere environments have been implemented in the same way. They have been represented by Cellular Automaton in which each cell is occupied the Hybrid Automaton (HA) shown in Fig. 5. A finite state machine with two states describes the HA discrete behavior. When in the state unsaturated, the water flow from higher cells to lowers cells is simulated by the exponential growth equation of the form $dW(t)/dt = k * W(t)$, where $dW(t)/dt$ is the instantaneous water flow, k is the flow coefficient in [0, 1] and W(t) is the amount of water stored in the higher at time t. In the saturated discrete state, the HA continuous behavior simulates the distribution of water among lower neighbor cells, that is, lower neighbor cells receives the equal amount of water coming from a higher central. Therefore, this water flow is describe by the equation $dW(t)/dt = W(t)/N$, where N is the number of neighbor cells. The HA changes from the unsaturated state to the saturated state when the storage capacity where it is embedded is reached. In the opposite direction, when the stored amount of water is under the cell capacity, the HA goes back to the unsaturated discrete state.

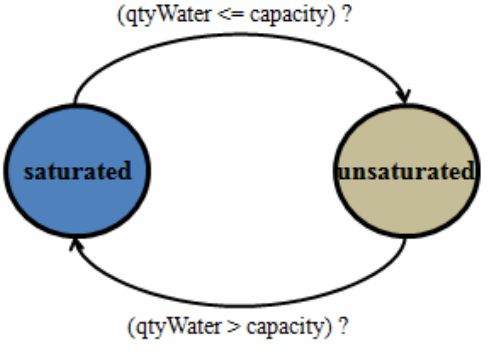


Fig. 5. Representation of the water flow hybrid automaton

The flows of water between scales, that is, the evaporation, precipitation and drainage processes have been implemented as Agents objects embedded in the "world" environment. These agents are also represented by the HA shown in Fig. 5.

5.4 Model Implementation Using TerraME GIMS

During the model geographic database construction, a SRTM image was used to represent the topography of the study area, a small size coastal region containing areas in the continent and ocean (Fig. 6). Then, the image was used to generate three regular grids of cells with the local attribute "altimetry" (Fig. 7). In TerraML, these grids of cells use to model space properties are named "CellularSpace".

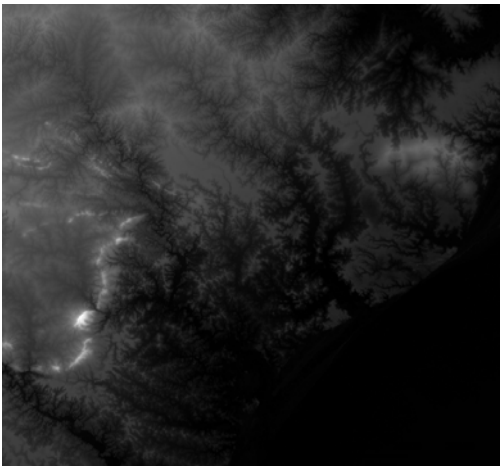
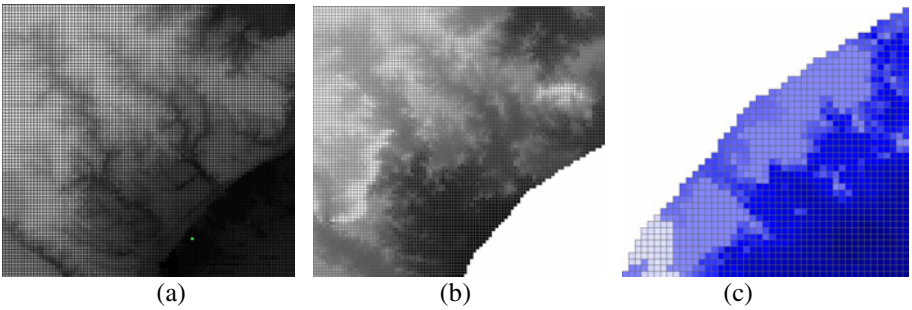


Fig. 6. SRTM image of the study area (SC–24–ZB). Source: Embrapa[4]

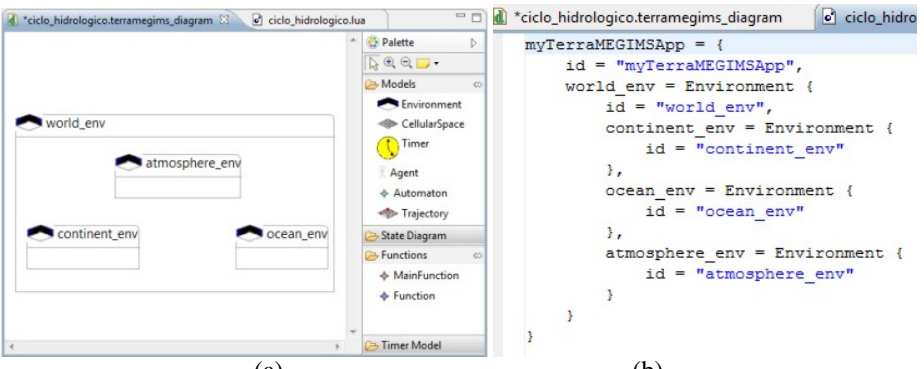


(a) (b) (c)

Fig. 7. Cellular spaces used to represent the study area topography in the three model environments: (a) atmosphere, (b) continent, and (c) ocean

In the TerraML, the Nested-CA concept of scale or micro-world is designated by the reserved word "Environment". In accordance with the Nested-CA structure, "Environment" objects are container of "Timer", "CellularSpace", "Automaton" and "Agent" objects. This way, the conceptual model was implemented as four nested Environment objects. The "world" environment represents the Earth and involves the environments "continent", "ocean" and "atmosphere" in its interior. For each one of these internal environments, the spatial and temporal behaviors of their inner processes were also implemented. A TerraME GIMS application is defined by an object named TerraMEGIMSApp located at the higher level of the model component hierarchy. This object is a container of "Function" and "Environment" objects. Using the TerraME GIMS graphical editor, one can visually specify the nested structure of the hydrologic cycle, as Fig. 8a shows. The automatically generated TerraML source code is shown in Fig. 8b.

[4] http:// www.relevobr.cnpm.embrapa.br/



(a) (b)

Fig. 8. Definition of nested environments using TerraME GIMS: (a) graphical representation; (b) corresponding TerraML source code

Then, the spatial model of each internal environment was implemented by the addition of CellularSpace objects to their interior. The properties of each CellularSpace object were edited directly in the TerraME GIMS Properties view, as illustrated in Fig. 9.

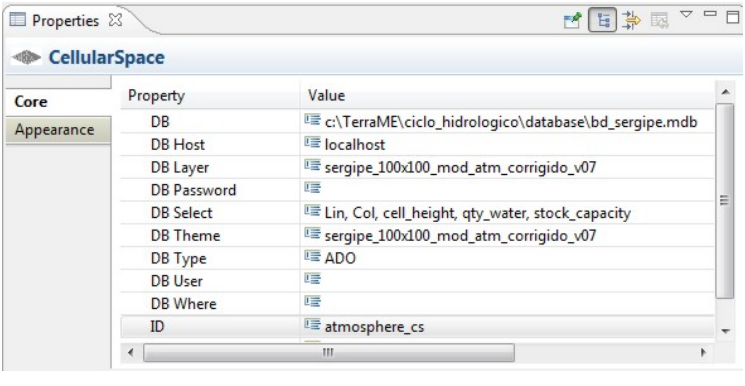
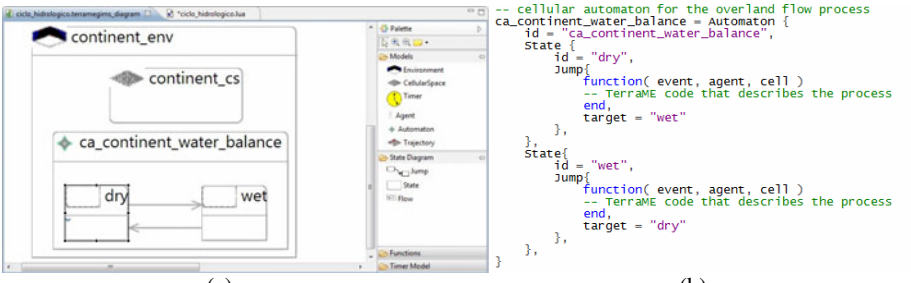


Fig. 9. Definition of spatial models

The behavioral model, shown in Fig. 5, has been implemented using the graphical metaphors presented in Fig. 10a. The finite state machine inside the continent environment has two states equivalent to the internal discrete states unsaturated (wet) and saturated (dry) discussed in section 5.3. The automaton's automatic generated TerraML code is also illustrated (Fig. 10b).



(a) (b)

Fig. 10. The cellular automaton used to simulate the drainage process: (a) the graphical representation, and (b) the corresponding source code

Once defined the spatial and behavioral models, it remains only to define the temporal model. The temporal models for the environments continent, atmosphere and ocean are similar. They are composed by two pairs {event, message} (Fig. 11): (i) the first event occurs only in the model startup to initialize cell attributes, such as, waterQuantity and waterCapacity; (ii) the second event occurs at each simulation step to simulate the internal flows of water.

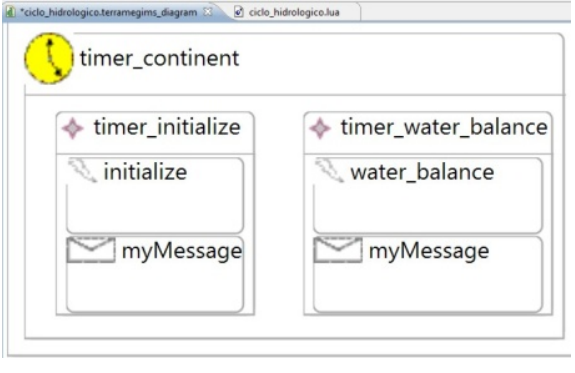


Fig. 11. Definition of continent temporal model

The flows of water between the environments occur through the processes of evaporation, precipitation and drainage and are implemented respectively by the agent agent_rain, agent_sun and agent_debouch. The agent agent_rain moves through the cellular space of atmosphere transferring part of water from each cell (in a saturated state) to the continent cell in the same geographical location. The agent agent_sun travels through the ocean cellular space transferring part of water from each cell to the corresponding cell in the atmosphere. The agent agent_debouch transfers part of water from each cell in coastal continent region to those cells in the ocean that are immediately adjacent to the continent cells. The temporal model of the environment "world" consists of three pairs {event, message} those execute the sun, rain and debouch agents. The graphical representation of hydrologic cycle model, including the spatial, temporal and behavioral dimensions, is presented in Fig. 12.

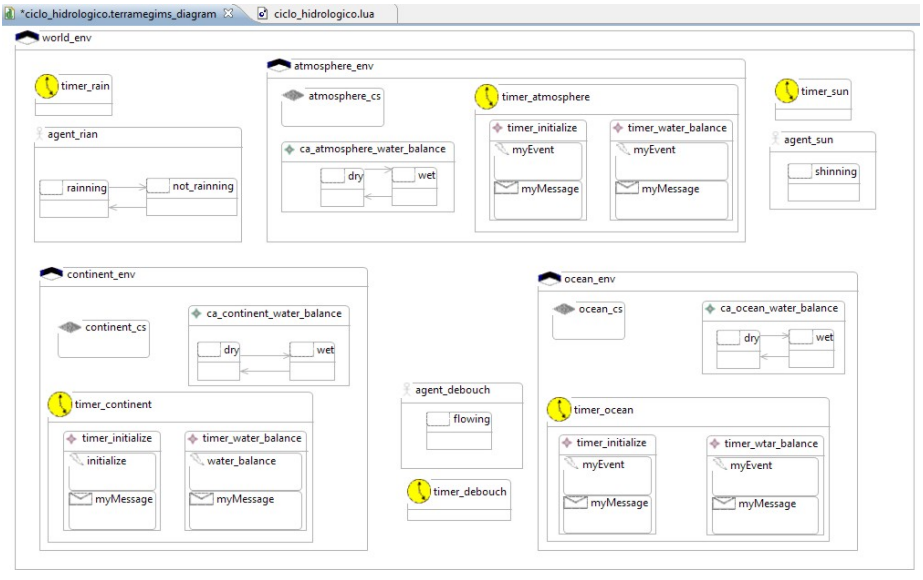
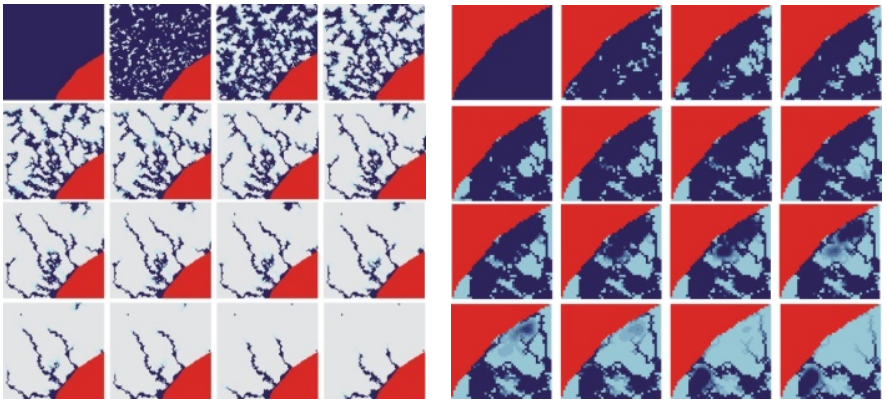


Fig. 12. Overview of the graphical representation of the hydrologic cycle model

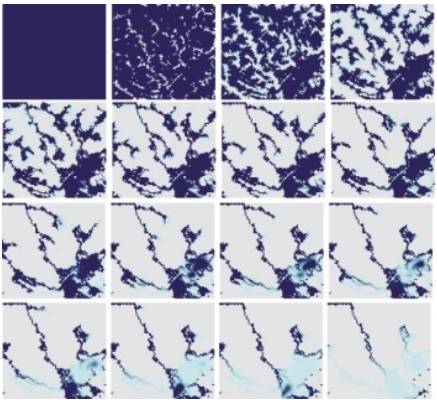
5.5 Simulation Outcome Analysis

The model validation was done only to verify if the total quantity of water flowing in the model was keep constant (mass conservation) during all the simulation steps. The simulation outcomes are shown in Fig. 13. It is possible to observe the emergence of global patterns in the higher organization levels from the interaction of simple rules executed at the local levels.



(a) (b)

Fig. 13. The images showing the flow of water in each environment, from the left to right and from the top to the bottom: (a) water drainage in the continent, (b) water drainage in the ocean, (c) water vapor convection in the atmosphere. Legend: Red means “no data”, gray means “no water” and blue varies from dark to light to mean the amount of water in the cells.



(c)

Fig. 13. (*continued*)

6 Final Remarks

In order to evaluate the advantages brought by the use of a visual IDE in the development of spatial dynamic models of Earth System phenomena, we have outlined a methodology for model development using the TerraME modeling and simulation platform. Then, we have implemented an Eclipse plug-in, named the TerraME GIMS, which supports the use of this methodology. It allows graphical and interactive specification of spatial dynamic models whose source code is automatically generated. Finally, we have developed a spatial dynamic model to the hydrologic cycle using the TerraME GIMS plug-in. This didactic model illustrates the use of main functionalities of the TerraME platform and some of the challenges found in the development of multi scales models to Earth System phenomena.

One limitation in the use of graphical interface for modeling is the lack of flexibility regarding the customization of the automatically generated code. On one hand, the standardized structure of the generated code may limit users to freely organize their model. On other hand, it can be a great benefit to users which do not have much experience in programming and code organization, contributing to the model readability and maintainability. Moreover, another advantage of automatic code generation is the reduction of the user's effort to understand and use the rigid syntax of programming languages, allowing them to focus on the phenomenon representation.

Moreover, the communication of the model conception and design through a graphical representation allows modelers to identify the model components and the relationships between them in an easier and more intuitive way. The graphical representation to the hydrologic cycle model is very close to the conceptual model shown in section 5.2.

Although TerraME GIMS is still a prototype, it is capable of graphically represent Nested-CA basic concepts. It can be use to completely describe model structure. However, some effort is still required to make possible the complete description of the model behavior through graphical metaphors. Future works also include the

evaluation of the TerraME GIMS in experiments involving groups of users. Measures such as modeling speed, frequency of errors, effectiveness of model communication still have to be collected. Both TerraME and TerraME GIMS are under continuous development at TerraLAB[5] [23].

Acknowledgements. We would like to thank CAPES and CNPq (CTINFO-2010) for the financial support.

References

1. Turner, M.G., Gardner, R.H., O'Neill, R.V.: Landscape Ecology in Theory and Practice: pattern and process. Springer, New York (2001)
2. Bratley, P., Fox, B.L., Schrage, L.E.: A guide to Simulation, 2nd edn. Springer, Heidelberg (1987)
3. Pedrosa, B.M., Câmara, G.: Modelagem dinâmica e sistemas de informações geográficas. In: Meirelles, M.S.P., Câmara, G., Almeida, C.M. (org.) Geomática - Modelos e aplicações ambientais, Embrapa, Brasília, DF, Brasil, vol. 5, pp. 235–280 (2007)
4. Carneiro, T.G.S.: Nested-CA: a foundation for multiscale modeling of land use and land change. Ph.D. thesis, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil (2006)
5. Lima, T.F.M., Carneiro, T.G.S., Faria, S.D.: Desenvolvimento de uma plataforma gráfica para a descrição de modelos de sistemas ambientais. In: Proceedings of the X Brazilian Symposium on GeoInformatics, pp. 121–126. Sociedade Brasileira de Computação (SBC), Brasil (2008)
6. Ierusalimschy, R., de Figueiredo, L.H., Filho, W.C.: Lua - an extensible extension language. Software Practice and Experience 26(6), 635–652 (1996)
7. Gibson, C.C., Ostrom, E., Ahn, T.K.: The concept of scale and the human dimensions of global change: a survey. Ecological Economics 32(2), 217–239 (2000)
8. Neumann, J.V.: Theory of self-reproducing automata. Mathematics of Computation 21(100) (1996)
9. Wooldridge, M.J., Jennings, N.R.: Intelligent agents: Theory and practice. Knowledge Engineering Review 10(2), 115–152 (1995)
10. Henzinger, T.A.: The theory of hybrid automata. In: Proceedings 11th Annual IEEE Symposium on Logic in Computer Science, pp. 278–292. IEEE Computer Society Press, Los Alamitos (1996)
11. Carneiro, T.G.S., Câmara, G., Maretto, R.V.: Irregular cellular spaces: Supporting realistic spatial dynamic modeling over geographical databases. In: Proceedings of the X Brazilian Symposium on GeoInformatics, pp. 109–120. Sociedade Brasileira de Computação (SBC), Brasil (2008)
12. Zeigler, B.P.: Theory of Modeling and Simulation, 1st edn. John Wiley, Chichester (1996)
13. Andrade, P.R., Monteiro, A.M.V., Câmara, G., Sandri, S.: Games on cellular spaces: How mobility affects equilibrium. Journal of Artificial Societies and Social Simulation 12(1), 5 (2009)
14. Almeida, R.M., Macau, E.E.N., França, H., Ramos, F.M., Carneiro, T.G.S.: Simulando padrões de incêndios no parque nacional das Emas, estado de Goiás, In: Proceedings of the X Brazilian Symposium on GeoInformatics, pp. 183–194. Sociedade Brasileira de Computação (SBC), Brasil (2008)

[5] http://www.terralab.ufop.br/

15. Moreira, E.G., de Aguiar, A.P.D., Costa, S.S., Câmara, G.: Spatial relations across scales in land change models. In: Proceedings of the X Brazilian Symposium on GeoInformatics, pp. 95–107. Sociedade Brasileira de Computação (SBC), Brasil (2008)
16. Pimenta, P.F., Coelho, A., Costa, S.S., Moreira, E.G., Aguiar, A.P., Câmara, G., Araújo, R., Ribeiro, A.: Land change modeling and institutional factors: heterogeneous rules of territory use in the Brazilian Amazonia. In: Proceedings of X Brazilian Symposium on GeoInformatics, pp. 81–93. Sociedade Brasileira de Computação (SBC), Brasil (2008)
17. Hannon, B.M., Ruth, M.: Dynamic Modeling (Modeling Dynamic Systems), 2nd edn. Springer, New York (2001)
18. Aguiar, A.P.D., Câmara, G., Escada, M.I.S.: Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. Ecological Modelling 209(2-4), 169–188 (2007)
19. Geist, H.J., Lambin, E.F.: Proximate causes and underlying driving forces of tropical deforestation. BioScience 52(2), 143–150 (2002)
20. Fraga, L.M., Lana, R.M., Carneiro, T.G.S., Guimarães, F.G.: Calibração em modelagem ambiental na plataforma TerraME usando algoritmos genéticos. In: Proceedings of LXII Brazilian Symposium on Operational Research (2010)
21. da Silveira, A.L.L.: Ciclo Hidrológico e Bacia Hidrográfica. In: Tucci, C.E.M. (org.) Hidrologia: Ciência e Aplicação, ch. 2, pp. 35–51. UFRGS, Porto Alegre, RS, Brasil, 3 (2003)
22. Ward, A.D., Trimble, S.W.: Environmental Hydrology, 2nd edn. CRC Press, Boca Raton (2003)
23. Carneiro, T.G.S., de Lima, T.F.M., Faria, S.D.: TerraLAB - Using free software for Earth system research and free software development. In: Proceedings of the X Workshop of Free Software, pp. 35–40. Sociedade Brasileira de Computação (SBC),Porto Alegre, RS, Brasil (2009)

Visual Quality Control of Planar Working Pieces: A Curve Based Approach Using Prototype Fitting

Georg Maier and Andreas Schindler

FORWISS, University of Passau,
Innstr. 43, 94032 Passau, Germany
{gmaier,schindler}@forwiss.uni-passau.de
http://www.forwiss.uni-passau.de

Abstract. In this paper we elucidate a method for visual quality control and vision metrology of planar objects in manufacturing. We present an efficient approach to comparing a masterpiece with an arbitrary working piece detected on a line of a factory. After extracting the contour of the object, we use prototype fitting to determine the best transformation which maps the masterpiece to the extracted contour points. In contrast to classic ICP like methods, we suggest an encoding of the masterpiece by a curve, which allows a fast computation of point-curve distances, in order to guarantee a more accurate, faster and less memory intensive performance.

Keywords: Prototype Fitting; ICP; Visual Quality Control; Shape Recognition; Arc Splines; Automated Visual Inspection.

1 Introduction

Visual quality assessment has become a growing concern for industrial manufacturing in the last decades (e.g. [1] or [2]). In almost every production process very high standards of quality are applied. Therefore, fast and highly accurate non-contact measurement systems are needed.

Many approaches to these challenges assume that the working pieces are observed by a camera or an image scanner and are planar (e.g. [3]). Then, the contours of a digital image are extracted, which results when capturing the object. Since we are situated in an industrial environment, we can take proper illumination and high accuracy of the detecting system concerning the contour point extraction as a starting point. Thus, accurate input data can be expected, possibly after a preprocessing step.

We subsequently discuss an approach to the online comparison between an arbitrary object with a *masterpiece*. A masterpiece is a specimen of a working piece with the exact shape required. Often, the corresponding CAD-layout is also called masterpiece.

As indicated above, we focus on (quasi) planar objects and we assume that the extracted contours of the currently observed object are already given. We

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 91–102, 2011.

then show how our method serves in optical quality control and vision metrology, and how it can be used for target-performance comparisons of planar geometries like laminations, panes of glass or planks of shelves.

Our procedure uses a generalization of the *iterated closest point (ICP) algorithm* algorithm (see [4]), the so-called *prototype fitting*, which was introduced in [5]. Using this strategy we determine the best transformation which maps the masterpiece to the extracted contour points of an observed working piece. Regarding the admissible transformations we focus mainly on translations, rotations and scalings. In order to guarantee a faster, more accurate and less memory-intensive approach, we suggest an encoding of the masterpiece by a curve which allows a fast computation of point to curve distances and which is composed of a preferably low number of segments.

The outline of this paper is as follows. First, we dwell on related work and motivate our algorithmic approach. Then, we describe the basic facts of prototype fitting, and we also give a brief sketch to the generation of a suitable curve description of a masterpiece. As the initial transformation is crucial for prototype fitting, Section 5 describes a stable and fast method for determining an initial guess needed for fitting the prototype to the current extracted contour points. While some test results are presented in the Section 6, we discuss further extensions and possible future work in Section 7.

2 Related Work and Motivation

Shape matching or shape registration is the basis for many computer vision techniques, such as image segmentation and pose estimation with applications in object recognition and quality assurance tasks. Therefore, several publications on shape matching can be found in the literature. A survey and a short summary can be found in [6] or [7]. Most of these approaches rely on classic explicit shape representations given by points, which are possibly connected by lines or other types of curve segments in order to form a shape. Different approaches as representing data sets by zero sets of *implicit polynomials* (e.g. [8]) are not suitable for the described application since the degree of the implicit polynomial must be determined before handling a fitting procedure and they often result in deterioration in local accuracy.

The most common method working on explicit shape representations is the ICP algorithm (cf. [4]): Given two shapes and an error metric, the task is to find a mapping in an admissible class of transformations which leads to the minimum distance between the two shapes. The ICP algorithm is then searching for an optimal rigid motion $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ matching a point set C to another set P as follows: For each point $y \in C$ the closest point $x_y \in P$ is calculated. Then, the optimal transformation Φ' is determined, that minimizes the sum of squared distances $\|\Phi'(y) - x_y\|^2$ between pairs of closest points y, x_y. Having this transformation applied to the point set C, the three steps explained above are repeated until the algorithm converges. The convergence of this algorithm is ensured to the next local minimum of the sum of squared distances between

closest points. Hence a good initial estimate is required to ensure reaching the sought solution. In order to improve the rate of convergence and to match partially overlapping point sets, several variants of the ICP algorithm have been developed in the last decades (cf. [9]).

For instance, in [5] generalizations to broader classes of admissible transformations and algorithmic improvements have been made and the corresponding theoretical bounds and convergence behaviors are analyzed. The reference geometry, namely the masterpiece, is encoded there as a compact subset $P \subset \mathbb{R}^2$, e.g. as a union of boundary curves, and it is called *prototype*. If points $y \in C$ have been extracted while a certain measurement process, the *prototype fitting problem* is the challenge to find a feasible transformation Φ minimizing

$$\sum_{y \in C} \operatorname{dist}(\Phi(P), y)^2,$$

where dist denotes the euclidean distance.

In general, the existence of such an optimal mapping can only be assured when enlarging the set of feasible mappings in a suitable way. A more detailed description can be found in [5]. In particular, it is important to transform the prototype and not the point set C as practically all other ICP like methods do. We will take a closer look to this strategy in Section 3.

The advantages of ICP and prototype fitting algorithms are obvious: They are easy to implement and they provide good results if a satisfying initial solution has been provided. Regardless the explicit method applied, the determination of closest points $x_y \in P$ for each $y \in C$ is the bottleneck of the computational time. Therefore, we here focus on a suitable encoding of P and give a strategy for providing a good initial guess.

3 Prototype Fitting

When we consider an arbitrary working piece, its position may not be the same as the location of the masterpiece, i.e. the observed object might be rotated, translated and even scaled.

Let the prototype be denoted by $P \subset \mathbb{R}^2$. If $y_1, \ldots, y_n$ are the available contour points of an observed working piece, we search for a motion $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ minimizing the sum of squares $\sum_{i=1}^{n} \operatorname{dist}(\Phi(P), y_i)^2$. Naturally, the existence of such optimal motions can only be assured if we make some restrictions and assumptions on the feasible transformations. Here, we focus on motions that consist of translations, isotropic scalings and rotations, which are mappings of the form

$$T : \mathbb{R}^2 \to \mathbb{R}^2,\ x \mapsto \lambda \cdot \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \cdot x + \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix},$$

where $\varphi \in [0, 2\pi[$ and $\lambda, \vartheta_1, \vartheta_2 \in \mathbb{R}$. We denote the set of all these mappings by $\mathcal{T}$ and note that these functions can also be written in the form

$$T_{c,s,t} : \mathbb{R}^2 \to \mathbb{R}^2,\ x \mapsto A_{c,s} \cdot x + t$$

with scaled rotation

$$A_{c,s} := \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \text{ for all } c, s, \in \mathbb{R}.$$

and translation $t \in \mathbb{R}^2$. Hence we search for optimal parameters $c, s \in \mathbb{R}$ and $t \in \mathbb{R}^2$.

Note that even for a larger class of admissible transformations, the pose estimation regarding rotation, scaling and translation is important to have a first match. Afterwards, possibly more sophisticated transformations, like non-isotropic scalings, projections or spline deformations can be taken into account. Thus, focusing on the class $\mathcal{T}$ is sufficient within this scope. Then, the problem formulated above can be solved very fast by an iterative approach.

To begin with, let us assume that some initial transformation parameter c_0, s_0, t_0 are known such that the transformed masterpiece approximately fits to the contour points. In order to obtain such an initialization we refer to Section 5.

In any case, after having found an initial transformation, we can compute the best approximating points x_i of y_i with respect to the set $P^{(1)} := T_{c_0,s_0,t_0}(P)$ and proceed as follows:

Using the abbreviation

$$\widetilde{x_i} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x_i \text{ for all } i = 1, \ldots, n$$

and denoting the barycenters of $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ by

$$\mu_x := \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \mu_y := \frac{1}{n} \sum_{i=1}^{n} y_i,$$

the optimal values $s_1, c_1 \in \mathbb{R}$ and $t_1 \in \mathbb{R}^2$, i.e.

$$\sum_{i=1}^{n} \|T_{c_1,s_1,t_1}(x_i) - y_i\|^2 = \min_{s,c \in \mathbb{R}, t \in \mathbb{R}^2} \sum_{i=1}^{n} \|T_{c,s,t}(x_i) - y_i\|^2$$

can be derived in a closed form:

$$c_1 = \frac{1}{\rho} \sum_{i=1}^{n} (x_i - \mu_x)^T y_i, \;\; s_1 = \frac{1}{\rho} \sum_{i=1}^{n} (\widetilde{x_i} - \mu_x)^T y_i,$$

where $\rho = \sum_{i=1}^{n} \|x_i - \mu_x\|^2$, and the optimal translation is given by $t_1 = \mu_y - A_{c_1,s_1}\mu_x$. Then, the value $E := \frac{1}{n} \sum_{i=1}^{n} \|T_{c_1,s_1,t_1}(x_i) - y_i\|^2$ indicates the fitting quality. Again, we can compute the best approximating points $x_i^{(2)}$ of y_i with respect to

$$T_{c_1,s_1,t_1}(P^{(1)}) =: P^{(2)}$$

and solve the least squares problem as above. We carry on with this alternating procedure while E is greater than a given threshold E_0 or the difference between

the predecessor error and the current error is not too small. Assuming that the algorithm has terminated in the k-th iteration step, we can additionally check if $\mathrm{dist}(y_i, P^{(k)}) > \varepsilon$ for some $i = 1, \ldots, n$ and $\varepsilon > 0$ in order to guarantee a good dimension accuracy satisfying the quality of the product. Alternatively, also the Hausdorff distance between $P^{(k)}$ and the compact set C including all y_i can be applied. In any case, if E is still larger than a given tolerance, the object doesn't satisfy the quality requirements and has to be sorted out. As a matter of course, the parameters and thresholds depend very much on the real application itself and can't be discussed within this scope.

4 Prototype Encoding

In order to achieve an efficient computation of this iterative method it is decisive that the calculation of the best approximating points with respect to P is very fast. As already discovered in [4] and [5] the necessity of a fast determination of best approximating points doesn't depend on a special choice of the optimization method but is also crucial when using any other non-linear optimization solver, like Gauß-Newton or Levenberg-Marquardt (cf. [10]).

Indeed, the efficiency of the calculation of closest points depends on the encoding of the prototype P. Therefore, we suggest a description of P as a union of curves having a preferably low number of segments and providing a fast calculation of best approximating points. Furthermore, high flexibility for modeling the desired geometric pattern is needed. Since almost all ICP methods are based on point encodings of the prototype, sophisticated point selection approaches are needed to achieve efficiency improvements. Obviously, a curve representation, as described above, has considerable advantages over these techniques regarding accuracy, time and storage space.

One possible choice of such a curve, fulfilling these criteria, is an *arc spline* which is a composition of circular arcs and line segments $\gamma := \gamma_1 \ldots \gamma_N$. Since arc splines are determined by only a few parameters and satisfy important invariance criteria, like invariance with respect to rotations, scalings and translations, they can be applied well to measurement tasks, as shown in [11] or [12]. The best approximating point x_i of y_i to the nearest circular segment γ_j can be calculated by intersecting γ_j with the line segment defined by y_i and the center of the circle corresponding to γ_j, and it is therefore very cheap in comparison to parametric curves, which need iterative strategies. If the nearest segment γ_j is a line segment, we only have to compute the orthogonal projection of y_i on γ_j.The correspondence between a point y_i and a segment γ_j can be quickly and efficiently determined if the number of segments N is low. For instance, by a (transformed) tree structure like a *quadtree decomposition* (e.g. [13]), which can be established beforehand, the assignment to the corresponding segment can be improved. Above all, smoothness at the breakpoints can be required in order to enable a realistic modeling.

However, the question remains how to get a representation of the masterpiece in form of an arc spline composed of as few as possible segments. In [11] a

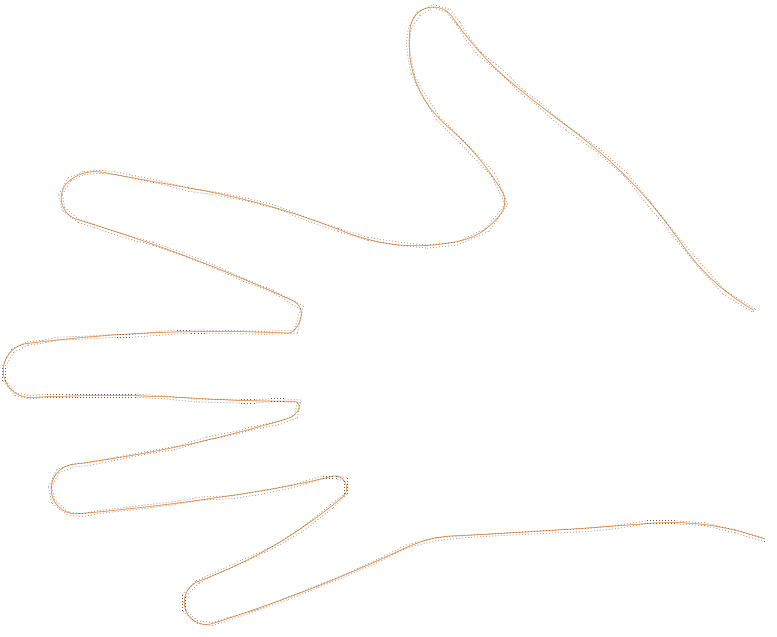


Fig. 1. SMAP-Approximation of the shape of a hand. The dashed lines indicate the corresponding tolerance channel.

method is proposed for the computation of a smooth arc spline approximating a sequence of points guaranteeing a minimum number of segments as well as a freely definable approximation quality given by a so-called tolerance channel along the points. Due to the relationship of this problem to the *minimum link problem* the resulting arc spline is called *smooth minimum arc path*, abbreviated by *SMAP*. In addition, it is possible to integrate preliminary knowledge by inserting automatically or manually vertices and line segments without losing the minimal number of segments. If we can completely pass on smoothness, [11] also presents an algorithm for generating a *continuous minimum arc path (CMAP)*, which is a continuous arc spline encoding the prototype with a minimum number of segments regarding a given maximum error.

Though these algorithms guarantee the minimally possible number of segments with respect to any accuracy, they don't satisfy real time requirements. In fact, they have a quadratic worst case complexity with respect to the number of input points n, but perform in $O(Nn)$ in the most practical applications. In any case, as the prototype can be generated off-line, the computational time doesn't play an important role.

Figure 1 shows an example of the shape of a hand captured with a resolution of 640x480 pixels enclosed by a tolerance channel with a width of 0.5 pixel. A corresponding SMAP (orange) needs 35 segments. Note that in general an

infinite number of SMAPs can exist regarding the same tolerance error but the minimal possible number is always unique. In comparison, a CMAP needs 28 segments for the displayed example.

5 Initial Solution

As in every nonlinear optimization problem, a good start value is crucial for the performance of the solving method. Hence we deal in this section with the problem of estimating an initial transformation, which provides a suitable first match of the prototype and the observed object.

If such an initial transformation can't be determined by preliminary knowledge, we suggest the following strategy in case of a translated, rotated and possibly scaled object: An initial translation t_0 is due to the difference of the barycenter of the extracted points and the barycenter b of the prototype, i.e.:

$$t_0 := \frac{1}{n} \sum_{i=1}^{n} y_i - b.$$

For instance, if the masterpiece is described by an arc spline $\gamma := \gamma_1 \cdots \gamma_N$, the barycenter is given by

$$b = \frac{1}{l} \sum_{i=1}^{N} l_i \cdot b_i,$$

where l_i is the arc length, $l := \sum_i l_i$ and b_i is the barycenter of γ_i. The arc length l_i can be computed easily using the opening angle and the radius and the distance of the start and end point in case of a line, respectively.

Then, it is reasonable to compute an initial rotation R_0 by the angle between the principal axes of the prototype and the contour points. The calculation of the principal direction of a set of points meets calculating the best approximating line. Likewise, we can approximately compute the principal direction of the curve describing the masterpiece by the best fitting line of some sampling points of the curve. In case of an arc spline, there even exists a closed form solution. As generally known, the computation of a best approximating line is a eigenvalue problem of a symmetric positive definite 2×2 matrix, which can be solved very fast. However, if the two eigenvalues are equal, the problem sketched above can't be solved uniquely. In this case, another strategy based on preliminary knowledge has to be taken into account. In fact, an ambiguity of 180 degrees is possible, as we obtain only lines, not directions, but using the fitting error the wrong direction can be identified easily.

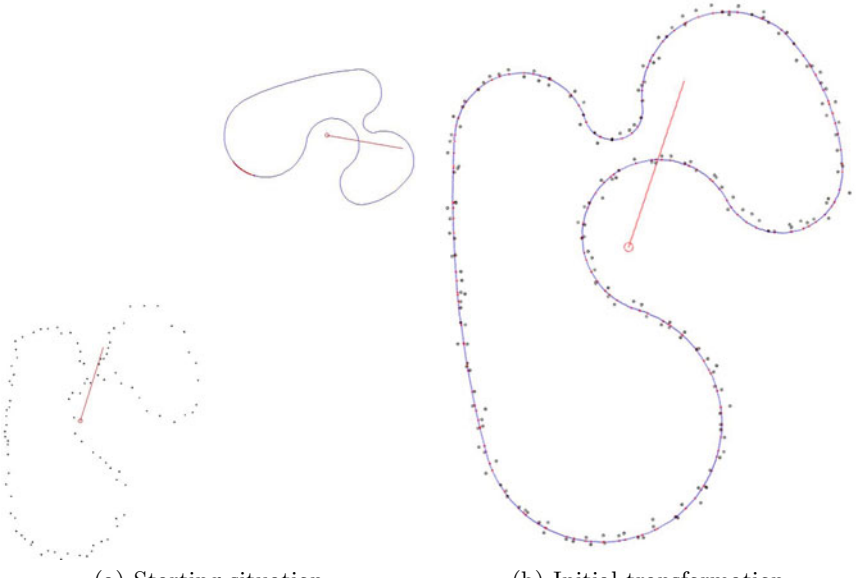
The root mean square deviation δ_C of the point set $C := \{y_1, \ldots, y_n\}$ and the corresponding barycenter μ_C is given by

$$\delta_C := \sqrt{\sum_{i=1}^{n} \operatorname{dist}(y_i, \mu_C)^2}$$

Again, the likewise value δ_P of the prototype P can be obtained by the sampling points on the curve or can be computed in a closed form in case of an arc spline. Thus, a suitable choice for an initial scaling factor λ_0 is the ratio

$$\lambda_0 := \frac{\delta_C}{\delta_P}.$$

The whole proceeding for determining an initial transformation is illustrated in Figure 2. Our method provides a satisfying starting solution such that the whole iterative strategy works very efficiently and converges to the global minimum of the error function.



(a) Starting situation (b) Initial transformation

Fig. 2. Computation of an initial transformation as described above. Black: Contour Points; blue: Prototype; red: Corresponding barycenters and principal directions; on the prototype: Best-approximating points.

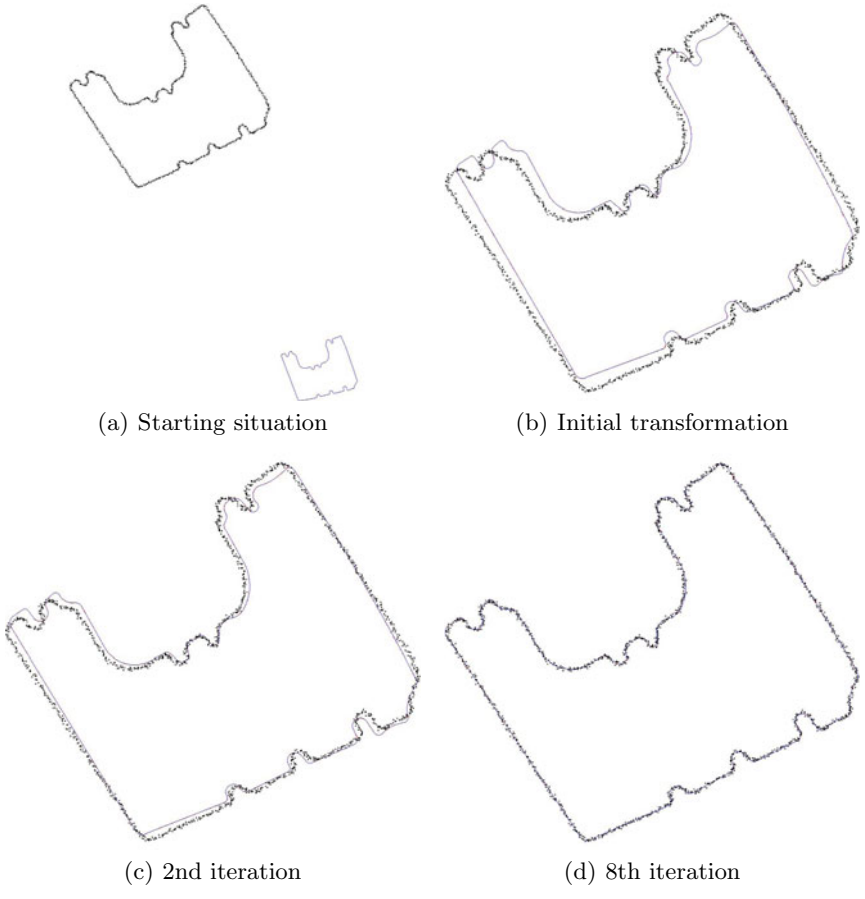
Besides, note that this approach can also be used for shape recognition tasks. For instance, the requirements formulated in [14] are also satisfied this way. Our method can be applied the same way to this problem if the reference object, which has to be detected, has a suitable curve encoding.

6 Results

To illustrate the performance of our method we consider the quality control of a window frame cross-section. A prototypical masterpiece in form of an arc

spline has been generated. Then the contour points of a working piece have been extracted. Additionally, we applied Gaussian noise of different deviations to the points.

Figure 3 shows some iterations of our prototype fitting method.



(a) Starting situation (b) Initial transformation

(c) 2nd iteration (d) 8th iteration

Fig. 3. Different iteration steps for the prototype fitting

In our test scenarios we applied a wide range of changes in rotation, translation, scaling and added Gaussian noise of a deviation up to a third of the pieces diameter. As the prototype fitting is based on a least-square optimization, the working piece could be fitted in all cases even under addition of extreme noise. Figure 4 shows the fitting error as a function of the number of iterations for the example considered in Figure 3. After the initial transformation the fitting error decreases consecutively. Figure 5 shows a plot of the fitting error as a function of additive Gaussian noise. As expected, the prototype fitting error grows linearly with the value of the added Gaussian noise.

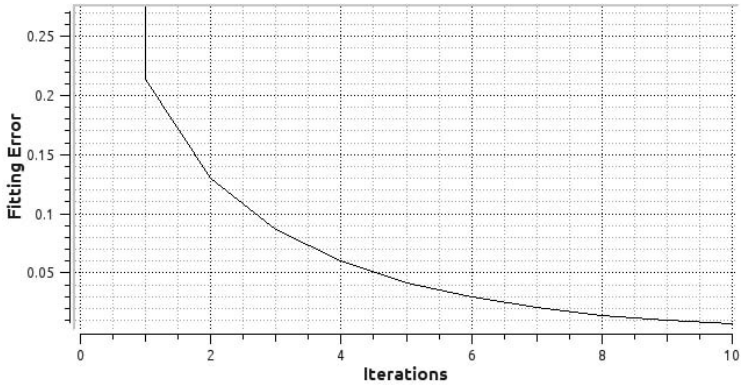


Fig. 4. Number of iterations against fitting error

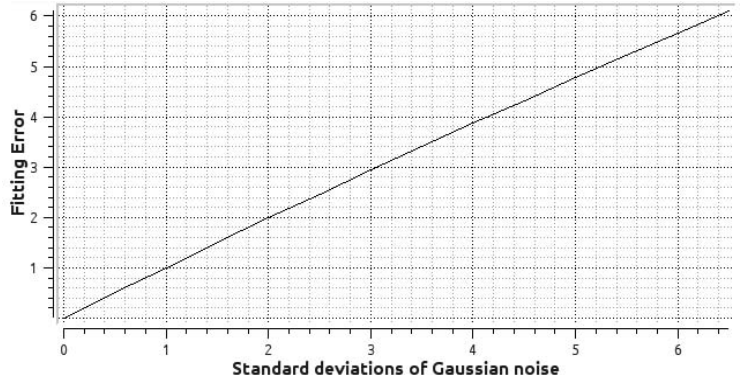


Fig. 5. Standard deviation of added Gaussian noise against fitting error

As a matter of course, if the points are too noisy, the resulting maximum error is quite large. As we have already indicated, the maximum error is also an important quantity in practical applications in order to detect local shape deviations.

Though a quantitative comparison of our method with traditional approaches is hardly possible, we obviously achieve advantages regarding timing and accuracy. Point based methods have to deal with a considerably higher number of points than segments needed in our approach in order to have an acceptable accuracy. Thus, solving point correspondences is much slower than computing the best-approximating points. This problem reduces to a logarithmic scale when using quadtrees but is still considerable for relatively complex working pieces. Contrariwise, reducing the number of points for describing the prototype would result in a worse accuracy. In any case, encoding the prototype by a discrete point set cannot be as accurate and realistic as using a curve encoding.

Therefore, corresponding best-approximating points of the curve representation are generally better than those taken from a point cloud.

7 Conclusions

We have presented an efficient approach to visual quality control of planar working pieces in manufacturing industry. Based on prototype fitting techniques we have developed a fast and stable method which is very suitable for target-performance comparisons.

In particular, we introduced an encoding of the prototype based on a curve description allowing a very fast calculation of closest points, and we presented an efficient method for estimating an initial value for determining an optimal transformation, which matches a prototype to an arbitrary point set. In comparison to state of the art methodologies, we could improve the accuracy, the run time and the space complexity.

Nevertheless, there are some situations, in which we might even improve the performance of our algorithmic approach. Thus, we conclude by mentioning a few research avenues for future work which seem to be particularly important and promising.

As indicated in Section 5, the proposed determination of the initial rotation can be ill-conditioned in a few cases, when the eigenvalues are almost equal. In this case, we surly may keep on improving our strategy. Taking into account the Fourier transformation of the curvature characteristics is an imaginable idea.

Another promising extension of our procedure would be considering the barycenters and principal directions of the convex hull of the points or of the polygon resulting when connecting adjacent points. In particular, this probably yields better results if the contour points have not been extracted homogeneously.

References

1. Chin, R.T., Harlow, C.A.: Automated visual inspection: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4, 557–573 (1982)
2. Hocenski, Z., Baumgartner, A.: Image comparison method for visual quality control based on matrix decomposition. In: Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE 2000), vol. 2, pp. 580–585 (2000)
3. Hong, D., et al.: A prototype indexing approach to 2D-object description and recognition. Pattern Recognition 31, 699–725 (1998)
4. Besl, P., McKay, H.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
5. Donner, K.: Image interpretation based on local transform characterization. Pattern Recognition and Image Analysis 7 (1997)
6. Veltkamp, R.C., Hagedoorn, M.: State-of-the-art in shape matching. Technical report, Principles of Visual Information Retrieval (1999)
7. Rosenhahn, B., Brox, T., Cremers, D., Seidel, H.-P.: A comparison of shape matching methods for contour based pose estimation. In: Reulke, R., Eckardt, U., Flach, B., Knauer, U., Polthier, K. (eds.) IWCIA 2006. LNCS, vol. 4040, pp. 263–276. Springer, Heidelberg (2006)

8. Zheng, B., Takamatsu, J., Ikeuchi, K.: An adaptive and stable method for fitting implicit polynomial curves and surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 561–568 (2010)
9. Rusinkiewicz, S., Levoy, M.: Efficient Variants of the ICP Algorithm. In: Proceedings of the Third Intl. Conf. on 3D Digital Imaging and Modeling, pp. 145–152 (2001)
10. Nocedal, J., Wright, S.J.: Numerical Opitimization. Springer, Heidelberg (1999)
11. Maier, G.: Smooth Minimum Arc Paths. Contour Approximation with Smooth Arc Splines. Shaker, Aachen (2010)
12. Aichholzer, O., et al.: Computational and structural advantages of circular boundary representation. Int'l Journal of Computational Geometry & Applications (in press 2011)
13. de Berg, M., et al.: Computational Geometry. Springer, New York (2000)
14. Junding, S., Heli, X.: Contour-shape recognition and retrieval based on chain code. In: CIS 2009: Proceedings of the 2009 International Conference on Computational Intelligence and Security, pp. 349–352. IEEE Computer Society, Washington, DC, USA (2009)

New Approaches for Model Generation and Analysis for Wire Rope

Cengiz Erdönmez[1] and Cevat Erdem İmrak[2]

[1] Istanbul Technical University, Institute of Informatics, Computational Science and Engineering Program, 34469 Maslak, Istanbul, Turkey
[2] Istanbul Technical University, Faculty of Mechanical Engineering, Mechanical Engineering Department, Gümüşsuyu, 34437, Istanbul, Turkey

Abstract. Independent wire rope cores are composed by helically wrapping a wire strand over a straight wire strand. Outer strand of the wire rope is composed with nested helical geometry which is difficult to model for analysis. In this paper a wire by wire based, a more realistic analysis model determination of an independent wire rope core is defined with the parametric equations of the nested helical geometry. The locations of the single and nested helical wires are created and the meshed model of each wire is constructed separately. Wire rope is assembled and the axial loading model is constructed and analyzed using finite element analysis. The obtained numerical results are compared with the theoretical results. The results have in good agreement and the wire by wire analysis gives insight about the wire loads acting within an independent wire rope core.

Keywords: Wire strand, wire rope, independent wire rope core, wire rope modeling.

1 Introduction

Wire rope analysis mostly relies on the classical treatise on the mathematical theory of elasticity of Love [1]. Nonlinear equilibrium equations are solved analytically by a number of researchers and theoretical models are developed in different aspects [2]-[9].

While solving the equilibrium equations of a simple straight strand using the theory of Love, frictionless theory is mostly used due to the difficulties of the equilibrium equations and geometric considerations of the helical structure. Costello has brought together his comprehensive studies over wire rope theory in his reference book [3]. Theoretical investigations of wire rope theory based on the analysis of a simple straight strand. Static behavior of a simple straight strand is investigated in [2]. For more complex geometries such as wire ropes, theoretical models are applied using some homogenization hypothesis.

To predict the axial static response, a Seale IWRC is analyzed in [7]-[8]. A design methodology for multi-lay wire strands is investigated in [9]. Among them the analysis of IWRC in [4] is remarkable which represents excellent correlation with the available experimental results in the literature.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 103–111, 2011.

In finite element models given in the literature, a basic sector of a simple straight strand is studied at first in [10]. Then it is developed to a three-layered straight helical wire strand in [11]. A new theoretical model simulating the mechanical response of a wire rope with an IWRC which fully consider the double-helical configuration of individual wires is investigated in [12]. Also the advantage of wire-by-wire modeling approach based on the general rod theory is introduced and compared with the fiber models in [13].

Lately modeling difficulties of double helical wires is investigated in [14]. Some of the encountered modeling problems are solved by the referred modeling scheme by Erdönmez and İmrak. Analysis of an IWRC including plastic deformations is investigated in [15]. The robustness of the generated solid model is mentioned.

Independent wire rope core (IWRC) is a special type of wire rope. It is composed by using a straight wire strand as a core, wrapped by six outer strands as shown in Fig. 1. The core strand of the IWRC is composed by (1+6) wires. The core wire of the core strand is a straight wire and the outer 6 wires are single helical wires. In the composition of the outer strand core wire of the outer strand is a single helical shape while the outer wires of the outer strand is in nested helical form as shown in Fig. 1. IWRC has been used in a variety of applications and also become a core strand for some of the wire ropes, such as Seale IWRC and Warrington IWRC. Large tensile force strength of the wire ropes is very important in application areas where as the small bending and torsional stiffness.

In this paper first of all a nested helical geometry definition is described. Then modeling of an IWRC with using the nested helical geometry is presented. Using the proposed geometrical model a numerical solution of the IWRC is presented and compared with the theoretical results.

2 New Approach for Wire Rope Geometry

2.1 Definition of the Nested Helical Geometry

An IWRC is composed by a core strand wrapped by six outer strands as shown in Fig. 1. Due to the complexity of the outer wires of the outer strand of the IWRC, it is difficult to model in 3-D form the outside nested helical geometries via using the computer aided design software. Even the generated models via CAD softwares have problematic surfaces and it is not coherent to analyze such problematic body using finite element analysis FEA codes such as Abaqus™. To overcome this, a new procedure for modeling wire rope is proposed using the mathematical definition of the nested helical structure. In this paper first of all a single helical wire composition will be described and then IWRC definition will be given.

The general construction of a simple straight strand and an IWRC is shown in Fig. 1 and Fig. 2. Straight core strand is composed by a straight center wire of radius R_1 surrounded by six single helical wires of radius R_2 around it. Center wire raius of the outer strand is given by R_3 and the nested helical wire radius is given by R_4 as shown in Fig. 2. The relation of the helix angle and the pitch length can be given with,

$$\alpha_i = \arctan\left(p_i / \left(2\pi r_i\right)\right), \tag{1}$$

where p_i is the pitch length of the helical wire and r_i is the helix radius for the *i*'th wire. To define the location of a single helix centerline, Cartesian coordinate system (X, Y, Z) is used with the Cartesian frame $\{e_X, e_Y, e_Z\}$ and the location along the centerline of a single helix is defined as,

$$\begin{aligned} X_s &= r_s \cos(\theta_s), \\ Y_s &= r_s \sin(\theta_s), \\ Z_s &= r_s \tan(\alpha_s)\theta_s, \end{aligned} \tag{2}$$

where e_Z is the rope axis, r_s is the radius of the single helix, α_s is the single helix laying angle and $\theta_s = \theta_0 + \theta$. Free angle θ is used to define the location of the wire around the rope axis e_Z, relative to e_X. Single helix phase angle is defined by $\theta_0 = \theta_{(z=0)}$. The outer double helical wires are wound around a single helical wire by using the location along the centerline of a single helix given in equation (2) and the location of the nested helices can be defined as [12],

$$\begin{aligned} X_d &= X_s(\theta_s) + r_d \cos(\theta_d)\cos(\theta_s) - r_d \sin(\theta_d)\sin(\theta_s)\sin(\alpha_s), \\ Y_d &= Y_s(\theta_s) + r_d \cos(\theta_d)\sin(\theta_s) + r_d \sin(\theta_d)\cos(\theta_s)\sin(\alpha_s), \\ Z_d &= Z_s - r_d \sin(\theta_d)\cos(\alpha_s), \end{aligned} \tag{3}$$

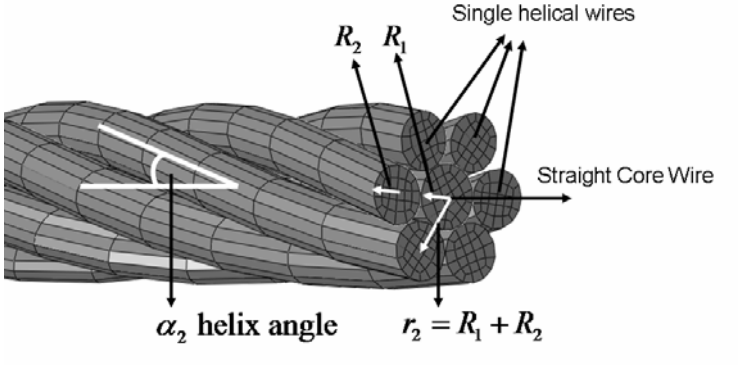


Fig. 1. A simple straight strand geometry

where $\theta_d = m\theta_s + \theta_{d0}$ and r_d is the distance along the nested helix wire centerline and single helix strand centerline shown in Fig. 2. The construction parameter m is a constant value that can be estimated by $m = h_s / (h_w \cos\alpha_s)$, where h_s and h_w are the lay lengths of the outer strands and outer wires of the outer strands respectively [13] and θ_{d0} is the wire phase angle. According to (1)-(3) a right lang lay wire rope structure can be constructed and to construct a left lang lay wire rope, it is enough to negate one of the coordinate values of X_d, Y_d or Z_d given in equation (3).

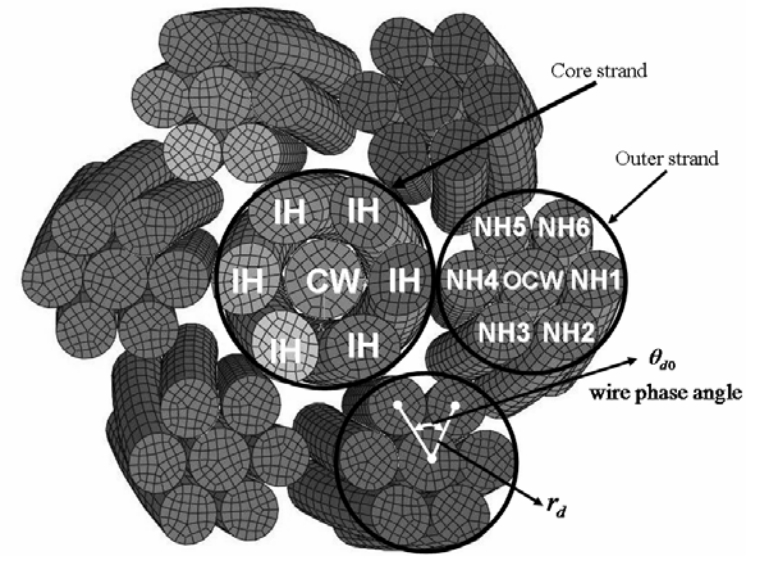


Fig. 2. Geometry and wires within an IWRC

2.2 Wire Rope Generation and Modeling

Due to the complex wire geometry, nested helical wires over the outer strand of the IWRC is not in the scope of the commercial solid modeling softwares for the moment to model a nested helical wire. To model a nested helical wire a code was developed by the authors in Matlab™. Using this code each wire centerline is created using the control nodes. Control nodes are generated by using the parametric equations of the both single and nested helical structures. To compose the simple straight strand, a straight center wire and a single helical wire is enough. The single helical wire will be wrapped by the six single helical wires which are identical. This procedure completes the construction issue of the simple straight strand.

IWRC geometry is composed by a straight strand and with six identical outer strands. In Fig. 2 the combination of the wires are shown. The core strand has two titles as CW for core wire and IH for six identical inner helical wires. For the outer

strand each wire has been titled with different code. OCW stands for outer center wire which is a single helical shape while the other six different nested helical wires are titled as NH1-NH6.

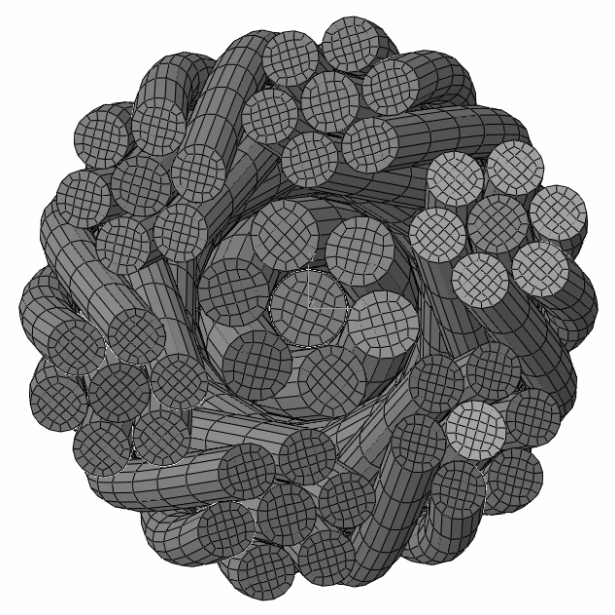


Fig. 3. Meshed form of a Right Regular Lay (RRL) IWRC

Each of the nested helical wire has to be generated separately to compose the outer strand of the IWRC. Once these wires are constructed, all of the wires are imported in Abaqus/CAE™ and assembled to complete the IWRC.

In the generated Matlab™ code each nested helical wire control nodes are created separately. After wire control nodes are imported in a modeling and meshing software HyperMesh™. Using HyperMesh™ each wire is constructed with the meshed form and exported to the analysis software Abaqus/CAE™.

Each wire is assembled and the complete meshed form of the IWRC is generated in Abaqus/CAE™. There are two benefits of this procedure. First one, the geometry is ready to analyze, no need to further meshing of the geometry and the geometry is error free. Second one is that the proposed method has no length limit.

It is possible to produce longer parts using this procedure. Using the generated wire rope geometry it is possible to analyze the model and find out the wire by wire behavior of the IWRC easily. A meshed form of a right regular lay IWRC is presented in Fig.3 by using Eqn. 3. It can be analyzed by applying the necessary boundary conditions.

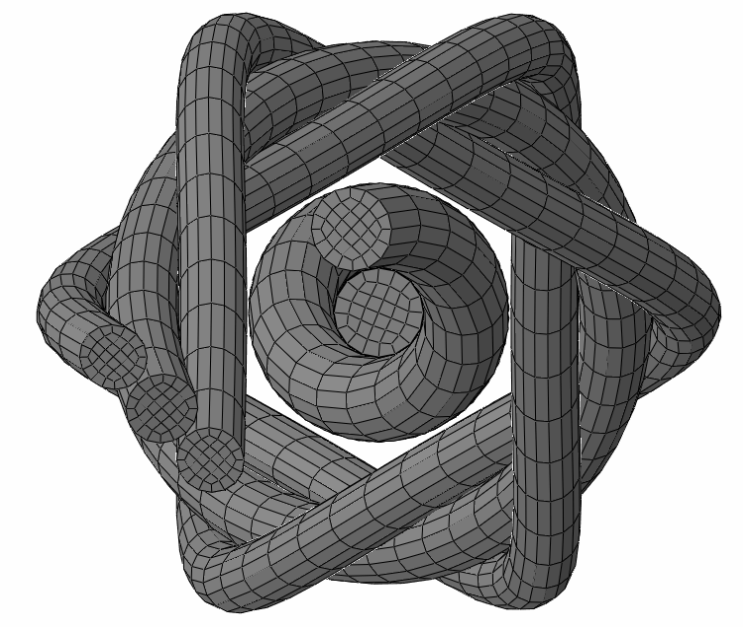


Fig. 4. Indentations over the double helical wires are presented for a Right Regular Lay (RRL) IWRC

2.3 Wire by Wire Analysis of the IWRC

The constructed IWRC is analyzed under the axial loading conditions using the commercial analysis code Abaqus™. The geometrical parameters of the wire rope are given in Table 1 [13].

Table 1. Geometric Parameters of IWRC

Parameter	Description	Value
d	IWRC diameter	29.80mm
R_1	Core strand center wire diameter	3.94mm
R_2	Core strand outer wire diameter	3.73mm
R_3	Outer strand center wire diameter	3.20mm
R_4	Outer strand double helical wire diameter	3.00mm
h	IWRC length used in the model	18mm
p_2	Pitch length of the inner helix	70mm
p_3	Pitch length of the outer center single helical wire	193mm
p_4	Pitch length of the outer nested helical wire	70mm
α_2	Helix angle	71.01^o
α_3	Helix angle	74.45^o
α_4	Helix angle	71.46^o

One side of the model is constraint with encastre boundary condition and the other end is constraint to be non rotating as $\Theta = 0$. Material properties are defined as follows; Young's modulus of 188000 N/mm^2, plastic modulus of 24600 N/mm^2, yield stress of 1540 N/mm^2, limit stress of 1800 N/mm^2, Poisson's ratio of 0.3, friction of 0.115.

Surface to surface contact controls defined for the interactions of the each wire with the neighbor wires. The analytical results obtained by the theory of Costello and finite element results are presented and compared in Table 2.

From the results it can be seen that the analysis results shows good agreement with the theory for the frictionless case. In frictional case the plastic behavior of the wire rope has been seen.

In Table 3, wire by wire forces of the IWRC have been presented. From the results, it can be concluded that core strand carries the maximum load within the strands. Within the outer wires of the outer strand NH4 in average carries the maximum load and NH5 and NH3 follows it.

From Fig. 2 the location of the NH4 can be seen and it shows that the wires close to the center strand carry more loads that the other wires in the outer strand. This information may help while optimization processes of the wire rope production.

Table 2. IWRC Force Results [N]

Strain 10^{-3} (mm)	Theory of Costello	Frictionless FEA	Frictional FEA
1	53852.46	53733	53549
2	107704.92	107392	107002
3	161557.38	160977	160360
4	215409.83	214487	213623
5	269262.29	267924	266792
6	323114.75	321285	319863
7	376967.21	374572	372579
8	430819.67	427784	423340
9	484672.13	480922	466319
10	538524.59	533984	494556
11	592377.04	586971	512383
12	646229.50	639882	525727
13	700081.96	692776	536805
14	753934.42	745551	546611
15	807786.88	798252	555685

Table 3. Wire by Wire Forces of the IWRC [N]

Strain 10^{-3} (mm)	CW	IH	OCW	NH1	NH2	NH3	NH4	NH5	NH6
0.8	1761	1202	952	712	742	826	882	845	760
1.5	3519	2402	1903	1423	1483	1650	1762	1689	1518
3.0	7030	4801	3803	2845	2965	3297	3522	3376	3035
4.5	10523	7189	5694	4260	4439	4937	5272	5054	4545
6.0	13998	9566	7576	5671	5907	6568	7014	6725	6049
7.5	17486	11829	9470	7089	7382	8208	8766	8405	7562
9.0	19182	13432	11177	8391	8682	9591	10381	9973	8966
10.5	19754	14406	12104	9336	9534	10306	10943	10776	9929
12.0	20291	15072	12560	9919	10042	10683	11205	11173	10488
13.5	20819	15585	12884	10328	10404	10959	11458	11464	10876
15.0	21335	16020	13172	10654	10701	11207	11708	11726	11185

3 Conclusion

Due to its complex geometry nested helical wires in an IWRC needs special handling while solid modeling. In this paper an IWRC is modeled taking into account the complex nature of the outer nested helical wires within the outer strand of the IWRC. This issue enables one to obtain results on wire by wire bases. A wire by wire analysis of an IWRC has been modeled and analyzed. First of all wire by wire modeling procedure using the nested helical structure of the IWRC has been described. Then the finite element analysis is constructed with the given boundary conditions. An example of the axial loading of an IWRC has been analyzed. Both theoretical and finite element analysis shows good agreement. Using the wire by wire analysis results, loads acting on each wire of the IWRC has been reported. Results show that the center wire of the core strand has been loaded with the maximum force. From the result it can be seen that the outer nested helical wires close to the center strand faced with more load than the other wires in the outer strand.

References

1. Love, A.E.H.: A treatise on the mathematical theory of elasticity, 4th edn., ch. XVIII-XIX, pp. 381–426. Dover Publications, New York (1944)
2. Costello, G.A., Sinha, S.K.: Static Behaviour of Wire Rope. Proceedings ASCE, Journal of Engineering Mechanical Division 103(EM6) 103, 1011–1022 (1977)
3. Costello, G.A.: Theory of wire rope. Springer, Heidelberg (1990)
4. Phillips, J.W., Costello, G.A.: Analysis of wire ropes with internal-wire-rope cores. Transactions of the ASME 52, 510–516 (1985)

5. Jolicoeur, C., Cardou, A.: A numerical Comparison of current mathematical models of Twisted wire cables under axisymmetric loads. Journal of Energy Resources Technology 113, 241–249 (1991)
6. Velinsky, S.A.: On the design of wire rope. Transactions of the ASME, Journal of Mechanics, Transmissions, and Automation in Design 111, 382–388 (1989)
7. Velinsky, S.A., Anderson, G.L., Costello, G.: Wire rope with complex cross sections. Journal of Engineering Mechanics 110(3), 380–391 (1984)
8. Velinsky, S.A.: General nonlinear theory for complex wire ropes. International Journal of Mechanical Science 27, 497–507 (1985)
9. Velinsky, S.A.: On the design of wire rope. Transactions of the ASME, Journal of Mechanics, Transmissions, and Automation in Design 111, 382–388 (1989)
10. Jiang, W.G., Yao, M.S., Walton, J.M.: A concise finite element model for simple straight wire rope strand. Int. Journal of Mechanical Sciences 41, 143–161 (1999)
11. Jiang, W.G., et al.: A concise finite element model for three-layered straight wire rope strand. International Journal of Mechanical Sciences 42, 63–86 (2000)
12. Elata, D., Eshkenazy, R., Weiss, M.P.: The mechanical behavior of a wire rope with an independent wire rope core. Int. Journal of Solids and Structures 41, 1157–1172 (2004)
13. Usabiaga, H., Pagalday, J.M.: Analytical procedure for modelling recursively and wire by wire stranded ropes subjected to traction and torsion loads. International Journal of Solids and Structures 45(21), 5503–5520 (2008)
14. Erdönmez, C., İmrak, C.E.: Modeling Techniques of Nested Helical Structure Based Geometry for Numerical Analysis. Strojniški vestnik - Journal of Mechanical Engineering (2011)
15. İmrak, C.E., Erdönmez, C.: On the problem of wire rope model generation with axial loading. Mathematical and Computational Applications 15(2), 259–268 (2010)

High-Quality Real-Time Simulation of a Turbulent Flame

Piotr Opioła

Institute of Computer Science, Jagiellonian University
Łojasiewicza 6, 30-348 Kraków, Poland

Abstract. We consider a 2-dimensional model of a turbulent, methane flame. The model is based on the compressible Navier-Stokes equations extended with a temperature equation and chemical reaction properties. In order to achieve a high-quality real-time simulation, a fully adaptive finite element method is used to solve the considered system of equations. The method performs adaptive mesh refinement, local adjustment of the approximation order, and multilevel adaptation. The structure and composition of the flame, along with the numerical properties of the method, such as the mesh density, are studied. The results are compared to results obtained with a direct numerical simulation.

Keywords: combustion, flame, Navier-Stokes equations, adaptive finite element method.

1 Introduction

Flame behavior has been extensively studied e.g. for the purpose of jet engine design [3]. Flame and smoke propagation models are used in fire-protection systems [4]. Simplified flame models are also of interest of the entertainment industry [8], [10]. In this paper, we present a flame simulation technique that can be used in either case. We use a 5-step reduced mechanism of methane-oxygen combustion, based on [13], which gives already reliable results while remaining relatively simple. The dynamics of the system is modeled with a variation of Navier-Stokes equations, presented e.g. in [5].

The model equations are solved with an adaptive finite element method. It is common to use adaptive mesh refinement for flame simulation, e.g. [5], [2], [1]. The refinement is usually performed to obtain high-resolution flame fronts. We propose an algorithm which additionally performs coarsening of the mesh at the regions of fast flow in order to obtain stability of the algorithm without decreasing the time step length. Further improvement of the convergence rate of the method is achieved by adjustment of the approximation order [12]. In order to keep track of small-scale features of the flow, we use a multilevel adaptation [15]. All the adaptive techniques are used to control the approximation error and ensure the stability of the algorithm.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 112–122, 2011.

2 Model

The main model equations are [5]

$$\frac{\partial}{\partial t}\rho_i + \operatorname{div}(\rho_i v) = \operatorname{div}(\rho \mathsf{D}_i \nabla Y_i) + \omega_i,\ i = 1 \dots N_s\,, \tag{1}$$

$$\frac{\partial}{\partial t}(\rho v) + \operatorname{div}(\rho v \otimes v) = -\nabla p + f_b\,, \tag{2}$$

$$\frac{\partial}{\partial t}(\rho T) + \operatorname{div}(\rho T v) = \frac{1}{c_p}\left(\operatorname{div}(\lambda \nabla T) - \sum_{i=1}^{N_s} \omega_i h_i\right)\,, \tag{3}$$

where ρ_i is the density of the i-th species, ρ the density of the mixture, v the velocity vector, p the pressure, T the temperature, Y_i the mass fraction of the i-th species, ω_i the reaction rate of the i-th species, N_s the number of species, f_b the buoyancy force, h_i the specific enthalpy of the i-th species, D_i the diffusion coefficient of the i-th species, c_p the specific heat at constant pressure, and λ the conduction coefficient. All the variables are functions defined on $\Omega \times [0, t_{max}]$, where $t_{max} < \infty$, and $\Omega \subset \mathbb{R}^2$. The mass fraction is given by

$$Y_i = \frac{\rho_i}{\rho}\,. \tag{4}$$

Pressure is given by the ideal gas law

$$p = R\rho T\,, \tag{5}$$

where R is the individual gas constant of the mixture. The specific enthalpy is given by

$$h_i = h_i^0 + \int_{T_0}^{T} c_{p,i}\, dT\,, \tag{6}$$

where h_i^0 is the standard enthalpy of formation of the i-th species, T_0 the standard temperature, and $c_{p,i}$ the specific heat of the i-th species at constant pressure.

The reaction properties are given by [7]

$$\omega_i = \sum_{j=1}^{N_r} \nu_{ij} q_j\,, \tag{7}$$

$$q_j = A_j T^{\beta_j} e^{\frac{-E_{aj}}{R_0 T}} \prod_{i=1}^{N_s} \left(\frac{\rho_i}{M_i}\right)^{-min\{\nu_{ij},0\}}\,, \tag{8}$$

where N_r is the number of reactions, q_j the rate of progress for the j-th reaction, $A_j > 0$ the pre-exponential factor for the j-th reaction, β_j the temperature exponent for the j-th reaction, ν_{ij} the stoichiometric coefficient of the i-th species for the j-th reaction, $E_{aj} > 0$ the activation energy for the j-th reaction, R_0 the

gas constant, and $M_i > 0$ the molar mass of the i-th species. The stoichiometric coefficients ν_{ij} are positive for products, and negative for reactants, so the following equation holds

$$\sum_{i=1}^{N_s} \omega_i = 0\,. \tag{9}$$

The chemical reactions are

$$\mathrm{CH_4} \rightarrow \mathrm{CH_3} + \frac{1}{2}\mathrm{H_2}\,, \tag{10}$$

$$2\mathrm{CH_3} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO} + 3\mathrm{H_2} + \mathrm{C}\,, \tag{11}$$

$$\mathrm{C} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO}\,, \tag{12}$$

$$\mathrm{CO} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO_2}\,, \tag{13}$$

$$\mathrm{H_2} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{H_2O}\,. \tag{14}$$

The mechanism is based on the 7-step mechanisms listed in [13] for lean, stoichiometric and rich methane-air mixtures. The first three reactions, in the case of a stoichiometric mixture, could be replaced by $\mathrm{CH_4} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{CO} + 2\mathrm{H_2}$. Splitting this into three reactions is important especially in the case of a lean mixture, when methyl takes part in soot formation in the 2nd reaction. In the 3rd reaction, by further oxidation, the soot forms carbon monoxide, which turns into carbon dioxide in the 4th reaction. Finally, the remaining hydrogen reacts with oxygen to produce water. Noticeably, all the reactions except the first consume oxygen, and thus limited oxygen supply may lead to release of soot and carbon monoxide as products of an incomplete combustion. The values of the reaction rate coefficients are given in Table 1.

Table 1. The reaction rate coefficients, calculated from GRI-Mech [14]

Reaction number	A_j	β_j	E_{aj} (cal/mol)
1	6.6E+8	1.6	10840
2	4E+13	0	0
3	5.8E+13	0	576
4	1.8E+10	0	2385
5	3.87E+4	2.7	6260

The initial and boundary conditions are as follows. A methane-oxygen mixture is injected into the model domain from the bottom, as in Figure 1. For all the boundary, we use Dirichlet boundary conditions: $p = const$, $\rho_i = const$. The initial temperature at the place of injection is raised above the minimum autoignition temperature.

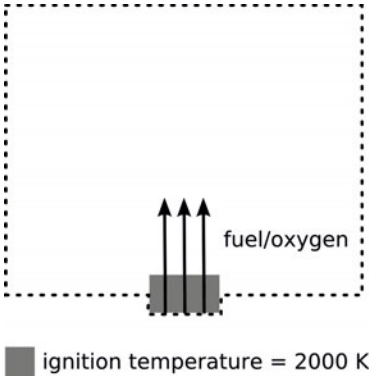


Fig. 1. The initial and boundary conditions

3 Algorithm

The model equations are solved with a fully adaptive finite element method, based on [11]. We define a mesh as a finite set $\mathcal{M} = \{x_i\}_{i \in I \subset \mathbb{N}} \subset \Omega$ and a neighborhood function $\mathcal{N} : \Omega \to \mathcal{P}(\mathcal{M})$. The neighborhood function maps a point $x \in \Omega$ into a nonempty set of its neighbors. The set of neighbors is chosen to minimize the condition

$$\gamma \sum_{x_i \in \mathcal{N}(x)} \|x - x_i\| + \max_{x_i \in \mathcal{N}(x)} \min_{x_j \in \mathcal{N}(x)} \measuredangle(x_i, x, x_j)\,, \tag{15}$$

where $\gamma > 0$ is a constant, $\|\cdot\| \equiv \|\cdot\|_2$ is the Euclidean norm, and $\measuredangle(x, y, z)$ is equal to the angle $\alpha \in [0, 2\pi]$ between $\overline{xy}$ and $\overline{yz}$. If $x = y$ or $y = z$ the angle is assumed 0. If $x = z \neq y$ the angle is assumed 2π.

The domain is divided into elements. Two points x, y belong to the same element iff $\mathcal{N}(x) = \mathcal{N}(y)$. This means that every mesh point itself is a degenerate element.

Having a mesh $(\mathcal{M}, \mathcal{N})$, we define the approximation $\tilde{u} : \Omega \to \mathbb{R}$ as

$$w_i(x) = 1 - (|\mathcal{N}(x)| - 1)\frac{\|x - x_i\|}{\sum_{x_j \in \mathcal{N}(x)} \|x - x_j\|}\,, \tag{16}$$

$$\Phi_i(x) = \frac{w_i(x)^{p_i}}{\sum_{x_j \in \mathcal{N}(x)} w_j(x)^{p_j}}\,, \tag{17}$$

$$\tilde{u}(x) = \sum_{x_i \in \mathcal{N}(x)} \Phi_i(x)\,(u_i + f_i \cdot (x - x_i))\,, \tag{18}$$

where $p_i \in \mathbb{R}$, $u_i \in \mathbb{R}$, $f_i \in \mathbb{R}^2$ are the approximation parameters. The parameters, as well as the mesh points $\mathcal{M}$, are sought to minimize the monitor function E, defined as

$$E = \frac{1}{a} E_\varepsilon + \frac{1}{b} E_I + E_\delta\,, \tag{19}$$

where

$$E_\varepsilon = \max_{x\in\Omega} \|\tilde{u}(x) - u(x)\|\,, \tag{20}$$

$$E_I = |I| = |\mathcal{M}|\,, \tag{21}$$

$$E_\delta = \max_{i\neq j} \frac{1}{\|x_i - x_j\|}\,, \tag{22}$$

$u : \Omega \to \mathbb{R}^2$ is the approximated function, and a, b are parameters of the method. An approximation analogous to the above can be defined for a vector function as well, that is for $u : \Omega \to \mathbb{R}^2$. The parameters a and b have the meanings of the expected error and the expected number of mesh points respectively. The parameters provide a convenient way to control the balance between the precision and the efficiency of the algorithm.

In the considered flame model, there are $N + 2$ variables: ρ_i, v, and T. For each variable, we seek for a solution, i.e. a mesh and an approximation defined on the mesh. The algorithm goes as follows

1. Generate a regular mesh on Ω for each model variable, $t \leftarrow 0$
2. Set the initial values
3. Refine each mesh by adding a new point in the center of every non-degenerate element and at each free node of every element
4. Solve the model equations on the given meshes
5. Coarsen the meshes by removing unimportant points
6. Add sub-levels to the meshes where the mesh density increases over an acceptable value
7. If the monitor function did not change significantly for any of the meshes: $t \leftarrow t + \Delta t$
8. Go to step 3

A free node of an element is a node which is not a mesh point. Deciding whether or not a mesh point is important is done by comparing the value of the monitor function before and after removing the point. If the value decreases, the point is unimportant. Otherwise, the point has to be restored. A significant change of the monitor function, mentioned in step 7 of the algorithm, is indicated by $|E^i - E^{i-1}| > \epsilon$, where i is the current iteration number and $\epsilon > 0$ is an arbitrary constant. The approximation error E_ε is estimated by the formula

$$E_\varepsilon \approx \max_{i\in I} \max_{x_j\in\mathcal{N}^\star(x_i)} \|\tilde{u}(\frac{x_i + x_j}{2}) - \tilde{u}^\star(\frac{x_i + x_j}{2})\|\,. \tag{23}$$

The star * denotes values calculated for $\mathcal{M}\backslash\{x_i\}$, that is without considering x_i as a mesh point.

3.1 Finding the Neighbors and the Elements

The set of neighbors $\mathcal{N}(x)$ for a given point $x \in \Omega$ and a given mesh $\mathcal{M}$ is found in constant average time. We use a Cartesian grid to localize quickly the

close mesh points in a given area. The neighbors are first sought in the grid cell which contains x. Next, the coincident cells are checked, and so on, until it is clear that no better set of neighbors can be found in respect to the optimization condition (15). For every checked grid cell, all the mesh points in the cell are tested as neighbors. If adding a given mesh point to the set of neighbors decreases (15), the point stays in the set. Otherwise, it is removed. If, for a current set of neighbors, the highest possible gain is lower than the lowest possible loss, the search ends. The side length of a grid cell is set to $\sqrt{\frac{|\Omega|}{|I|}}$. The parameter γ is set to $\frac{\pi}{6d} \ln |I|$, where d is the average distance between the mesh points. This way, the expected number of mesh points in a single grid cell is 1, and the expected number of neighbors for a given point is 3. Thus checking up to 9 grid cells is usually sufficient to find the set of neighbors. Generating the Cartesian grid and placing the mesh points in the correct cells takes linear time in respect to the size of the mesh, but it is only done once for a given mesh.

The elements are found in linear time in respect to the number of them. The edges of the elements are given by (x_i, x_j), $x_i \in \mathcal{M}$, $x_j \in \mathcal{N}^\star(x_i)$. Some of the edges may intersect. At each intersection a new element node is added, and the intersected edges are split. To find the intersections in linear time, again we use the Cartesian grid. Only the edges that share at least one coincident cell are tested for intersection. Once the element edges are establish, we can calculate the center points of the elements as the geometric centers of their nodes.

3.2 The Numerical Solver

Each model equation is a convection-diffusion equation, and thus it can be written in the general form

$$\frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u v) = \operatorname{div}(\mathsf{D}\nabla u) + s\,, \tag{24}$$

where u is a model variable in an intensive form (Y_i, v or T). This can be transformed to

$$\frac{\partial}{\partial t}(\rho u) + \rho v \cdot \nabla u + u v \cdot \nabla \rho + \rho u \operatorname{div} v = \mathsf{D}\Delta u + \nabla D \cdot \nabla u + s\,, \tag{25}$$

where s is the source term. After rearrangement the equation reads

$$(\rho v - \nabla \mathsf{D}) \cdot \nabla u = \mathsf{D}\Delta u + s - \frac{\partial}{\partial t}(\rho u) - u v \cdot \nabla \rho - \rho u \operatorname{div} v\,. \tag{26}$$

Let us introduce the following additional notation

$$f := \mathsf{D}\Delta u + s - \frac{\partial}{\partial t}(\rho u) - u v \cdot \nabla \rho - \rho u \operatorname{div} v\,, \tag{27}$$

$$g := \rho v - \nabla \mathsf{D}\,. \tag{28}$$

The equation (26) is solved in the following steps

1. $u_i^* \leftarrow u_i$, $i \in I$
2. $f_i^* \leftarrow \frac{f(x_i)}{|g(x_i)|} g(x_i)$, $i \in I$
3. $u_i^* \leftarrow \tilde{u}^\star(x_i)$, $i \in I$
4. Find p_i^* that minimizes $\| \frac{f(x)}{|g(x)|} g(x) - \nabla \tilde{u}^*(x) \|$, $i \in I$, $x := \frac{x_i + x_j}{2}$, $x_j \in \mathcal{N}^\star(x_i)$
5. If any of the approximation parameters changed significantly, go to step 2

The asterisk * denotes values calculated for the next time step. All the derivatives in f and g are approximated with the basic schemes. The time derivative in f is approximated using $\tilde{u}^*$ as

$$\frac{\partial}{\partial t}(\rho u) \approx \frac{\tilde{\rho}^* \tilde{u}^* - \tilde{\rho} \tilde{u}}{\Delta t} \,. \tag{29}$$

3.3 Multi-level Approach

If the following condition holds for any $x_i, x_j \in \mathcal{M}$, $x_i \neq x_j$

$$\| x_i - x_j \| < \delta \,, \tag{30}$$

where $\delta > 0$ is an arbitrary constant, then a new mesh level is added between x_i and x_j. The new level domain is defined as the convex hull of $\mathcal{N}(\frac{1}{2}(x_i + x_j))$. The calculation for every level is analogous to calculations for the root level. The height of the level tree is limited by an arbitrary constant.

3.4 Ensuring Real-Time Simulation

The simulation is said to run in real time if the time spent for calculating one time step is not greater than the time step length, $\Delta t_{\text{real}} \leq \Delta t$. This property is ensured by adjusting the method parameters a, b at every time step according to

$$b \leftarrow b \frac{\Delta t}{\Delta t_{\text{real}}} \,, \tag{31}$$

$$a \leftarrow \frac{a}{b} \,. \tag{32}$$

The initial values of the parameters are $a = \epsilon$, $b = 1$.

4 Results

We have tested the algorithm on three test cases corresponding to different equivalence ratios of the methane-oxygen mixture: $\phi = 0.6$ (a lean mixture), $\phi = 1$ (a stoichiometric mixture) and $\phi = 1.5$ (a rich mixture). The method accuracy has been verified by comparing the results to the those obtained with the direct numerical simulation and to the results presented in [2], [5], [6], [1] and [9]. The structure and composition of the flame, along with the numerical properties of the method, such as the mesh density, have been studied.

The mixture is injected into the domain with speed of 30 cm s^{-1}, through a hole with area of 12.56 mm^2. The temperature at the ignition area is set to 2000K at the beginning of the simulation. The numerical error tolerance ϵ is set to 10^{-10}, the minimum distance between mesh points δ is set to 10^{-3}, and the time step length Δt is set to 0.05s. The size of the domain is 20cm × 30cm.

Table 2. Frames per second (FPS), the number of mesh points, the error estimate, and the error with respect to direct numerical simulation for various configurations of the method parameters, at $t = 5$s.

ϕ, a, b	FPS	$\|I\|$	E_ε	DNS error
$\phi = 0.6$				
a = 1E-10, $b = 1$	15.2	3303	1E-6	0.3086
$a = 0.01$, $b = 1$	138.2	69	0.0773	0.6026
a = 1E-10, $b = 1000$	23.5	2410	0.0001	2.398
$a = 0.01$, $b = 1000$	45.6	735	0.0025	9.573
$\phi = 1$				
a = 1E-10, $b = 1$	19.8	2836	1E-7	0.2579
$a = 0.01$, $b = 1$	150.5	63	0.0238	0.5496
a = 1E-10, $b = 1000$	15.9	3508	1E-8	0.1162
$a = 0.01$, $b = 1000$	60.3	612	0.0002	2.4803
$\phi = 1.5$				
a = 1E-10, $b = 1$	20	2018	1E-7	0.2112
$a = 0.01$, $b = 1$	151.2	66	0.0639	0.597
a = 1E-10, $b = 1000$	15.4	2922	1E-7	0.0769
$a = 0.01$, $b = 1000$	63.1	585	0.0003	1.0587

Table 2 shows how some of the simulation properties depend on the method parameters for the three test cases. We can see that the simulation speed, expressed in frames per second, is highly correlated with the number of mesh points, as well as the approximation error estimate. In most cases, the simulation speed is above the critical 16 frames per second. The lowest approximation error estimate corresponds to the lowest error with respect to DNS, even though the correlation between these two properties is generally weak. Better accuracy is obtained for higher equivalence ratios.

Table 3 describes the mesh properties with respect to values of the equivalence ratio. Surprisingly, the average number of sub-levels is highest in the stoichiometric case and the lowest in the lean mixture case. This can be explained by higher concentration of error in the case of a lean flame. For the stoichiometric mixture, the maximum error is lower but there are more high-error points. The maximum depth of the level tree is limited to 4 and it reaches the limit. Most of the mesh points in the lean- and rich-mixture cases have 3 neighbors, with 5 being the maximum. In the stoichiometric case, most of the mesh points have 6

Table 3. The mesh properties

Property	$\phi = 0.6$	$\phi = 1$	$\phi = 1.5$
Avg. number of sub-levels	95	144	116
Max. level tree depth	4	4	4
Avg. number of neighbors	3.54	5.37	3.8
Max. number of neighbors	5	12	5
Avg. number of elements	6506	11244	3936

neighbors. The average number of elements shows a strong correlation with the number of points and the average number of neighbors.

Table 4. The peak values of soot and carbon monoxide mass fractions in the post-flame zone

ϕ	Our results		Values from [9]	
	Y_{CO}	Y_{C}	Y_{CO}	Y_{C}
0.6	0.2	0.001	0.05	0
1	0.01	0	0.002	0
1.5	0.3	0.002	0.1	0.002

Fig. 2. The CO_2 density for lean, stoichiometric and rich methane-oxygen mixtures

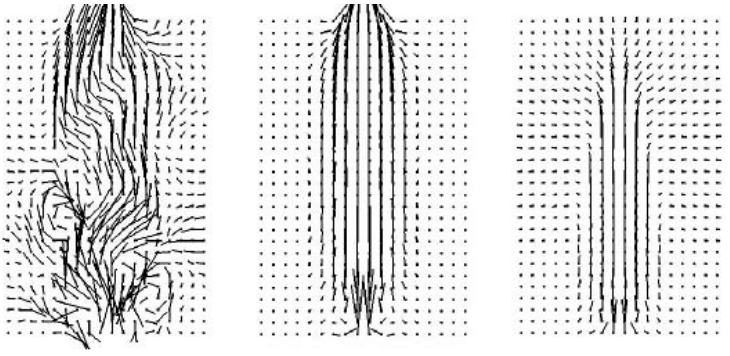


Fig. 3. The velocity vector field for lean, stoichiometric and rich methane-oxygen mixtures

Fig. 4. The light radiation in gray scale for lean, stoichiometric and rich methane-oxygen mixtures. The black color corresponds to the strongest radiation.

Table 4 compares the soot and CO release results with values presented in [9]. We can see that the soot release is highest in the rich-mixture case. Similar behavior is observed for carbon monoxide. Rich-mixture case is related to lower oxygen supply, which results in lower reaction rates of the reactions (2), (3), (4) and (5).

The CO_2 density presented on Figure 2 indicates the location and shape of the flame front. The highest concentration of CO_2 corresponds to the flame zone. The low CO_2 concentration region inside the flame corresponds to the pre-heat zone. The low CO_2 concentration region outside the flame corresponds to the post-flame zone. Figure 3 shows the velocity vector field. We can see that the flow is turbulent in the case of lean mixture, when the reaction is relatively slow as well as the flow itself. For the stoichiometric and rich flames the flame speed is higher and the flow becomes laminar.

References

1. Burman, E., Ern, A., Giovangigli, V.: Bunsen flame simulation by finite elements on adaptively refined, unstructured triangulations. Combustion Theory and Modelling 8(1), 65–84 (2003)
2. Consul, R., Pérez-Segarra, C.D., Claramunt, K., Cadafalch, J., Oliva, A.: Detailed numerical simulation of laminar flames by a parallel multiblock algorithm using loosely coupled computers. Combustion Theory and Modelling 7(3), 525–544 (2003)
3. Dunn, M., Masri, A., Bilger, R., Barlow, R.: Finite rate chemistry effects in highly sheared turbulent premixed flames. Flow, Turbulence and Combustion 85(3/4), 621–648 (2010)
4. Fukuchi, N., Takao, J., Hu, C.: Thermal properties and smoke diffusion of oil pool fires in engine room for fire safety design. International Journal of Offshore and Polar Engineering 15(1) (2005)
5. Graziadei, M.: Using local defect correction for laminar flame simulation. Eindhoven University of Technology (2004)
6. van 't Hof, B.: Numerical Aspects of Laminar Flame Simulation. Ph.D. thesis, Eindhoven University of Technology (1998)
7. Kee, R., Rupley, F., Meeks, E., Miller, J.: Chemkin-III: A fortran chemical kinetics package for the analysis of gasphase chemical and plasma kinetics(1996), http://www.reactiondesign.com/products/open/chemkin.html
8. Lamorlette, A., Foster, N.: Structural modeling of flames for a production environment. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques (2002)
9. Mcenally, C., Schaffer, A., Long, M., Pfefferle, L., Smooke, M., Colket, M., Hall, R.: Computational and experimental study of soot formation in a coflow, laminar ethylene diffusion flame. In: Proceedings of the Combustion Institute, pp. 1497–1505 (1998)
10. Nguyen, D., Fedkiw, R., Jensen, H.: Physically based modeling and animation of fire. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques (2002)
11. Opiola, P.: A new strategy for finite element method adaptation (2010), http://www.ii.uj.edu.pl/~opiola/adaptive_fem.pdf
12. Rao, V., Das, P., Sundararajan, T.: An adaptive hp-version of the finite element method applied to flame propagation problems. International Journal for Numerical Methods in Engineering 40(17), 3181–3203 (1997)
13. Skevis, G., Goussis, D., Mastorakos, E.: Understanding methane flame kinetics from reduced mechanisms. International Journal of Alternative Propulsion 1(2/3), 216–227 (2007)
14. Smith, G., Golden, D., Frenklach, M., Moriarty, N., Eiteneer, B., Goldenberg, M., Bowman, C., Hanson, R., Song, S., Gardiner Jr., W.C., Lissianski, V., Qin, Z.: Gri-mech 3.0 (1996), http://www.me.berkeley.edu/gri_mech
15. Teigland, R., Eliassen, I.: A multiblock/multilevel mesh refinement procedure for cfd computations. International Journal for Numerical Methods in Fluids 36(5), 519–538 (2001)

Security Improvement on a Group Key Exchange Protocol for Mobile Networks*

Junghyun Nam[1], Kwangwoo Lee[2], Juryon Paik[2], Woojin Paik[1], and Dongho Won[2,**]

[1] Department of Computer Engineering, Konkuk University, Korea
jhnam@kku.ac.kr,wjpaik@kku.ac.kr
[2] Department of Computer Engineering, Sungkyunkwan University, Korea
kwlee@security.re.kr,wise96@ece.skku.ac.kr,dhwon@security.re.kr

Abstract. A group key exchange (GKE) protocol is designed to allow a group of parties communicating over a public network to establish a common secret key called a *session key*. As group-oriented applications gain popularity over the Internet, a number of GKE protocols have been suggested to provide those applications with a secure multicast channel. Among the many protocols is the GKE protocol presented by Dutta and Dowling for mobile ad hoc networks. In this paper, we are concerned with the security of the Dutta-Dowling protocol. Their protocol carries a proof of security in the standard adversarial model which captures unknown key-share attacks. But unlike the claim of provable security, the Dutta-Dowling protocol fails to achieve unknown key-share resilience. We here reveal this security vulnerability of the protocol and show how to address it.

Keywords: Mobile ad hoc network, group key exchange, cluster-based network, unknown key-share attack.

1 Introduction

Mobile ad hoc networks have matured as a viable means to provide anytime-anywhere networking services for infrastructure-free communication over a shared wireless channel. With the broad range of wirelessly connected mobile devices being used today, security in mobile ad hoc networks is of essential and increasing importance. However, it still remains a difficult task to provide strong protection for mobile ad hoc networks where nodes may be resource constrained. Although mobile computing technology has become more powerful and accessible than ever before, mobile devices are typically characterized by low processing capability and limited power supply [14] which are inherent to the mobility nature. It is thus necessary that the cost due to security-related operations should be

* This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0020210).
** Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 123–132, 2011.

minimized for mobile devices in such a way that the required security goals are not compromised. This requirement makes it especially difficult to secure mobile ad hoc networks. In fact despite all the research efforts made so far, security is still a major limiting factor towards the eventual success of mobile ad hoc networks [2,8,13,12].

Recently, Dutta and Dowling [7] proposed a group key exchange (GKE) protocol for securing mobile ad hoc networks. The Dutta-Dowling protocol divides the whole nodes in a mobile ad hoc network into a set of clusters. Each cluster consists of a single sponsor node and multiple regular nodes. A sponsor node, as the representative of its cluster, is responsible for communicating with the sponsors of other clusters. Such cluster-based networks can be found in many applications including military missions and environmental monitoring. For instance, a military network consists of mobile devices carried by soldiers, automatic weapons, sensing devices, etc. In this setting, a platoon commander may act as a sponsor of its cluster and may be able to communicate with platoon commanders of other clusters.

The GKE protocol presented by Dutta and Dowling was claimed to be provably secure in a formal security model which captures unknown key-share attacks. But despite the claim of provable security, their GKE protocol is in fact insecure against an unknown key-share attack. In the current paper, we report this security problem with the Dutta-Dowling protocol and figure out how to solve it. Our result implies that the claimed proof of security for the Dutta-Dowling protocol is invalid.

The remainder of this paper is organized as follows. Section 2 reviews the GKE protocol proposed by Dutta and Dowling. Section 3 presents our known key-share attack on the Dutta-Dowling protocol and offers a security patch for the protocol. Section 4 provides an overview of the formal proof model for the protocol, along with an associated definition of security. Section 5 breaks the so-called "AKE security" of the protocol.

2 The Dutta-Dowling Protocol

Dutta and Dowling [7] construct their protocol in two steps: first, they present an unauthenticated GKE protocol UP; then, they transform UP into an authenticated protocol AP. The transformation from UP to AP is done by using a simplified variant of the compiler presented by Katz and Yung [11].

2.1 Unauthenticated Protocol UP

Let $\mathbb{G}$ be be a cyclic (multiplicative) group of prime order q, and let g be a random fixed generator of $\mathbb{G}$. Let $H : \mathbb{G} \rightarrow \mathbb{Z}_q^*$ be a cryptographically secure hash function. The protocol UP is designed to be secure against passive adversaries. The users participating in UP are divided into into small groups called clusters. Assume there are n clusters $C_1, \ldots, C_n$. Let S_i denote the sponsor of cluster C_i. UP runs in two phases:

Phase 1 (Cluster Key Generation): The users in each C_i generate a cluster key CK_i by running the protocol Setup (described below).

Phase 2 (Group Key Generation): Once all cluster keys CK_1, ..., CK_n have been generated, the sponsors $S_1, \ldots, S_n$ run the Setup protocol again to establish a group session key SK. But this time, each S_i sets its random exponent x_i equal to $H(CK_i)$ so that other users in C_i can compute the session key. In this run of Setup, it is assumed that the broadcast messages are received not only by sponsors but also by all other users.

In the following description of protocol Setup, all indices are to be taken in a cycle, i.e., $U_{n+1} = U_1$, etc.

Protocol Setup

Round 1: Each user U_i chooses a random $x_i \in \mathbb{Z}_q^*$ and broadcasts $y_i = g^{x_i}$.

Round 2: Each user U_i computes the left key $K_i^L = y_{i-1}^{x_i}$ and the right key $K_i^R = y_{i+1}^{x_i}$ and broadcasts $Y_i = K_i^R / K_i^L$.

Key Computation: Each U_i computes

$$\begin{aligned} K_{i+1}^R &= Y_{i+1} K_i^R, \\ K_{i+2}^R &= Y_{i+2} K_{i+1}^R, \\ &\vdots \\ K_{i+n-1}^R &= Y_{i+n-1} K_{i+n-2}^R. \end{aligned}$$

U_i verifies that K_{i+n-1}^R is equal to K_i^L. U_i aborts if the verification fails. Otherwise, U_i computes the session key SK as:

$$\begin{aligned} SK &= K_1^R K_2^R \cdots K_n^R \\ &= g^{x_1 x_2 + x_2 x_3 + \cdots + x_n x_1}. \end{aligned}$$

2.2 Authenticated Protocol AP

Let $\Sigma = (\mathsf{Kgen}, \mathsf{Sign}, \mathsf{Vrfy})$ be a signature scheme which is strongly unforgeable under adaptive chosen message attack. Here, Kgen is the key generation algorithm, Sign is the signature generation algorithm, and Vrfy is the signature verification algorithm. During the initialization phase of AP, each user U_i generates its long-term verification/signing keys (PK_i, SK_i) by running $\mathsf{Kgen}(1^\kappa)$ and makes the verification key PK_i public. Each user U_i should maintain a counter c_i to participate in the protocol AP. The counter c_i is incremented every time U_i participates in a new run of AP. Let $U_1, \ldots, U_n$ be the users who wish to establish a common session key. As part of the initialization, each U_i sets $\mathsf{pid}_i = \{U_1, \ldots, U_n\}$. In the protocol AP, the users simply perform the protocol UP, but signing outgoing messages and verifying the correctness of incoming messages as follows:

- Let m_i^t be the t-th message that U_i is supposed to broadcast as part of protocol UP. Then, U_i computes the signature $\sigma_i^t = \mathsf{Sign}_{SK_i}(U_i|t|m_i^t|c_i)$ and broadcasts $M_i^t = U_i|t|m_i^t|c_i|\sigma_i^t$ as the replacement of m_i^t.
- When U_i receives message $M_j^t = U_j|t|m_j^t|c_j|\sigma_j^t$ from U_j, it checks that: (1) $U_j \in \mathsf{pid}_i$, (2) t is the next expected sequence number for messages from U_j, and (3) $\mathsf{Vrfy}_{PK_j}(U_j|t|m_j^t|c_j, \sigma_j^t) = 1$. U_i aborts if any of these conditions fail to hold. Otherwise, U_i continues as it would in UP upon receiving message M_j^t.
- Each non-aborted user U_i computes the session key as in UP and defines the session identifier $\mathsf{sid}_i = \langle (U_1, c_1), \ldots, (U_n, c_n) \rangle$.

The above procedure for converting UP into AP is a variant of the Katz-Yung compiler [11] which transforms any GKE protocol secure against a passive adversary into one that is secure against an active adversary. But unlike the Katz-Yung compiler, the transformation here makes use of counters instead of random nonces and accordingly does not require an additional round for exchanging random nonces among users.

3 Security Analysis

This section conducts a security analysis on the above-described protocol AP. We first reveal the vulnerability of AP to an unknown key-share attack and then figure out how to fix the protocol.

3.1 Unknown Key-Share Attack

The protocol AP is vulnerable to an unknown key-share attack in the presence of an active adversary. An adversary U_A is said to succeed in an unknown key-share attack if at the end of the attack, there exist two parties U_i and U_j such that: *(1) U_i and U_j have computed the same key; (2) U_i is unaware of the "key share" with U_j and mistakenly believes that its key is shared with U_A; and (3) U_j correctly believes that its key is shared with U_i.* As implied by this definition, U_A need not obtain any session key to benefit from an unknown key-share attack [6,1,9,4]. Consider a protocol session S to be run by the users of group $\mathcal{G} = \{U_1, U_2, U_3\}$. Now suppose that U_1 and U_2 accept the invitation by U_A to participate in a new concurrent session S', thus forming the group $\mathcal{G}' = \{U_1, U_2, U_A\}$. We will consider the users of two sessions as sponsors who have already established their respective cluster keys and yet need to establish a common group key by running the protocol Setup. Hereafter, we use Π_i and Π_i' to denote U_i's instances participating respectively in S and S'. Let $\textsc{Exp}_i$ and $\textsc{Div}_i$ be the first and second messages sent by U_i in the authenticated version of Setup. Let α_i and β_i be the signatures contained in $\textsc{Exp}_i$ and $\textsc{Div}_i$. The attack works as follows:

1. As session S starts, Π_1, Π_2 and U_3 will send their first messages $\textsc{Exp}_1 = U_1|1|y_1|c_1|\alpha_1$, $\textsc{Exp}_2 = U_2|1|y_2|c_2|\alpha_2$ and $\textsc{Exp}_3 = U_3|1|y_3|c_3|\alpha_3$, respectively. The adversary U_A intercepts $\textsc{Exp}_3$ while blocking $\textsc{Exp}_1$ and $\textsc{Exp}_2$ from

reaching U_3. In other words, Π_1 and Π_2 never receive EXP_3 and U_3 receives neither EXP_1 nor EXP_2.

2. As a participant of session S', the adversary U_A sends the message $\text{EXP}_A = U_A|1|y_3|c_A|\alpha_A$ to Π_1' and Π_2' and receives the messages $\text{EXP}_1' = U_1|1|y_1'|c_1'|\alpha_1'$ and $\text{EXP}_2' = U_2|1|y_2'|c_2'|\alpha_2'$ respectively from Π_1' and Π_2'. Notice that EXP_A contains y_3 (which is generated by U_3 and is obtainable from EXP_3). We mean by this that U_A has computed its signature α_A as $\alpha_A = \mathsf{Sign}_{SK_A}(U_A|1|y_3|c_A)$.
3. U_A forwards the received messages EXP_1' and EXP_2' to U_3 as if they are sent by Π_1 and Π_2, respectively. These messages will pass U_3's verification since α_1' (resp. α_2') is a valid signature on $U_1|1|y_1'|c_1'$ (resp. $U_2|1|y_2'|c_2'$) under the verification key PK_1 (resp. PK_2). Hence, U_3 will send out its second message $\text{DIV}_3 = U_3|2|Y_3|c_3|\beta_3$. If we let $y_1' = g^{x_1'}$ and $y_2' = g^{x_2'}$, then clearly

$$Y_3 = g^{x_3x_1' - x_2'x_3}.$$

4. U_A intercepts DIV_3, computes $\beta_A = \mathsf{Sign}_{SK_A}(U_A|2|Y_3|c_A)$ (using Y_3 from U_3), and sends $\text{DIV}_A = U_A|2|Y_3|c_A|\beta_A$ to Π_1' and Π_2'. Meanwhile, Π_1' and Π_2' will send U_A their respective messages $\text{DIV}_1' = U_1|2|Y_1'|c_1'|\beta_1'$ and $\text{DIV}_2' = U_2|2|Y_2'|\ c_2'|\beta_2'$ where

$$Y_1' = g^{x_1'x_2' - x_3x_1'} \text{ and } Y_2' = g^{x_2'x_3 - x_1'x_2'}.$$

5. U_A forwards DIV_1' and DIV_2' to U_3 as if they are from Π_1 and Π_2, respectively. These messages will pass U_3's verifications since the signatures β_1' and β_2' are both valid and the following equation holds: $K_3^L = Y_2'Y_1'K_3^R$, where $K_3^L = g^{x_2'x_3}$ and $K_3^R = g^{x_3x_1'}$.
6. Consequently, Π_1', Π_2' and U_3 will compute the same session key $SK = g^{x_1'x_2' + x_2'x_3 + x_3x_1'}$.

At the end of the attack: (1) U_1, U_2 and U_3 have computed the same session key SK; (2) U_1 and U_2 believe that SK is shared with U_A, while in fact it is shared with U_3; (3) U_3 believes that SK is shared with U_1 and U_2. This shows that the protocol AP is vulnerable to an unknown key-share attack when two protocol sessions are running concurrently with some joint participants.

3.2 Countermeasure

The vulnerability of AP to the unknown key-share attack stems from the fact that U_3 cannot distinguish between the signatures generated by Π_1 (resp. Π_2) and those generated by Π_1' (resp. Π_2'). A simple way to fix the protocol is to change the computations of the signatures α_i and β_i to $\alpha_i = \mathsf{Sign}_{SK_i}(U_i|1|y_i|c_i|\mathsf{pid}_i)$ and $\beta_i = \mathsf{Sign}_{SK_i}(U_i|2|Y_i|c_i|\mathsf{pid}_i)$. α_i and β_i are now computed by including participants' identities as part of the messages being signed. With this modification applied, the signatures from Π_1' and Π_2' can no longer pass U_3's verification since the participants are different between the two sessions S and S'. Hence, the unknown key-share attack is not valid against the improved protocol.

4 Overview of Security Model

The protocol AP comes along with a claimed proof of its security in a formal model of communication and adversarial capabilities. The proof model used for AP is based on Bresson et al.'s 2002 model for dynamic group key exchange [3]. Here we break the security of AP in the context of the proof model.

4.1 Participants

Let $\mathcal{U}$ be a set of all users who can participate in a group key exchange protocol. The users in any subset of $\mathcal{U}$ may run the group key exchange protocol at any point in time to establish a session key. Each user may run the protocol multiple times either serially or concurrently, with possibly different groups of participants. Thus, at a given time, there could be many instances of a single user. We use Π_i^π to denote the π-th instance of user U_i. Before the protocol is executed for the first time, each user $U_i \in \mathcal{U}$ creates a long-term public/private key pair (PK_i, SK_i) by running a key generation algorithm $\mathcal{K}(1^\kappa)$. All instances of a user share the public/private keys of the user even if they participate in their respective sessions independently. Each private key is kept secret by its owner while the public keys of all users are publicized.

4.2 Adversary

The adversary is in complete control of every aspect of all communications between participants, and may ask, at any time, them to open up access to their long-term secret keys. These capabilities and others of the adversary are modeled via various oracles to which the adversary is allowed to make queries.

- $\mathsf{Send}(\Pi_i^\pi, m)$: This query sends message m to instance Π_i^π. The instance Π_i^π proceeds as it would in the protocol upon receiving m; the instance updates its state performing any required computation, and generates and sends out a response message as needed. The response message, if any, is the output of this query and is returned to the adversary. This models active attacks on the protocol, allowing the adversary to control at will all message flows between instances. A query of the form $\mathsf{Send}(\Pi_i^\pi, \langle U_1, \ldots, U_n \rangle)$ prompts Π_i^π to initiate a protocol run in which the participants are $U_1, \ldots, U_n \in \mathcal{U}$.
- $\mathsf{Execute}(U_1, \ldots, U_n)$: This query lets the protocol be executed among the users $U_1, \ldots, U_n$ (or rather among their instances). The transcript of the protocol execution is returned to the adversary as the output of the query. This oracle call represents passive eavesdropping of a protocol execution.
- $\mathsf{Reveal}(\Pi_i^\pi)$: This query returns to the adversary the session key SK_i^π held by instance Π_i^π. This oracle call captures the idea that exposure of some session keys should not affect the security of other session keys.
- $\mathsf{Corrupt}(U_i)$: This query returns all long-term secrets of U_i. The adversary can issue this query at any time regardless of whether U_i is currently executing the protocol or not. This oracle call captures the idea that damage due to

loss of U_i's long-term secrets should be restricted to those sessions where U_i will participate in the future.
- $\mathsf{Test}(\Pi_i^\pi)$: This oracle call does not model any capability of the adversary, but is used to define the semantic security of the session key SK_i^π. The output of this query depends on the hidden bit $\mathbf{b}$ chosen uniformly at random from $\{0,1\}$. The Test oracle returns the real session key held by Π_i^π if $\mathbf{b} = 1$, or returns a random key drawn from the session-key space if $\mathbf{b} = 0$. The adversary is allowed to access the Test oracle only once, and the only Test query must be directed to a *fresh* instance (see Section 4.3 for the definition of freshness).

Remark 1. A common misunderstanding regarding the $\mathsf{Corrupt}$ oracle is that it is included in the model only for the purpose of dealing with (perfect) forward secrecy. The truth is that $\mathsf{Corrupt}$ queries not only model forward secrecy but also capture a variety of impersonation attacks. For example, key exchange protocols proven secure in a model that allows $\mathsf{Corrupt}$ queries ought to be secure against unknown key share attacks [4]. Another example is attacks by malicious insiders, though we focus on outsider attacks in this work. We refer the interested reader to [10] for details on how $\mathsf{Corrupt}$ queries are used to model insider attacks. Basically, the corruption of a principle U_i should not lead the adversary to have the ability to impersonate any principle other than U_i, because such ability would endanger even those sessions where U_i is not invited to participate.

4.3 Security Definition

Definition of security of a GKE protocol is based on the notion of *freshness* which in turn is defined using the notion of *partnership*.

Partnership. Like most of previous works, we define partnership between instances via *session IDs* and *partner IDs*. Literally, a session ID is a unique identifier of a protocol session. A partner ID is the identities of a group of users who intend to establish their common session key. Session IDs are typically computed during protocol runs whereas partner IDs should be given as input to protocol participants. We let sid_i^π be the session ID computed by instance Π_i^π and let pid_i^π be the partner ID provided to instance Π_i^π. An instance is said to *accept* when it successfully computes a session key in a protocol execution. Let acc_i^π be a boolean variable that evaluates to TRUE if Π_i^π has accepted, and FALSE otherwise. We say that any two instances Π_i^π and Π_j^ω are *partnered* if all the following three conditions are satisfied: (1) $\mathsf{sid}_i^\pi = \mathsf{sid}_j^\omega$, (2) $\mathsf{pid}_i^\pi = \mathsf{pid}_j^\omega$, and (3) $\mathsf{acc}_i^\pi = \mathsf{acc}_j^\omega =$ TRUE.

Freshness. An unfresh instance is an instance whose session key can be exposed by trivial means. More precisely:

Definition 1. *The instance Π_i^π is unfresh if any of the following conditions holds:*

1. *The adversary queried $\mathsf{Reveal}(\Pi_i^\pi)$ or $\mathsf{Reveal}(\Pi_j^\omega)$ where Π_i^π and Π_j^ω are partnered.*

2. *A* Corrupt *query was asked for some user in* pid_i^π *before a* Send *query has been asked for an instance of some user in* pid_i^π.

All other instances are considered fresh.

As already mentioned, the adversary is disallowed to make a Test query against an unfresh instance.

Security. The security of a GKE protocol P against an adversary $\mathcal{A}$ is defined in terms of the probability that $\mathcal{A}$ succeeds in distinguishing a real session key established in an execution of P from a random session key. That is, the adversary $\mathcal{A}$ is considered successful in attacking P if it breaks the semantic security of session keys generated by P. More precisely, the security is defined in the following context. The adversary $\mathcal{A}$ executes the protocol exploiting as much parallelism as possible and asking any queries allowed in the model. During executions of the protocol, the adversary $\mathcal{A}$, at any time, asks a Test query to a fresh instance, gets back a key as the response to this query, and at some later point in time, outputs a bit $\mathbf{b}'$ as a guess for the value of the hidden bit $\mathbf{b}$ used by the Test oracle.[1] Then the advantage of $\mathcal{A}$ in attacking protocol P is denoted by $\mathsf{Adv}_\mathsf{P}(\mathcal{A})$, and is defined as

$$\mathsf{Adv}_\mathsf{P}(\mathcal{A}) = |2 \cdot \Pr[\mathbf{b} = \mathbf{b}'] - 1|.$$

This notion of security is commonly termed as "AKE security". Protocol P is said to be *AKE-secure* if the advantage $\mathsf{Adv}_\mathsf{P}(\mathcal{A})$ is negligible (as a function of the security parameter) for all probabilistic polynomial time adversaries $\mathcal{A}$.

5 Breaking AKE Security

AKE-secure protocols should be secure against unknown key-share attacks. The protocol AP carries a claimed proof of its AKE security, but as we have seen, it is not secure against an unknown key-share attack. This implies that the security proof for AP is flawed. In this section, we interpret our attack in the context of the formal proof model to show that the attack does indeed break the AKE security of the protocol.

Our attack on AP is well captured in the proof model. Table 1 shows the sequence of oracle queries corresponding to the attack scenario given in Section 3.1. The goal of the adversary $\mathcal{A}$ is to break the AKE security of AP. $\mathcal{A}$ begins by asking $\mathsf{Corrupt}(U_A)$ to obtain the signing key SK_A of U_A. Then $\mathcal{A}$ initiates two sessions $S : \{U_1, U_2, U_3\}$ and $S' : \{U_1, U_2, U_A\}$ by asking the initial Send queries: $\mathsf{Send}(\Pi_1^1, \langle U_1, U_2, U_3\rangle)$, $\mathsf{Send}(\Pi_2^1, \langle U_1, U_2, U_3\rangle)$, $\mathsf{Send}(\Pi_3^1, \langle U_1, U_2, U_3\rangle)$, $\mathsf{Send}(\Pi_1^2, \langle U_1, U_2, U_A\rangle)$, $\mathsf{Send}(\Pi_2^2, \langle U_1, U_2, U_A\rangle)$. Notice that no instance of U_A needs to be asked this form of Send query because $\mathcal{A}$ will simulate by itself the actions of U_A. The rest of the queries are straightforward from the attack scenario.

[1] Of course, the adversary is prohibited from asking a Reveal query to the instance that is partnered with the tested instance, because such Reveal query would trivially expose the test session key.

Table 1. The sequence of oracle queries corresponding to the unknown key-share attack described in Section 3.1

	Query	Response
1	$\mathsf{Corrupt}(U_A)$	SK_A
2	$\mathsf{Send}(\Pi_1^1, \langle U_1, U_2, U_3\rangle)$	$\textsc{Exp}_1 = U_1\|1\|y_1\|c_1\|\alpha_1$
3	$\mathsf{Send}(\Pi_2^1, \langle U_1, U_2, U_3\rangle)$	$\textsc{Exp}_2 = U_2\|1\|y_2\|c_2\|\alpha_2$
4	$\mathsf{Send}(\Pi_3^1, \langle U_1, U_2, U_3\rangle)$	$\textsc{Exp}_3 = U_3\|1\|y_3\|c_3\|\alpha_3$
5	$\mathsf{Send}(\Pi_1^2, \langle U_1, U_2, U_A\rangle)$	$\textsc{Exp}'_1 = U_1\|1\|y'_1\|c'_1\|\alpha'_1$
6	$\mathsf{Send}(\Pi_2^2, \langle U_1, U_2, U_A\rangle)$	$\textsc{Exp}'_2 = U_2\|1\|y'_2\|c'_2\|\alpha'_2$
7	$\mathsf{Send}(\Pi_1^2, \textsc{Exp}_A = U_A\|1\|y_3\|c_A\|\alpha_A)$	
8	$\mathsf{Send}(\Pi_1^2, \textsc{Exp}'_2 = U_2\|1\|y'_2\|c'_2\|\alpha'_2)$	$\textsc{Div}'_1 = U_1\|2\|Y'_1\|c'_1\|\beta'_1$
9	$\mathsf{Send}(\Pi_2^2, \textsc{Exp}_A = U_A\|1\|y_3\|c_A\|\alpha_A)$	
10	$\mathsf{Send}(\Pi_2^2, \textsc{Exp}'_1 = U_1\|1\|y'_1\|c'_1\|\alpha'_1)$	$\textsc{Div}'_2 = U_2\|2\|Y'_2\|c'_2\|\beta'_2$
11	$\mathsf{Send}(\Pi_3^1, \textsc{Exp}'_1 = U_1\|1\|y'_1\|c'_1\|\alpha'_1)$	
12	$\mathsf{Send}(\Pi_3^1, \textsc{Exp}'_2 = U_2\|1\|y'_2\|c'_2\|\alpha'_2)$	$\textsc{Div}_3 = U_3\|2\|Y_3\|c_3\|\beta_3$
13	$\mathsf{Send}(\Pi_1^2, \textsc{Div}_A = U_A\|2\|Y_3\|c_A\|\beta_A)$	
14	$\mathsf{Send}(\Pi_1^2, \textsc{Div}'_2 = U_2\|2\|Y'_2\|\ c'_2\|\beta'_2)$	(accept)
15	$\mathsf{Send}(\Pi_2^2, \textsc{Div}_A = U_A\|2\|Y_3\|c_A\|\beta_A)$	
16	$\mathsf{Send}(\Pi_2^2, \textsc{Div}'_1 = U_1\|2\|Y'_1\|c'_1\|\beta'_1)$	(accept)
17	$\mathsf{Send}(\Pi_3^1, \textsc{Div}'_1 = U_1\|2\|Y'_1\|c'_1\|\beta'_1)$	
18	$\mathsf{Send}(\Pi_3^1, \textsc{Div}'_2 = U_2\|2\|Y'_2\|\ c'_2\|\beta'_2)$	(accept)
19	$\mathsf{Reveal}(\Pi_1^2)$ (or $\mathsf{Reveal}(\Pi_2^2)$)	SK_1^2 (or SK_2^2)
20	$\mathsf{Test}(\Pi_3^1)$	$\overline{SK}$

$\overline{SK}$: either the real session key SK_3^1 or a random session key.

The instance Π_3^1 accepts when the 18-th query is asked while Π_1^2 (resp. Π_2^2) accepts when the 14-th (resp. 16-th) query is asked. The computed session keys SK_1^2, SK_2^2 and SK_3^1 are all set equal to $g^{x'_1x'_2+x'_2x_3+x_3x'_1}$. The instance Π_3^1 is fresh under Definition 1; no one in $\mathsf{pid}_3^1 = \{U_1, U_2, U_3\}$ has been sent a $\mathsf{Corrupt}$ query and no Reveal query has been made against Π_3^1 or its partners. (Notice that by definition, Π_3^1 is not partnered with Π_1^2 and Π_2^2 since pid_3^1 is different from pid_1^2 and pid_2^2.) Thus, $\mathcal{A}$ may test (i.e., ask the Test query against) the instance Π_3^1. Because $\mathcal{A}$ may query $\mathsf{Reveal}(\Pi_1^2)$ and $\mathsf{Reveal}(\Pi_2^2)$ without affecting the freshness of Π_3^1, it follows that $\Pr[\mathbf{b} = \mathbf{b}'] = 1$ and hence $\mathsf{Adv}_{\mathsf{AP}}(\mathcal{A}) = 1$. Therefore, $\mathcal{A}$ achieves its goal of breaking the AKE security of protocol AP.

References

1. Blake-Wilson, S., Menezes, A.: Unknown key-share attacks on the station-to-station (STS) protocol. In: Imai, H., Zheng, Y. (eds.) PKC 1999. LNCS, vol. 1560, pp. 154–170. Springer, Heidelberg (1999)
2. Borisov, N., Goldberg, I., Wagner, D.: Intercepting mobile communications: the insecurity of 802.11. In: 7th ACM Conference on Mobile Computing and Networking, pp. 180–189 (2001)
3. Bresson, E., Chevassut, O., Pointcheval, D.: Dynamic group Diffie-Hellman key exchange under standard assumptions. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 321–336. Springer, Heidelberg (2002)

4. Choo, K.-K., Boyd, C., Hitchcock, Y.: Errors in computational complexity proofs for protocols. In: Roy, B. (ed.) ASIACRYPT 2005. LNCS, vol. 3788, pp. 624–643. Springer, Heidelberg (2005)
5. Choo, K.-K.R., Boyd, C., Hitchcock, Y., Maitland, G.: On session identifiers in provably secure protocols. In: Blundo, C., Cimato, S. (eds.) SCN 2004. LNCS, vol. 3352, pp. 351–366. Springer, Heidelberg (2005)
6. Diffie, W., Oorschot, P., Wiener, M.: Authentication and authenticated key exchanges. Designs, Codes, and Cryptography 2(2), 107–125 (1992)
7. Dutta, R., Dowling, T.: Secure and efficient group key agreements for cluster based networks. In: Gavrilova, M.L., Tan, C.J.K., Moreno, E.D. (eds.) Transactions on Computational Science IV. LNCS, vol. 5430, pp. 87–116. Springer, Heidelberg (2009)
8. Johnston, D., Walker, J.: Overview of IEEE 802.16 security. IEEE Security and Privacy Magazine 2(3), 40–48 (2004)
9. Kaliski, B.S.: An unknown key-share attack on the MQV key agreement protocol. ACM Transactions on Information and System Security 4(3), 275–288 (2001)
10. Katz, J., Shin, J.: Modeling insider attacks on group key-exchange protocols. In: 12th ACM Conference on Computer and Communications Security (CCS 2005), pp. 180–189 (2005)
11. Katz, J., Yung, M.: Scalable protocols for authenticated group key exchange. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 110–125. Springer, Heidelberg (2003)
12. Nam, J., Kim, S., Won, D.: A weakness in the Bresson-Chevassut-Essiari-Pointcheval's group key agreement scheme for low-power mobile devices. IEEE Communications Letters 9(5), 429–431 (2005)
13. Ng, S.-L., Mitchell, C.: Comments on mutual authentication and key exchange protocols for low power wireless communications. IEEE Communications Letters 8(4), 262–263 (2004)
14. Potlapally, N.R., Ravi, S., Raghunathan, A., Jha, N.K.: Analyzing the energy consumption of security protocols. In: 2003 ACM International Symposium on Low Power Electronics and Design, pp. 30–35 (2003)

Energy and Path Aware Clustering Algorithm (EPAC) for Mobile Ad Hoc Networks

Waqar Asif and Saad Qaisar*

National University of Sciences and Technology, H-12 Islamabad, Pakistan
{09mscsewasif,saad.qaisar}@seecs.edu.pk

Abstract. Node clustering is a technique that mitigates the change in topology in Ad hoc communication. It stabilizes the end to end communication path and maximizes the path life time. In SWARM communication, each cluster is assigned an objective and expected to complete it in the available resources. Most of the algorithms previously designed assume that the assignment of tasks can be done in any arbitrary manner and does not depend on the energy resources. In this work, we have emphasized that the number of nodes in a cluster is fundamentally related to the energy requirement of the objective. With the help of this new algorithm, we minimize energy consumption in a cluster by improving the mechanism for selecting objective, depending upon the amount of energy present at the nodes of that cluster.

Keywords: SWARM, Pareto Optimality, Cluster Head, Optimality.

1 Introduction

Recent years have witnessed increased interest in the field of SWARM based cooperative communication for establishment and coordination in a Mobile Ad Hoc network. Swarm Communication plays an important role in creating a communication bridge between different independent devices working in a decentralized approach. Most of the current applications of SWARM based communication e.g. aerial swarm deployment [14] and disaster mitigation [13] require minimal energy wastage to maximize life time of the SWARM. Thus, while developing protocols for SWARM Communication, the prime objective is to reduce the energy consumption [6] and improve packet delivery ratio [2].

Clustering provides the flexibility in dividing tasks among multiple nodes to mitigate energy wastage and also to help in formulating stable routing mechanisms eventually resulting in minimal packet loss. A cluster in general constitutes of two types of nodes, a cluster head that forms the backbone of the network and is responsible for controlling a cluster and non-cluster head nodes that follow the instructions of a cluster head. Cluster heads negotiate among each other for a stable and reliable inter cluster routing path, hence whenever, there is a change in cluster head the inter cluster routing path has to be reconfigured, resulting in wastage of energy resources.This change in cluster head is known as

* *Member(IEEE).*

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 133–147, 2011.

re-clustering. Usually, it is fashionable for re-clustering to be performed when the present cluster head lacks the energy requirement for controlling and monitoring the cluster. The frequency of re clustering can be reduced by efficiently selecting the cluster size as the consumption of energy and the size of cluster are directly proportional. For a large sized cluster, the distance between two neighboring clusters will reduce but distance between nodes inside the cluster will increase resulting in extra energy consumption for communicating within the cluster. On the other hand, if size of the cluster is kept smaller, amount of energy required for communicating within the cluster will reduce but the amount of energy required for communicating between any two clusters will increase. Hence, to have a stable cluster, energy consumption of the cluster for inter-cluster and intra-cluster routing should be considered [16].

Along with cluster stability, efficient resource consumption is also an important parameter for SWARM based networks. For a network with multiple objectives and each of them having a different priority, it is of high concern which objective should the cluster execute first and how many cluster should there be on a particular instance. Else, there is an increased high likelihood that a cluster will die out on its way of trying to complete a task that has a lower priority. In order to avoid such a scenario, various techniques and optimization algorithms have been proposed [9] [8] . These algorithms consider priority of objectives and depending upon the geographic location and velocity of the nodes, select which objectives should be executed first. These algorithms are mostly build upon the basic Particle Swarm Optimization (PSO) [4] technique. They help in finding the optimal objective but do not provide a mechanism as to how these objectives should be achieved in an energy and time efficient manner. To solve this problem, we propose a new framework that focuses on improving energy efficiency and stability of the cluster by forming clusters based on the energy requirement per cluster.

The paper is further divided as: Section II explains some related work in this field, Section III defines the problem statement, Section IV gives the simulation and analysis of the algorithm and Section V gives the conclusion.

2 Related Work

For making a stable cluster, a clustering scheme should be devised that forms clusters based on some stability constrained mechanism. In [5], mobility of nodes is considered as cluster creation mechanism. When two cluster heads come close to each other, the node having the lower relative mobility and who is also close to all the cluster members becomes the new cluster head. Likewise in [7], nodes form clusters based on velocity constraint. Nodes having similar velocity form up a single cluster. Both these techniques lack in stability in a way that they do not select a stable cluster head during the initial cluster head selection procedure hence this results in frequent need for re-clustering and hence wastage of resources.

In [1], node mobility is considered as the constraint for the cluster head selection. In this algorithm every node monitors its relative speed with its neighbor

and selects a node that has the lowest speed to be its head. Likewise, in [4] speed entropy is used to accumulate the nodes that are in close proximity to each other. Among these nodes, the node having the maximum energy is selected as the cluster head. These algorithms improve cluster stability but face convergence issues i.e., as relative speed measured on a specific node has local significance, hence different nodes have different measurements. This hinders the algorithm from reaching a unanimous decision.

The Size-bounded Multi-hop Clustering (SMC) [15] considers the local topology information. The node elects a cluster head depending on a relative weight value. This weight value decreases as the number of hops from the cluster head increase. This surely restricts the size of the cluster but does not guarantee stability and does not cater for unstable nodes.

The Affinity propagations for vehicular network (APROVE) [12] uses the Affinity Propagation Algorithm to form cluster of nodes having minimum distance and minimum difference of speed. Likewise the Mobility and Energy Aware Clustering Algorithm (MEACA) [16] consider node's mobility and energy into consideration for selecting the cluster head. In this algorithm, the energy and mobility factor of each node is advertised to its immediate neighbor. Nodes receive the advertised values and form a table on their basis. Node having most energy and least mobility is selected as the cluster head. Both these algorithms improve stability of the cluster but they ignore the importance of priority of the objectives and the energy required by each objective.

Multi objective Particle Swarm optimization (MOPSO) and Game Strategies Multi-Objective Particle Swarm Optimization (GMOPSO) both include the importance of the optimality of an objective. In these algorithms the objective that is close to the current position of the node becomes the optimal objective of that node [8]. This also helps giving high priority to a specific node and low priority to some other node, but these algorithms lack in determining the energy requirement for accomplishing those objectives.

Some of these algorithms improve stability but to the best of our knowledge algorithms that assure the completion of objectives with minimal energy consumption are almost non existing. The existing literature mainly concentrates on stability of the cluster for maximum time. A specific cluster may totally run out of power before performing a particular objective because the cluster formation algorithm did not consider this parameter.

3 Problem Statement

In SWARM based Ad Hoc Network groups of nodes put in a joint effort in completing a set of tasks. These nodes have limited energy and time resources and are required to complete these tasks in the available resources [14]. The geographic location of tasks can be either predefined or dynamically set. For the scope of this work we consider that the location is fed in the nodes before they are deployed into the test area. Among these tasks, the nodes dynamically select the sequence in which the tasks have to be completed. To have a stable and

organized movement towards these tasks, these nodes make up clusters. Each cluster then selects its own head. The cluster head changes over time depending upon the circumstances defined, in our case being the energy constraints. Different algorithms are used to introduce fairness in the selection of task per cluster, e.g. Pareto Optimality, Multi Object particle Swarm optimization etc. These algorithms help all the nodes decide which path to move on , depending upon the geographically closest objectives location, relative to all the nodes in that cluster. Various algorithms have been proposed that consider selection of path and which also design special metrics for forming cluster, but to the best of our knowledge none of them have considered the relation between the total energy of the cluster and the energy required for completing the task that cluster has selected as its prime objective. When a cluster inefficiently selects a tasks which it cannot complete with the available energy then the cluster will perish on the way towards that task, resulting in wastage of energy. To avoid this problem an algorithm should be devised that considers the relation between the energy of the cluster and the energy required to complete the task.

4 EPAC

In this section we present the details of the proposed clustering algorithm called EPAC(Energy and Path Aware Clustering). In our algorithm a node in a cluster can be in one of the five possible states: Initial (I), Cluster Head (CH) , Gateway Node (GN), Initial Cluster Head (ICH) and Ordinary Node (ON) .An *I* state is when the node has not heard from any of its neighbors and has yet to become a part of any cluster.A *CH* is at the center of the cluster and it communicates with all the nodes inside the cluster. The *GN* communicates with multiple cluster heads of different cluster and also relays messages coming from one source cluster to another destination cluster. A *ICH* is a node that becomes the cluster head at the start of the algorithm and calculates all the paths. An *ON* is a node inside a cluster and is not the GN or the CH and only communicates with the cluster head of the cluster which it is a part of.

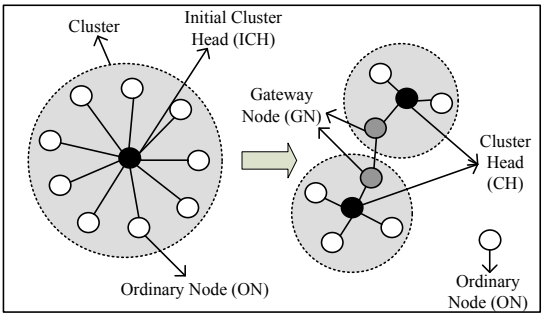


Fig. 1. Cluster Name Representation

CHs and *GNs* both forward messages hence this leads to more energy consumption of these nodes compared to *ONs*. We use path based clustering in order to reduce the energy consumption by forming intelligent clusters to retain the same *CHs* and *GNs* for a longer time period. Our clustering technique works in two phases. In the first phase a single cluster is formed from all the nodes that are present in a close by location and the head of this cluster calculates all the possible paths and then forms sub-clusters based on the number of objectives and the amount of energy present in the cluster. This phase is called the Initialization phase. In the second phase clusters are formed based on the path that cluster has been assigned and the number of nodes that will be required. This phase is called the Clustering phase.

4.1 Initialization Phase

At the start of the algorithm all the nodes are in the initial phase. These nodes broadcast a Hello message based on the CSMA/CA scheme. These Hello messages contain node *id* and the remaining energy E_r [10]. Node that has a locally lowest ID, broadcasts a CHR. On receiving the CHR message, all the nodes with the highest ID become the ONs. Any node that finds itself as having a lower ID will also broadcast a CHR with its *id*. If the nodes sending the CHR message receives another CHR then it compares its own ID with the ID of the message received. If it finds its own *id* to be less, it becomes an ON. If the node sending the CHR message does not receive any other CHR message for two broadcasting time intervals, it sets itself as the ICH. Once the ICH has been decided then the ICH starts calculating the possible paths depending on the amount of energy available in the form of number of nodes.

4.2 Path Computation Phase

The objective of this phase is to have maximum paths covered within the amount of energy present with all the nodes in the network. Among all the set of paths that can be covered in the same amount of energy, the set carrying the most parallel paths is the most preferable set. This is because the more the parallel paths the less will be the average amount of time required per task.

We model this scenario using graph theory. Consider a graph represented by G=G(V,E), where, tasks are represented by the vertices V and the paths, represented by edges E.

The graph shown in Fig 2 is a completely connected graph of the six given tasks. The priority of the nodes is represented by weights of the tasks(higher weights represent higher priority and lower weights represent lower priorities) and the edges represent the distances. Algorithm begins with ideal case of considering direct paths from starting point to all the tasks individually. Sum of

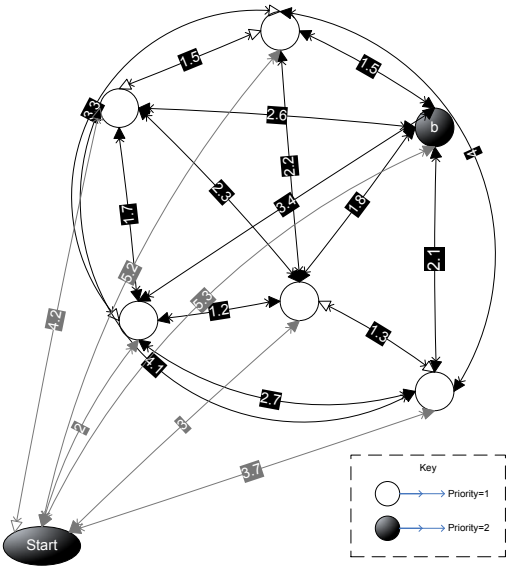


Fig. 2. A Graph Completely Connecting Six Target Locations

these distances will provide energy requirement for completing these tasks in parallel. Energy required for executing task i located at distance d is given by:

$$E_{t,i} = (\frac{4kd(nT_x + (n-1)R_x) + Comp + nE_d}{3}) \tag{1}$$

Here T_x represents the energy required for transmitting the signal from the non cluster head nodes to the cluster head and R_x represents the energy required for receiving the signal transmitted by the non cluster head nodes to the cluster head, Comp represents the energy required for the clustering process, E_d represents the energy required by each node for traveling a distance d and n represents the number of nodes. The sum of the energies required for completing all the paths is represented by $E_{t,t}$. If we assume that all the nodes contain an equal amount of energy then the total energy of all the nodes is represented by $E_{n,t}$. After the calculation of paths, we can have three different scenarios discussed below.

1. $E_{n,t} = E_{t,t}$ In the case of equality we prefer using the parallel paths for completing the tasks in the minimum time possible. Fig 3a shows the direct paths to all the tasks from the original start point where the initialization phase took place.
2. $E_{n,t} < E_{t,t}$ In this case, total energy of nodes is less than the energy required for completing tasks in parallel. In such a scenario, difference of the distances is calculated and the paths having the minimum priority and maximum weights are removed. These weights are added togather and compared with

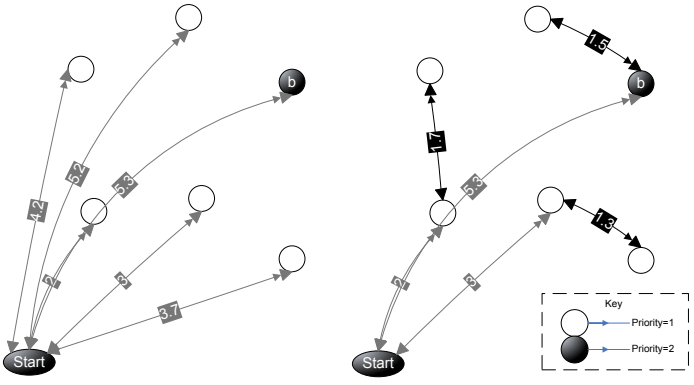


Fig. 3. a)Energy of Nodes = Energy of Paths, b)Energy of Nodes < Energy of Tasks

the difference between $E_{n,t}$ and $E_{t,t}$. When this weight increases the size of the difference then the tasks whose paths were removed are reconnected to the nearest task and the weights with the new connections are computed again. This algorithm repeats itself till the time.

$$E_{n,t} > E_{t,t}$$

Fig 3b shows that depending on energy of the nodes there are three parallel paths. Objectives that are geographically close to each other are laid on the same path and objectives that are away from each other are laid on a separate paths.

3. $E_{n,t} > E_{t,t}$ In this case the total energy of the nodes is larger than the energy required for completing the tasks in parallel. In such a scenario, since we have extra energy with the nodes hence we use it by placing extra nodes on a new self generated path. This path will be between two paths leading to two different tasks. The purpose of these specially designed paths will be to reduce energy required by one *GN* to transmit data to another cluster, by placing nodes that act as relay devices only. These nodes will help multi-hop communication between them and save energy of the gateway nodes present inside each cluster. As energy will increase, these relay paths will also increase till the time all the clusters are at one hop distance to each other.Fig 4 shows that there are 6 parallel paths on which the clusters will travel and the arrows in between the paths show the path where the relay nodes will travel. These paths as shown in the figure are in the center of the parallel paths.

Cost function for the energy of a path. The cost function of path energy is dependent on the energy consumed by initial clustering *Comp* and the energy required for the transmission and reception of the data signals T_x and R_x.The *Comp* function is defined by:

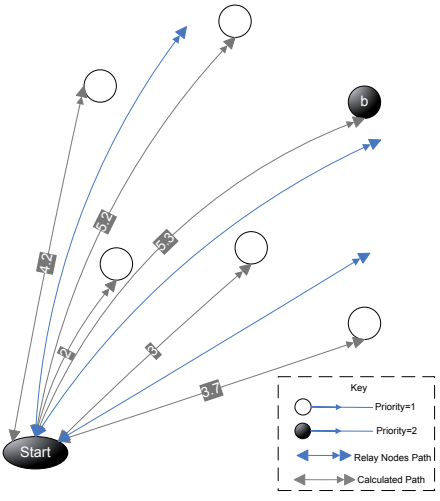


Fig. 4. Energy of Nodes > Energy of tasks

$$Comp = aE_cu + (N - a)E_{no-cu} \tag{2}$$

Here E_{cu} represents the energy consumed by the cluster head in forming a cluster and E_{no-cu} is the energy consumed by the non cluster head nodes in forming a cluster.

$$E_cu = b_l(2E_(elec) + E_(mp)M^4 + E_(fs)d^2) \tag{3}$$

$$E_{no-cu} = b_l((a+2)E_(elec) + E_(fs)d^2) \tag{4}$$

Here E_{elec} represents the energy required per bit and E_{fs} represents the energy per bit per unit distance. Number of bits in a packet are represented by b_l. E_{elec} and E_{fs} depends upon the environmental conditions in which the system is operating.

4.3 Clustering

Once paths have been calculated ICH starts forming sub-clusters. Number of these sub-clusters depends upon the number of parallel paths calculated. ICH uses available data to declare the authorities of a CH to a node having the highest energy. The new CH transmits a signal having the parameters $(id, E_r, P_i(d))$. This message initiates a three way handshake. The nodes receiving this message reply with a *join* message and then the cluster head replies with an accept message along with $(id, E_r, P_i(d))$. Whenever a new node joins the value of $P_i(d)$ is reduced. This informs the surroundings that a new node has entered the cluster. When the cluster has fulfilled its needs then the value of $P_i(d)$ is set to zero showing that there is no more need of any new node. To ensure fair division of nodes among all the selected paths Ant Colony Optimization algorithm is used.

4.4 Ant Colony Optimization

In ant colony optimization (ACO), an artificial colony of ants is used to construct solutions guided by the heuristic information and the pheromone trails. ACO is inspired by the foraging behavior of the real ants. Initially, ants explore surrounding area of the nest in a random manner [3]. When they find a food source, they evaluate quality and quantity of the food and carry it back to the nest. On the way back they drop a pheromone trail on the ground whose quantity and quality guides the other ants towards the food source. This behavior of ants helps them find the shortest path to the food sources from their nests. If the $n_i(d)$ is the number of nodes on path i at distance d then the total number of nodes is

$$m = \sum_{i=1}^{j} n_i(d) \tag{5}$$

Now the intensity of trail represents number of nodes that are acting as cluster members. As number of non cluster node increases and energy level reaches close to energy required for the path, intensity of the trail also reduces. Intensity of the trail is represented by

$$\tau(d+1) = \tau(d) - \delta\tau(d, d+1) \tag{6}$$

Here $\tau(d)$ represents the maximum energy required for that particular path and $\delta\tau(d, d+1) = \sum_{k=1}^{m} \delta^k(d, d+1)$ represents sum of energies of nodes that have already joined the cluster. As number of nodes inside the cluster increase, the requirement of energy decreases hence the intensity of that cluster decreases. The *visibility* in ant colony optimization is the inverse of the distance of that particular path. The visibility $\eta_i = 1/d_i$. Now the number of nodes that a particular cluster will have after a distance d will be

$$p_i(d) = \frac{\tau_i(d)^\alpha \eta_i(d)^\beta}{\sum_{j=1}^{m} [\tau_j(d)]^\alpha [\eta_j(d)]^\beta} \tag{7}$$

4.5 Choosing the Gateway Node

Gateway node selection also depends upon the selection of the path. A node will only become the gateway node if it has the sufficient energy and the suitable location. As all the cluster heads are aware of the paths that were originally calculated by the ICH, hence considering the path and the average speed of the nodes the cluster head can calculate the expected location of the neighboring cluster. This calculation helps select the node that is best suitable for sending the data in that direction. Hence the gateway node selection will also vary over time. The amount of power required to send data to the neighboring node will also be calculated based on the location.

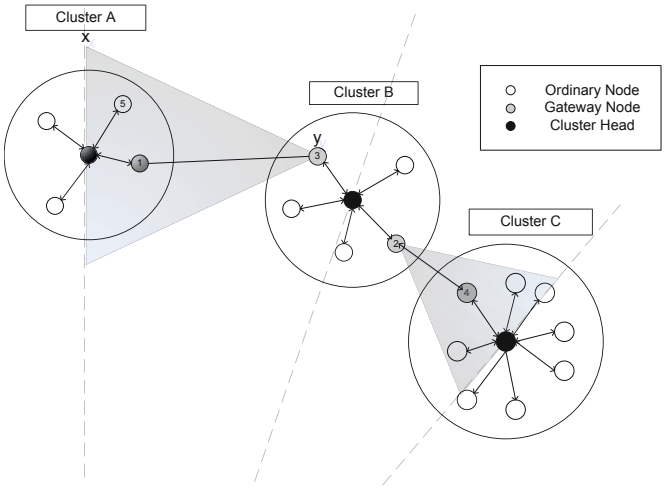


Fig. 5. Gateway Power Approximation

Fig 5 shows that the CH of Cluster B expects the Cluster A to be inside the left sided shaded region.If the distance between point y in cluster B and point x in cluster A is d, power required by the GN of cluster B to convey the information to cluster A in free space will be [11].

$$P_t = \frac{S_r(4\pi)^2 d^2}{G_r G_t \lambda^2} \tag{8}$$

Here S_r is the receiver sensitivity, G_r and G_t are the receiver and transmitter gain respectively and λ is the wavelength the system is working on. Suppose the transmission time is T then the sending energy will be $W_t = P_t T$. Equation 8 can be re-written as.

$$W_t = S_r T d^2 k, k = \frac{(4\pi)^2}{G_r G_t \lambda^2} \tag{9}$$

Now in this case cluster B will require W_t amount of energy to transmit data to cluster A. Likewise, as shown in Fig 5, the distance between cluster B and cluster C is less, hence, it will require less energy to transmit data from cluster B to cluster A. This adaptive mechanism will reduce the energy consumption of GN as GN will not be required to keep a constant connection with all the neighboring clusters.

5 Simulation Results

We conducted simulation experiments to provide a performance comparison of our framework with APROVE [12] and MEACA [16]. We take a test area of

$1000x1000m^2$ and consider a random uniform distribution of 50 tasks over the whole test area. Initially, it is assumed that each mobile node contains 10 joules of energy. For averaging results we have taken 10^7 iterations for each task. The number of mobile nodes varies from 5-200.

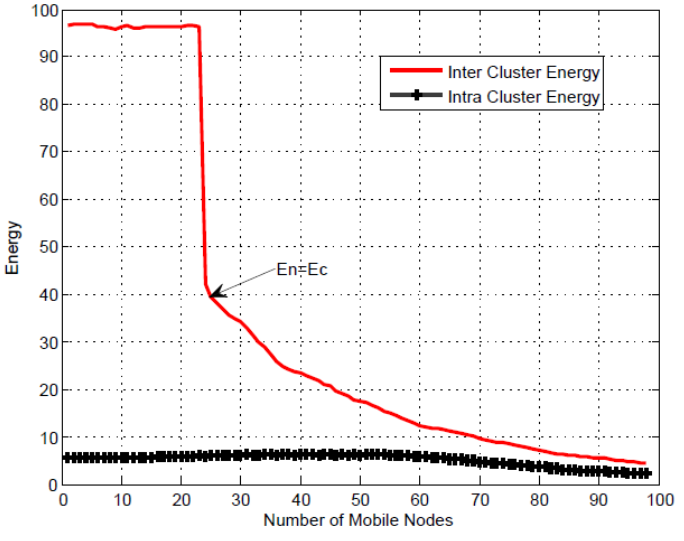


Fig. 6. Energy requirement for Inter-Cluster and Intra-cluster Communication

Fig 6 shows variation in energy required for inter-cluster and intra-cluster communication, based on the amount of resources present. The intra-cluster communication remains the same for different number of mobile nodes. This is due to the fact that all nodes inside a cluster are at a one hop distance. Thus, nodes can communicate while using a low amount of energy. On the other hand, the energy required for the inter-cluster communication varies with the amount of mobile nodes present. Initially, as the total energy of the nodes is less than amount of energy required for covering the paths in parallel, hence different paths are formed each having multiple tasks appended together. Hence, more amount of energy is required for inter-cluster communication. At a point when number of mobile nodes are 25, energy required for completing paths in parallel is equal to total energy of the nodes . After this point, the amount of available energy increases the required amount of energy, hence relay paths similar to those shown in Fig 3b are deployed. This causes the graph to decrease after the point where the amount of mobile number of nodes equal 25. The amount of relay paths increase as the number of extra nodes (that are not required by the cluster to complete the task) increases.

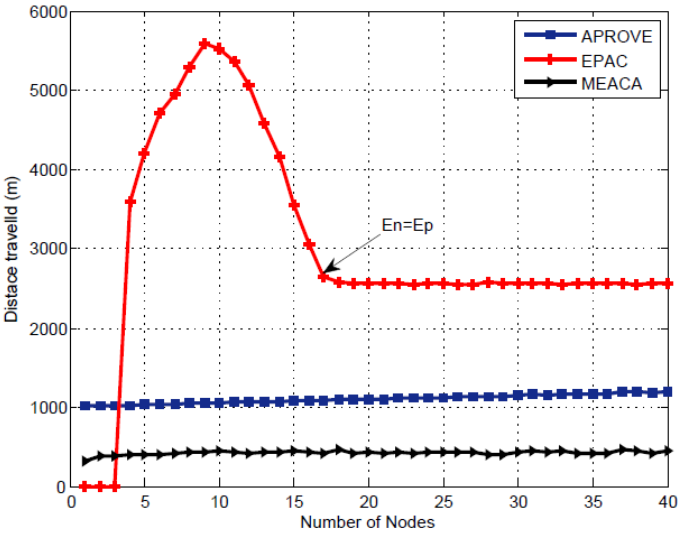


Fig. 7. Average maximum distance covered by different protocols

Fig 7 shows the difference in the average distance covered by each node under different algorithms. In EPAC, average distance covered per node increases for an increase in number of nodes from 0 to 12 because at this time the amount of energy required at the nodes is not sufficient to handle all the tasks in parallel. The tasks are executed in serial manner as shown in Fig 3b. As the number of nodes increases from 12 to 17, the average distance covered per node decreases as the amount of energy available at the nodes becomes sufficient enough to handle more and more tasks in parallel. After the number of nodes crosses 17, the average distance covered per node becomes constant, since, from this point onwards the E_n exceeds E_p and all the extra nodes start following the self generated paths as shown in Fig 4. On the other hand, when APROVE is executed, covered distance increases linearly as the number of nodes increase, but this increase is of a very low gradient. This is because APROVE does not consider the relation between amount of energy required for completing a task and amount of available energy. Likewise MEACA shows a linear behavior, the amount of distance covered remains same for any amount of nodes present.

Fig 9 shows the difference in the average maximum amount of clusters required to complete 50 tasks. In EPAC, number of clusters formed for less than 12 nodes increases as E_c is less than E_p. Thus, after performing each task, re-clustering is required. This results in increase in number of clusters. After the point when E_c is equal to E_p, number of clusters become stable as all tasks here onwards are completed with similar patterns. On the other hand, number of clusters in APROVE and MEACA increase exponentially. The reason behind such a behavior is that when a new node is added into a network operating under APROVE, all the nodes with the same velocity and in a close proximity to each

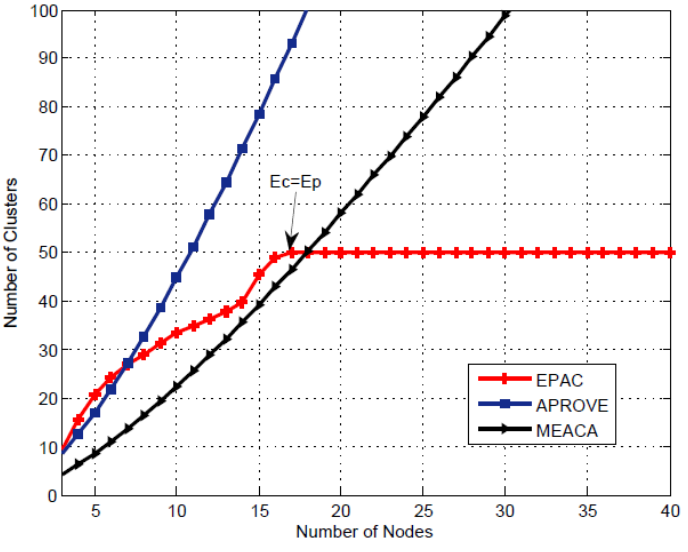


Fig. 8. Average maximum amount of clusters required

other join the same cluster resulting in excessive energy utilization of the cluster head for handling the cluster hence provoking extra re-clustering. Likewise the amount of clusters required to complete maximum tasks in MEACA increases as new nodes are entered into the network. Due to the addition of the new nodes size of a cluster increases as more nodes come into close proximity of each other, resulting in spending extra energy in controlling the cluster.

These results show that when the number of clusters is made dependent on number of objectives provided, we may have more efficient use of resources. The proposed EPAC algorithm gives us a more stable cluster by judicious use of energy for inter-cluster communication and undertaking clustering based on available energy. This algorithm will reduce the chance of wastage of energy by not selecting a task that has less priority or a task for which it does not have sufficient energy. This algorithm will also remove the chance for unnecessary network partitioning due to re-clustering since once a cluster has started moving towards a specific task then for re-clustering, a new cluster head will be selected from the nodes that are moving towards the same destination point. The new cluster head will not have to decide a new destination, hence network partitioning is intelligently avoided due to re-clustering. This algorithm also support parallel execution of tasks, hence it reduces the time consumed for the completion of the tasks.

6 Conclusion

In this paper a cluster head selection algorithm based on the amount of energy required per objective is proposed for mobile ad hoc networks. It produces

energy efficient and stable clusters that cover more tasks in parallel to reduce the time required to complete all the tasks. It also reduces the chance of network partitioning due to re-clustering or node failure by forming clusters strong enough to complete a task on its own. Clusters are only given tasks that they can handle, this reduces the chance of re-clustering. Hence by reducing the energy wastage, the clusters formed are more stable compared to the previously developed protocols.

References

1. An, B., Papavassiliou, S.: A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks. International Journal of Network Management 11(6), 387–395 (2001)
2. Chiang, C.C., Wu, H.K., Liu, W., Gerla, M.: Routing in clustered multihop, mobile wireless networks with fading channel. In: Proc./E SlCON, vol. 97, pp. 121–197 (1997)
3. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. IEEE Computational Intelligence Magazine 1(4), 28–39 (2006)
4. Gu, B., Hong, X.: Mobility identification and clustering in sparse mobile networks. In: IEEE Military Communications Conference (MILCOM 2009), pp. 1–7. IEEE, New York (2010)
5. Gunter, Y., Wiegel, B., Grossmann, H.P.: Cluster-based medium access scheme for vanets. In: IEEE Intelligent Transportation Systems Conference (ITSC 2007), pp. 343–348. IEEE, New York (2007)
6. Kawadia, V., Kumar, P.R.: Power control and clustering in ad hoc networks. In: Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003), IEEE Societies, vol. 1, pp. 459–469. IEEE, New York (2003)
7. Kayis, O., Acarman, T.: Clustering formation for inter-vehicle communication. In: IEEE Intelligent Transportation Systems Conference (ITSC 2007), pp. 636–641. IEEE, New York (2007)
8. Li, Z., Liu, S., Xiao, D., Chen, J., Li, K.: Multi-objective particle swarm optimization algorithm based on game strategies. In: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp. 287–294. ACM Press, New York (2009)
9. Mishra, S.K., Panda, G., Meher, S.: Multi-objective particle swarm optimization approach to portfolio optimization. In: World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), pp. 1612–1615. IEEE, New York (2010)
10. Nasim, M., Qaisar, S.: Hierarchical MIMO: A clustering approach for Ad hoc wireless sensor networks. In: 44th Annual Conference on Information Sciences and Systems (CISS 2010), pp. 1–6. IEEE, New York (2010)
11. Rappaport, T.S.: Wireless communications. Prentice Hall PTR, New Jersey (2002)
12. Shea, C., Hassanabadi, B., Valaee, S.: Mobility-based clustering in VANETs using affinity propagation. In: IEEE Global Telecommunications Conference (GLOBECOM 2009), pp. 1–6. IEEE, New York (2010)

13. Stirling, T., Floreano, D.: Energy Efficient Swarm Deployment for Search in Unknown Environments. In: Dorigo, M., Birattari, M., Di Caro, G.A., Doursat, R., Engelbrecht, A.P., Floreano, D., Gambardella, L.M., Groß, R., Şahin, E., Sayama, H., Stützle, T. (eds.) ANTS 2010. LNCS, vol. 6234, pp. 562–563. Springer, Heidelberg (2010)
14. Stirling, T., Floreano, D.: Energy-Time Efficiency in Aerial Swarm Deployment. In: Proceedings of the 10th International Symposium on Distributed Autonomous Robotics Systems (2010)
15. Wang, Y., Medidi, M.: A Distributed Size-bounded Multi-hop Clustering for Mobile Ad Hoc Networks (2006)
16. Wu, Y., Wang, W.: MEACA: Mobility and Energy Aware Clustering Algorithm for Constructing Stable MANETs. In: IEEE Military Communications Conference (MILCOM 2006), pp. 1–7. IEEE, New York (2007)

Employing Energy-Efficient Patterns for Coverage Problem to Extend the Network Lifetime

Manh Thuong Quan Dao, Ngoc Duy Nguyen, Vyacheslaw Zalyubovskiy, and Hyunseung Choo

School of Information and Communication Engineering
Sungkyunkwan University, Korea
dmtquan@skku.edu, duyngoc30@skku.edu, slava@ece.skku.ac.kr, choo@skku.edu

Abstract. The coverage problem is a fundamental issue in wireless sensor networks. It has attracted considerable attention recently. Most node scheduling patterns utilize the adjustable range of sensors to minimize the sensing energy consumption, and thus extend the network lifetime. However, a large source of the consumption of the sensor communication energy is not strictly taken into account. In this paper, we introduce two energy-efficient patterns that are used to minimize the communication energy consumption of a sensor network, and simultaneously, maintain a high degree of coverage. Moreover, the proposed patterns have a structure that is easy to design and apply to practical applications. Calculations and extensive simulation are conducted to evaluate the efficiency of the new patterns compared to existing ones in terms of various performance metrics.

Keywords: coverage problem; node scheduling; energy-efficiency; wireless sensor networks.

1 Introduction

Modern technology has enabled a new generation of wireless sensor networks that are feasible for a wide range of commercial and military applications. A wireless sensor network (WSN) is composed of a large number of autonomous sensors that are densely deployed into a target sensing field to monitor physical phenomena of interest [1]. Replenishing power resources is a difficult and impossible task in most cases, since each sensor has a limited power battery. Energy-saving optimization is an important criteria to evaluate the success of WSNs. Recent analysis [4] shows that each sensor uses a large portion of power for communication. Therefore, this paper considers the problem of how to minimize the communication energy consumption by adopting two strategic coverage patterns.

A fundamental issue in WSNs is the coverage problem, that concerns how well the target sensing field is monitored or tracked by sensors. In this paper, we consider the full coverage problem, in the sense that every point in the target sensing area is covered by at least one sensor. On the other hand, there are two mechanisms in sensor deployment: deterministic deployment and random deployment. In deterministic deployment,

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 148–158, 2011.

a sensor can be placed exactly at a defined position in the target sensing field. In contrast, the position of a sensor is not known a priori in random deployment. Our proposed strategic patterns are useful in deterministic deployment as well as random deployment. Another important consideration in WSNs is the node scheduling problem, as it has a significant impact on extending the network lifetime. A node scheduling mechanism operates, so that a set of active sensor nodes is selected to work in a round, another random set is selected in another around, as long as the coverage goal is met [12]. The results of this paper can be used as a guideline to select active sensors in each round, so that two conflicting goals are satisfied simultaneously: minimizing the communication energy consumption and keeping a high degree of coverage.

As shown in [4], power consumption can be divided into three domains: sensing, communication, and data processing. Recent coverage patterns only consider how to minimize the overlapped sensing areas of sensors, and thus optimize the sensing energy consumption of WSNs. In [3,12,14] the authors utilize the adjustable sensing range of sensors to achieve significant improvement in coverage efficiency. Recently, the authors in [14] proposed the optimal sensing energy patterns using two adjustable sensing ranges. These patterns outperform the existing ones with respect to various performance metrics. However, none of the previous work considered patterns that minimized communication energy consumption. Communication tasks are indispensable since sensors have to forward and receive data from other sensor nodes hop-by-hop. Communication energy consumption in WSNs becomes more critical in applications that require high rate data collection in real-time such as [10] and [8]. Therefore, communication energy consumption should be taken into account when designing energy-efficient coverage patterns.

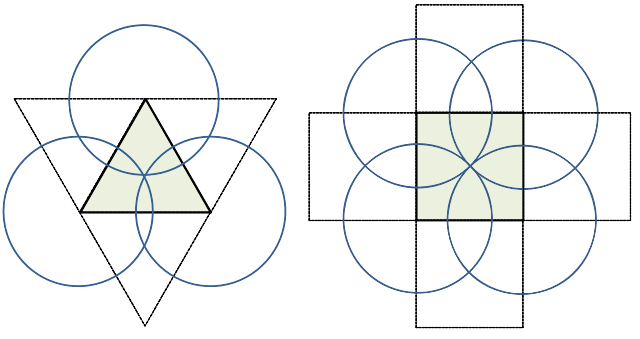
In this paper, we also propose a node scheduling pattern, but we concentrate on minimizing the communication energy consumption that is the most important resource of WSNs. In summary, this paper makes the following key contributions:

- Novel patterns are constructed, so that they can be used in deterministic deployment, as well as in random deployment. In deterministic deployment, the patterns become a strategic plan to design an efficient WSN. In cases where sensors are randomly deployed, our proposed patterns aid a node scheduling mechanism to select active nodes in each round.
- All the patterns have structures that are easy to implement and design due to their simplicity. This makes our patterns practical.
- To the best of our knowledge, when considering an energy-efficient pattern for the coverage problem, we provided the best patterns in terms of communication energy consumption.

The remainder of this paper is organized as follows. In section 2, we discuss related work. Section 3 presents the system model and assumptions while Section 4 presents our proposed patterns. Section 5 shows the performance evaluation result. Finally, we conclude our work in section 6.

2 Related Work

A survey on the energy-efficient coverage problem is researched by Cardei and Wui [4]. The paper summarizes various problems on coverage area as well as their corresponding solutions. One of the mechanisms to reduce the redundant energy is using a node scheduling strategy. In this strategy, the network is scheduled to operate in turn, in the sense that one set of sensors is selected to monitor fully the entire target sensing field, and another set will be selected at another time, after the current set of sensors goes into a dormant state. Coverage ratio is one of the measurements of the system's quality of service. It plays an important role in evaluating whether or not a WSN topology is good. There is always a lower bound of coverage ratio. If the coverage ratio falls below this threshold, the network may not operate correctly. Therefore, the major challenge for the success of WSNs is designing an energy-efficient model that provides a high degree of coverage.

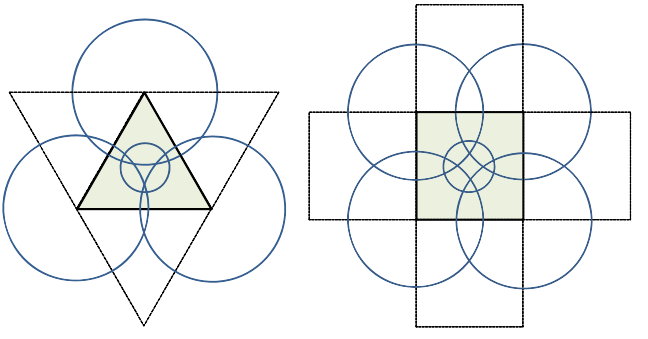


(a) Pattern A1. (b) Pattern B1.

Fig. 1. Coverage patterns with uniform sensing range

Another approach to extend the network lifetime is using sensors with adjustable ranges. In [12], two node scheduling patterns are proposed to reduce the sensing energy consumption of WSNs. The authors construct the patterns based on regular polygons-tiles that cover the entire target sensing field without overlap. As shown in Fig. 1, pattern A1 is based on a regular triangle tile and pattern B1 based on a square tile. The authors in [14] introduce the concept of coverage density that is used as a standard metric to evaluate the efficiency of a pattern in terms of coverage efficiency. Thus, they proposed the optimal sensing energy consumption patterns, using two adjustable sensing ranges, as shown in Fig. 2. The ratio between the large disk's sensing range and small disk's sensing range of patterns A2 and B2 are $\sqrt{31}$ and $\sqrt{5}$, respectively. These patterns have been shown to outperform prior ones with respect to various performance metrics. Similar to [3,12], and [14], we consider the problem of energy-efficient area coverage patterns. However, this paper supplements the important limitation of previous studies by introducing patterns that are considered to be the best among the existing ones in terms of communication energy consumption.

The analysis of the power usage for the WINS Rockwell seismic sensor shows that the power usage for communication is between 0.74 W and 1.06 W, for the idle state it is 0.34 W, for the sleep state 0.03 W, and for the sensing task 0.02 W [9]. Thus, the communication energy consumption is much higher than the sensing energy consumption. Recently, more applications such as monitoring industrial processes, geophysical environments, and civil structures (buidings, bridges, etc), require high-data rate signals [10, 8]. A key challenge in those applications is how to collect efficiently those fidelity data subject to limited radio bandwidth and the battery of sensors. Therefore, finding an energy-efficient pattern in terms of communication energy consumption is crucial in such those applications.



(a) Pattern A2. (b) Pattern B2.

Fig. 2. Optimal coverage patterns with 2 adjustable sensing range

In [2], Bai et al. propose deployment patterns to achieve full coverage with three-connectivity and full coverage with five-connectivity, under different ratios of sensor communication range over the sensing range for WSNs. The authors in [15] and [13] consider the *k-coverage* problem with any arbitrary sensing shape and find the weak sub-regions degrading the overall coverage performance. Various related coverage problems have been discovered recently. Examples include the coverage problem in three dimensional space [6], coverage for estimating localization error [11], and barrier coverage problem [13].

3 System Model and Assumptions

3.1 System Model

In this paper, we assume that the sensor nodes are randomly deployed in a two-dimensional target sensing field, where each node uses a Global Positioning System or a localization scheme to knows its position. The sensing area of each sensor is a disk of a given sensing range. The sensors are in charge of monitoring a target sensing field which is assumed very large compared to the sensing area of a sensor, and thus we can ignore the boundary effect of the target sensing field.

We construct a minimal spanning tree among active nodes when calculating the communication energy consumption. Each node adjusts its communication range to the farthest node on the tree to guarantee network connectivity. We also assume that the energy consumed by communication for a sensor is proportional to the square of the distance from itself to its farthest node in the minimum spanning tree.

Finally, since our sensor deployment is random deployment, we may not find a sensor that has the exact location in the pattern. In this case, we select a sensor that is the closest to the ideal position in the corresponding pattern. Similar to previous studies, we construct the patterns based on a regular polygons-tile that cover the whole target sensing field without overlap. We also suppose that all tiles are covered in the same manner. Sensors are placed at the vertices of the polygons and the circles represent the sensing areas of sensors. All patterns proposed in [12] and [14] have the structure of pattern A1 or pattern B1. Therefore, we select these two basic patterns as the starting point of our procedure to construct the new patterns.

3.2 Important Definitions

Before going into the details of the proposed patterns, we introduce three important metrics that are used in [14] to compare the efficiency among patterns in terms of coverage density and communication energy consumption.

Definition 1. Coverage density (D) is the ratio of the total area of the parts of disks inside the tile divided by the area of the tile. Given a coverage model as in Figure 3, the coverage density $D = \frac{S_1+S_2+S_3}{S_{I_1I_2I_3}}$, where S_1, S_2, and S_3 denote the areas of parts of disks of sensors I_1, I_2, and I_3 inside the tile, respectively; $S_{I_1I_2I_3}$ denotes the area of the triangular tile.

Definition 2. Sensing energy consumption per area (SECPA) is the part of the sensors' sensing energy used by the nodes inside a tile divided by the tile's area. We suppose that the sensing energy consumption is proportional to the area of sensing disks by a factor of μ_1, or the power consumption per unit. Then, SECPA is $SE = D.\mu_1$

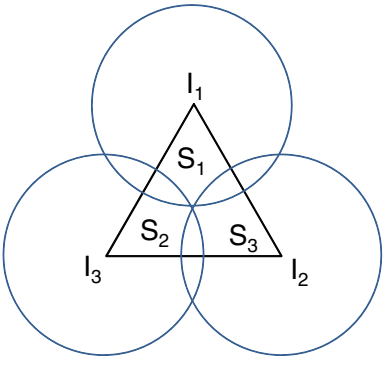


Fig. 3. Coverage Density

Definition 3. Similar to SECPA, communication energy consumption per area (CECPA) is the part of the sensors' communication energy used by the nodes inside a tile divided by the tile's area.

4 Proposed Patterns

4.1 Energy-Efficient Pattern Based on the Hexagonal Tile

As presented in [15], pattern A1 is the optimal topology, in the sense that it provides the minimum number of sensors used to cover fully the entire target sensing field, if all sensors have the same sensing range. However, the communication energy consumption of pattern A1 is considerably high. Therefore, to retain the advantages of pattern A1, and simultaneously improve the communication energy consumption, we construct another pattern, A3, based on a hexagonal structure, as shown in Fig. 4.

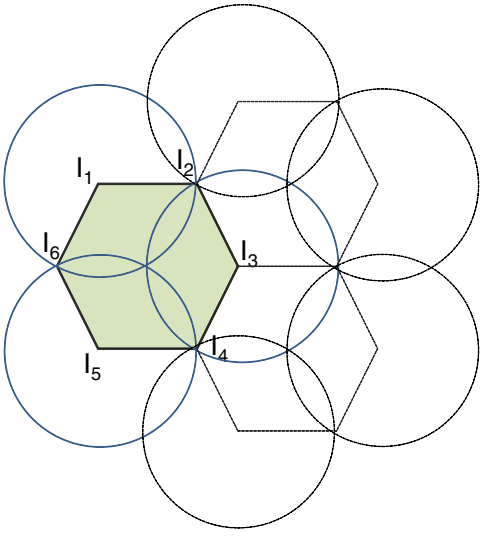


Fig. 4. Pattern A3

As opposed to the previous patterns, we assign sensors to different tasks. In Fig. 4, only three sensors placed at I_1, I_2, and I_5 retain both roles: sensing the monitored field and continuously communicating with other sensors. Conversely, sensors placed at I_2, I_4, and I_6 only need to turn on their communication function. The main reason for this strategy is that the topology of all sensors retaining both roles is the same with the topology of pattern A1. Therefore, the coverage density and the sensing energy consumption are equal to pattern A1's coverage density and sensing energy consumption, respectively. An interesting point we found here is that due to the symmetry, sensors in pattern A3 can take turns to switch on/off their sensing ability. For example, in the first round, sensors at I_2, I_4, and I_6 turn on their sensing ability, while sensors at I_1,

I_3, and I_5 turn off their sensing ability. In the next time slot, sensors placed at I_2, I_4, and I_6 take on the sensing responsibility, while sensors at I_1, I_3, and I_5 can safely turn the function off. Therefore, the energy consumption is more balanced between sensor nodes.

We apply the same method to calculate the coverage density D and sensing energy consumption per area, similar to [14]. For pattern A3, the coverage density and sensing energy consumption are:

$$D_{A3} = 2\Pi/3\sqrt{3} \approx 1.2091$$

$$SE_{A3} \approx 1.2091\mu_1,$$

where μ_1 is the sensing power consumption per unit.

We assume that all sensors are involved in communication and construct a minimal spanning tree that spans all the sensors to calculate the communication energy consumption of this pattern. We assume that the energy consumed by communication for a sensor is proportional to the square of the distance from itself to its farthest neighbor on the tree by a factor of μ_2, where μ_2 is the communication power consumption per unit. We ignore the edge effect and calculate CECPA for the case of infinite grid, as for the estimation of coverage density. The communication energy consumed by each node in each rectangle is $\frac{1}{3}{I_1I_2}^2\mu_2$, since each node contributes to 3 hexagonals. Finally, the CECPA of pattern A3 is calculated as follows:

$$CE_{A3} = \frac{\frac{6}{3}{I_1I_2}^2\mu_2}{S_{I_1I_2I_3I_4I_5I_6}} \approx 0.7698\mu_2,$$

where $S_{I_1I_2I_3I_4I_5I_6}$ denotes the area of the hexagonal tile.

4.2 Energy-Efficient Pattern Based on the Square Tile

According to [14], the patterns based on a square tile are optimal with respect to the communication energy consumption per unit area. Moreover, a rectangular placement grid seems to be more convenient in practice, especially in the case of covering a rectangular area. As shown in [14], pattern B1 has a SECPA that approximates to $1.57\mu_1$, while the CECPA is μ_2. Although the communication energy consumption of pattern B1 is efficient, we can further improve this pattern to lower communication energy consumption.

Based on coverage pattern B1, we construct another pattern, as shown in Fig 5. Similar to pattern A3, we also assign sensors to different roles. Sensors at I_1, I_3, I_5 and I_7 retain both sensing and communication functions, while sensors placed at I_2, I_4, and I_6 only need to turn on their communication function. The coverage density and SECPA of pattern B3 are similar to the coverage density and SECPA of pattern B1, which are $D_{B3} \approx 1.5708$ and $SE_{B3} \approx 1.5708\mu_1$, respectively, because the pattern B3 is similar to pattern B1 with respect to sensing energy consumption.

We also construct the MST that spans all the sensors to calculate the communication energy consumption of pattern B3. Each of the four nodes, I_1, I_3, I_5 and I_7, contributes to four rectangles. Thus, the communication energy consumed by each node in each

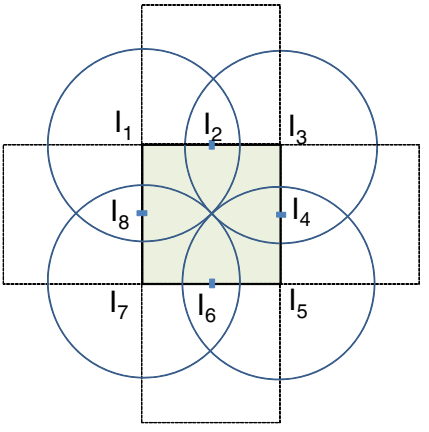


Fig. 5. Pattern B3

rectangle is $\frac{1}{4}I_1{I_2}^2\mu_2$. Each of the four nodes I_2, I_4, I_6 and I_8, contributes to only two rectangles. Thus, the communication energy consumed by each node in each rectangle is $\frac{1}{2}I_1{I_2}^2\mu_2$. Finally, the CECPA of pattern B3 can be calculated using the following equation:

$$CE_{B3} = \frac{\frac{4}{4}I_1{I_2}^2\mu_2 + \frac{4}{2}I_1{I_2}^2\mu_2}{S_{I_1I_3I_5I_7}} \approx 0.75\mu_2,$$

where $S_{I_1I_3I_5I_7}$ denotes the area of the square tile, and μ_2 is the communication power consumption per unit.

Table 1 summarizes the SECPA and CECPA of six patterns. We see that patterns A2 and B2 have the lowest value of sensing energy consumption, whereas patterns A3 and B3 are the best in terms of communication energy consumption. In the next section, we will compare the energy efficiency of all patterns with extensive simulation.

Table 1. Energy consumption per area for different patterns

Type	*SECPA*	*CECPA*
Pattern A1	$1.20\mu_1$	$1.15\mu_2$
Model A2	$1.10\mu_1$	$1.15\mu_2$
Pattern A3	$1.20\mu_1$	$0.7698\mu_2$
Pattern B1	$1.57\mu_1$	μ_2
Pattern B2	$1.17\mu_1$	μ_2
Pattern B3	$1.57\mu_1$	$0.75\mu_2$

5 Performance Evaluation

5.1 Simulation Environment

To evaluate the efficiency of our new patterns (patterns A3 and B3), we compare them to the previous ones: pattern A1 proposed by Zhang et al. [15], patterns B1, A2, and B2

proposed by Zalyubovskiy et al. We use the same simulation environment as in [12,14]. We randomly deployed 1000 sensors in a $50m \times 50m$ area. The sensing range of the large disk, R, varies from 4 m to 12 m. We first construct a minimal spanning tree among the working nodes to estimate the communication energy consumption. We assume that the energy consumed by communication for a working sensor is proportional to n's power of the distance to its farthest neighbor in the tree ($n = 2, 4$). As mentioned earlier, we suppose that we can find a sensor at any desirable position. Since this assumption may not hold in practical applications, we select sensors that are closest to the defined position in our ideal patterns. We use the following metrics to compare the performance of all patterns:

1. Sensing energy consumption per area (SECPA) in one round.
2. Communication energy consumption per area (CECPA) in one round.
3. Total energy consumption per area (TECPA) in one round.

We use the energy cost model as in [16], to estimate the TECPA of the entire network. In this model, the total energy consumption of each working node is calculated by the following formula:

$$E = kS(R)^x + (1 - k)T(R)^y + C,$$

where $S(R)$ and $T(R)$ denote the sensing range and communication range of a sensor, respectively; x and y are constants between 2 and 4; k is a constant, such that $0 \leq k \leq 1$. The energy consumed by the idling radio and processor of each sensor is a constant, C. Similar to [16], we select $x = y = 4$ and $C = 2000$ for our simulation.

5.2 Simulation Results

As shown in Figure 6, the simulation results about the sensing energy consumption are correlated with the theoretical analysis in section IV. Patterns A2 and B2 are the best patterns in terms of the sensing energy consumption. The sensing energy consumption of patterns A3 and B3 are equal to patterns A1 and B1, respectively, as the set of sensor nodes, which turns on sensing ability in patterns A3 and B3, is similar to the set of sensor nodes in patterns A1 and B1.

Figures 7 shows the communication energy consumption of all patterns when the path loss exponent is $n = 2$ and Figures 8 shows the communication energy consumption when the path loss exponent is $n = 4$. In both cases, patterns A3 and B3 are the best patterns with respect to communication energy consumption, since the average distance between neighbor nodes in patterns A3 and B3 is smaller than other patterns. Thus, the communication energy consumption is minimized.

Figure 9 shows the total energy consumption of six patterns. We can see that our proposed patterns A3 and B3 outperform the other patterns in terms of the total energy consumption in most cases. When the ratio $k > 0.7$, the total energy consumption of pattern B3 is not as good as for pattern A2 and A3 due to its high sensing energy consumption. Therefore, pattern B3 is the most preferred in a WSN that has frequent traffic, while pattern A2 and A3 are suitable for WSNs that have low traffic.

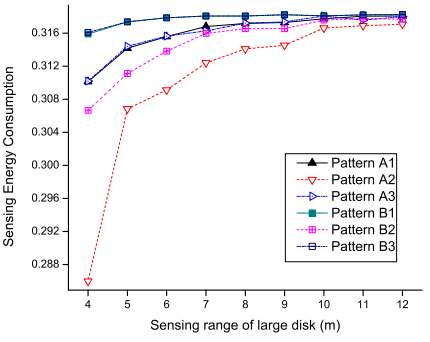


Fig. 6. Sensing energy consumption

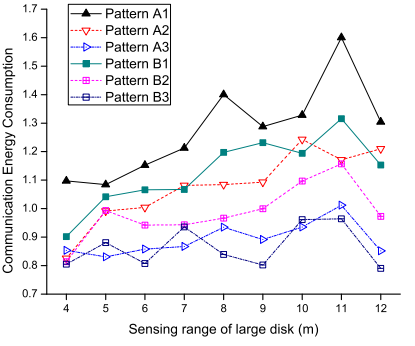


Fig. 7. Communication energy consumption ($n = 2$)

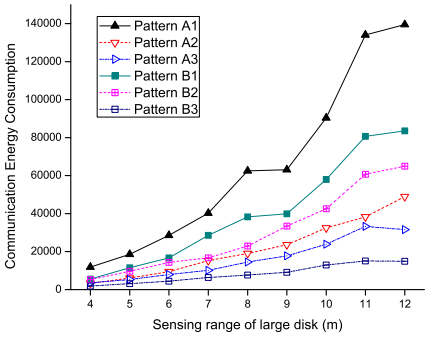


Fig. 8. Communication energy consumption ($n = 4$)

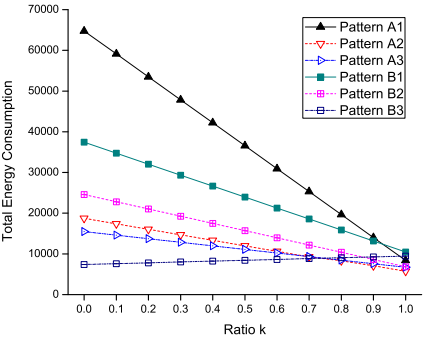


Fig. 9. Total energy consumption

6 Conclusion

In this paper, we constructed energy-efficient node scheduling patterns under the condition of a high ratio of sensing coverage. Our mathematical results and simulation show that our two new proposed patterns A3 and B3 significantly improve energy consumption compared to patterns A2 and B2 that were previously known as the best. Our future research considers scheduling algorithms employed on the proposed patterns and improve the sensing energy consumption as well as minimizing the number of deployed sensors.

Acknowledgment

This research was supported in part by MKE and MEST, Korean government, under ITRC NIPA-2011-(C1090-1121-0008), WCU NRF (No. R31-2010-000-10062-0) and PRCP(2010-0020210) through NRF, respectively.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Communications Magazine 40, 102–114 (2002)
2. Bai, X., Xuan, D., Yun, Z., Lai, T.H., Jia, W.: Complete optimal deployment patterns for full-coverage and k-connectivity $(k \leq 6)$ wireless sensor networks. In: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 401–410 (2008)
3. Cardei, M., Wu, J., Lu, M.: Improving network lifetime using sensors with adjustable sensing ranges. International Journal of Sensor Networks 1, 41–49 (2006)
4. Cardei, M., Wu, J.: Energy-efficient Coverage Problems in Wireless Ad-hoc Sensor Networks. International Journal of Computer Communications, 413–420 (2006)
5. Fan, G.J., Jin, S., Cheng, S.: An Approach to Finding Weak Regions for Wireless Sensor Networks with Arbitrary Sensing Areas. In: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 445–446 (2008)
6. Huang, C.F., Tseng, Y.C., Lo, L.C.: Coverage and Connectivity in Three-Dimensional Underwater Sensor Networks. Wireless Communication and Mobile Computing 8(8), 995–1009 (2008)
7. Meguerdichian, S., Koushanfar, F., Potkonjak, M., Srivastava, M.B.: Coverage problems in wireless ad-hoc sensor networks. In: Proceedings of the 20th IEEE INFOCOM, pp. 1380–1387 (2001)
8. Porta, L., IIIangasekare, T., Loden, P., Han, Q., Jayasumana, A.: Continuous Plume Monitoring Using Wireless Sensors: Proof of Concept in Intermediate Scale Tank. Journal of Environment Engineering 135(9), 831–838 (2009)
9. Raghunathan, V., Schurgers, C., Park, S., Srivastava, M.B.: Energy-Aware Wireless Microsensor Networks. IEEE Signal Processing Magazine 19, 40–50 (2002)
10. Song, W., Huang, R., Shirazi, B., LaHusen, R.: TreeMAC: localized TDMA MAC protocol for real-time high-date-rate sensor networks. In: Proceeding of the IEEE Internation Conference on Pervasive Computing and Communications, pp. 750–765 (2009)
11. Wang, W., Srinivasan, V., Wang, B., Chua, K.C.: Coverage for target localization in wireless sensor networks. IEEE Transactions on Wireless Communications 7(3), 667–676 (2008)
12. Wu, J., Yang, S.: Energy-efficient node scheduling models in sensor networks with adjustable ranges. International Journal of Foundations of Computer Science 16, 3–17 (2005)
13. Yang, G., Qiao, D.: Multi-Round Sensor Deployment for Guaranteed Barrier Coverage. In: Proceedings of the 30th IEEE INFOCOM, pp. 2462–2470 (2010)
14. Zalyubovskiy, V., Erzin, A., Astrakov, S., Choo, H.: Energy-efficient Area Coverage by Sensors with Adjustable Ranges. Sensors Journal 9, 2446–2460 (2009)
15. Zhang, H., Hou, J.C.: Maintaining Sensing Coverage and Connectivity in Large Sensor Networks. International Journal of Ad Hoc and Sensor Wireless Networks, 89–124 (2005)
16. Zhou, Z., Das, S.R., Gupta, H.: Variable radii connected sensor cover in sensor networks. ACM Transaction on Sensor Networks 5(1) (2009)

Cooperative Communication for Energy Efficiency in Mobile Wireless Sensor Networks*

Mehwish Nasim and Saad Qaisar

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology, Islamabad, Pakistan
{mehwish.nasim,saad.qaisar}@seecs.edu.pk
http://connekt.seecs.edu.pk

Abstract. The inherent challenges which static sensor networks face such as energy constraints, are also faced by mobile sensors. In this work, we present a hierarchical cooperative clustering architecture for ad hoc wireless sensor networks. Communication cost is a crucial factor in depleting the energy of sensor nodes. We propose a framework in which nodes cooperate to form clusters at each level of hierarchy ensuring maximal coverage and minimal energy expenditure with relatively uniform distribution of load within the network. Performance is enhanced by cooperative multiple-input multiple- output (MIMO) communication. We test our framework using TOSSIM over TinyOS-2.0.x on MICAz motes. We implement and compare the proposed framework with cooperative clustering scheme (CMIMO) and traditional multihop Single-Input-Single-Output (SISO) routing approach. Performance is evaluated on the basis of energy consumption. Experimental results show significant energy conservation as compared to existing schemes.

1 Introduction and Motivation

Wireless sensor networks (WSNs) demand energy conserving techniques from MAC to application layer. One of the main design challenges in wireless sensor networks is coping with resource constraints placed on individual sensor devices. Energy constraints end up creating limitations such as computational power and limited coverage which are an impediment to achieve the overall objective of a sensor network [1].

In large networks, data is communicated to sink via multihop routing. Typically, in static networks, nodes closer to sink get depleted off energy, despite energy conservation techniques. This is because these nodes are more loaded than other nodes in the network i.e. the reach back problem [2]. They act as relay nodes between sink and the network. If sink is mobile, the number of hops to communicate data can be reduced. This is because a mobile sink can reach in the vicinity of nodes in the sensor network. Similarly, because redundancy

* This work was supported by NUST grant nustrd100109.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 159–172, 2011.

requirement (*'connectivity'*) in sensor nodes can be relaxed by introducing one or several mobile sinks, therefore, a sparse network architecture may be considered. In a dense network deployment, nodes may die out or the connectivity is disrupted due to physical obstacles or damages; in such scenarios mobile sinks can effectively collect data from the sensory nodes.

Mobile wireless sensor networks also face the inherent design challenges which static WSNs face. Since sensor nodes are energy-constrained, network longevity is a major concern in designing topology protocols, especially for applications deployed in harsh environments. Communication models such as hierarchical models [3] and [4], cooperative schemes [5], planer design [6] have been proposed in literature, for energy efficient clustering. Combined clustering and routing has also been considered for maximizing networks lifetime [7].

In this work, we focus on two scenarios: sensor nodes are static with a fixed basestation and a second scenario where there is a mobile sink in the network. The mobile sink is a light weight airborne plateform containing one or more sensor nodes. It serves as the data sink.Most sensor networks are static in nature but in emergency scenarios such as after an earthquake or during a fire (in buildings or forest fires) it is impossible to safely deploy a traditional static sensor network. In such events, it may be unsafe for the rescue teams to enter the buildings and enclosed spaces. Light weight airborne sensor networks can be used in such situations to detect survivors and send alerts to rescue teams. In recent years, cooperative Multi-input multi-output (MIMO) has been proposed as a communication model to be used in ad hoc wireless sensor networks [5], [8], [9]. MIMO technology has the potential to increase channel capacity and reduce transmission energy consumption. Since large network deployments necessitate energy efficient communication techniques, a concept known as cooperative MIMO has attracted a growing interest. In a cooperative MIMO network, a group of sensor nodes cooperate to transmit and receive data. The participation of multiple transmitters and receivers in a transmission saves significant energy in long-range communication [10]. Energy can be conserved if nodes are allowed to cooperatively transmit data just like in MIMO mode of communication. In our previous work [11], we showed that energy is conserved for a single transmission when data is routed via hierarchical cooperative routing. We extend that work by deploying our scheme on a more complicated network where clusterheads are aggregating data from a number of nodes and sending it to mobile/static sinks. Results show an extension in network lifetime in a communication intensive network. We exploit the benefits of cooperative communication for relaying data in static nodes as well as on airborne sensory platform.

1.1 Contributions

The main contributions of this paper are as following:

- A novel architectural framework for energy efficient routing to be used with cooperative MIMO communication in ad hoc WSNs.

– Performance analysis of our framework which minimizes energy consumption by employing an energy efficient cooperative clusterhead(CH) selection/routing algorithm.

1.2 Organization

The remainder of this paper is organized as follows. Section 2 gives an overview of existing literature. In section 3 we explain our framework; in section 4 and 5 we present our network model and experimental results. Section 6 summarizes our conclusions.

2 Related work

In the past few years, MIMO has surfaced as a reliable and energy conserving technology in the field of wireless networks. Various studies have focused on use of MIMO in sensor networks in order to improve energy conservation, network throughput and reliability in fading channels. In [12], authors propose a multiple-input multiple-output (MIMO) technique where multiple nodes within a cluster cooperate in signal transmission and reception. A cross-layer design is applied to optimize routing in order to minimize the energy usage and delay. For the cooperative MIMO scheme, routing is optimized based on an equivalent single-input single-out (SISO) system, where authors have treated each cooperating cluster as a super node.

Recently, lightweight airborne sensor networks have gained considerable attention as an area of research, Sensorflock[13] and Sensorfly[14] being two examples. An airborne wireless sensor network provides the capability to enhance many applications of interest to scientific community. This is provided by three-dimensional sampling of phenomena of interest that would otherwise be infeasible. One such class of applications is chemical dispersion sampling. A deployment of a flock of airborne sensors sensing and communicating their data back to a network of ground stations enables scientists to study the rate of dispersion of a natural or man-made toxin, pollutant, or chemical[13]. Another protocol, MobiRoute [15], supports routing in wireless sensor networks (WSNs) with a mobile sink. The authors theoretically prove that moving the sink can improve network lifetime without sacrificing data delivery latency. In MobiRoute, simulation results show that a mobile sink, in most cases, improves the network lifetime with only a modestly degraded reliability in packet delivery.

To our knowledge, little work exists in cooperative clustering for sensor networks with mobile sinks. An on ground MIMO based protocol TwinsNet [16],is proposed which shows gains, in terms of link loss rate. The improved link loss rate implies greater availability of critical data where the authors assumed that continuous data is required by the application. A uniform model for link quality is unrealistic for complex, unstructured environments but simulation results show a decrease in link loss rate. Authors assumed random movements of robots in the network. However, in TwinsNet protocol, energy consumption for the network is not explored.

3 Hierarchical Cooperative MIMO

In conventional networks where on ground sensory systems are deployed, as the system size gets smaller, the antennas get closer to the ground. This detunes the antennas, reduces line-of-sight, and increases multi-path problems [16].Airborne sensors offer better reception and transmission. However, when the sensors are deployed on light weight platforms, which are mobile, battery conservation becomes an important concern.

In this work, we propose Hierarchical MIMO (HMIMO), a novel clustering topology framework with MIMO capabilities for mobile sinks. To our knowledge, cooperative MIMO gains in lightweight airborne sensory devices has not been addressed in past.

MIMO offers three types of gains: Multiplexing, Capacity and Diversity gain. In this work we focus on diversity gain. Diversity gain is the slope of average bit error rate (BER) curve versus signal-to noise ratio (SNR). Diversity gain for a targeted BER offers energy conservation in static sensor networks. In networks consisting of mobile nodes, next hop destination changes from time to time. Each time the nodes transmit data, they do not recalculate a path which consumes least energy. This is because, such a task is very energy intensive. In clustered networks, the clusterheads communicate with mobile sinks. Everytime nodes transmit data to the mobile sink, the energy consumption differs based on the distance between the mobile sink and clusterhead. When mobile sinks are in close vicinity of clusterheads, they communicate via SISO approach. As the distance changes a distance dependent transmission protocol is followed in which CHs communicate with mobile sinks using cooperative MIMO approach.

3.1 Architecture

Our network is a heterogenous hierarchical network as illustrated in Figure 1. Network consists of ordinary sensor nodes, clusterheads, cooperative nodes, beacon nodes and airborne sensors (ABS).

When the algorithm starts, each ordinary node sends its neighbor node a HELLO message broadcast at a low power level. The aim of using low power is to send message to only one hop neighbors which are in close geographic vicinity. The second goal is to save energy. Every node calculates its weight on the basis of neighborhood information which is the normalized sum of clustering parameters. The first parameter is the transmission range of a sensory node. A node is able to communicate with any other node within a given transmission range. As we increase transmission range, there is also an increase in the radius of a cluster. For sparse networks we may have to increase the transmission range but for dense networks the transmission range is kept less so as not to burden the clusterheads. The second parameter is the residual energies of nodes. A node with a higher amount of energy is a better candidate for becoming a clusterhead. The third factor is link quality. HELLO message consists of node ID, its weight and a list of its neighbors. The nodes continue to exchange this information for a finite number of rounds.

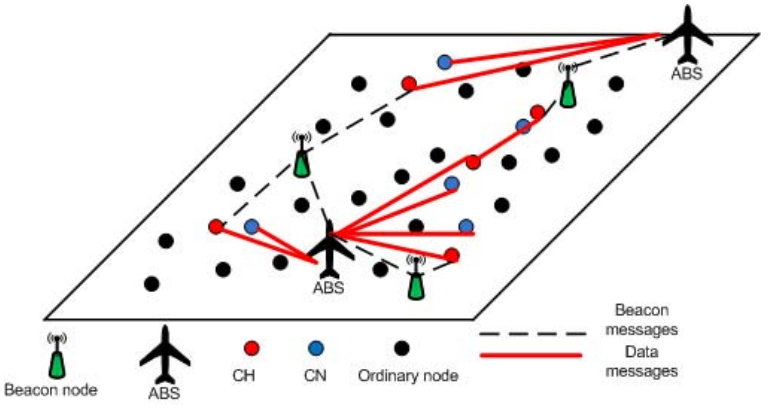


Fig. 1. HMIMO with mobile airborne sinks

Clusters formation. Clusters formation consists of following steps. It is discussed in a greater detail in our previous work [11].

Step 1: Each node compares its weight with that of its neighbors. If a node v has greatest amount of weight in it's neighborhood, it declares itself a clusterhead. If there is a tie, it is resolved on the basis of node ID.

Step 2: If v is not the best node then it sends a "Clusterhead message" to the node with the highest residual energy to become a clusterhead in case the best node has not already declared itself a clusterhead.

Step 3: After sending the "Clusterhead message" to the best node, node v waits for a fixed duration of time(δ) for the best node to reply back.

- If that best node and not any other node in the neighborhood sends a "Cluster formation" message then after a fixed time (δ), node v declares it self to be a clusterhead.
- Each clusterhead selects a cooperative node (CN).

Airborne sinks. We have assumed airborne sinks as sensory units consisting of two sensor nodes. The airborne sinks have the capacity to fly across the sensor network. In emergency situations where there is buildings debri or many obstacles, ABS cannot move in a fixed path. Therefore, in such scenarios they follow a random movement. We have also considered scenarios where ABS can move in a predefined path. After collecting data from the ABS send this data to the basestation which is assumed to be outside the network.

Beacon nodes. Few nodes in the network are assigned the duty of beacon nodes. Whenever an ABS comes in the vicinity of beacon nodes, this information is communicated to neighboring clusterheads.

3.2 Routing

For hierarchical routing, clusters join to form bigger clusters. Once clustering process at first level is completed, clusters may join to form bigger clusters for routing called Routing Clusters. The neighboring clusterheads exchange a ROUTE message. In this message, clusterheads announce the cluster ID of the cluster whose clusterhead possesses the highest residual energy. Such a cluster becomes a routing cluster. Routing clusterheads announce their presence to all neighbor CHs in the network. The clusterheads update their routing tables. If it takes less energy from a CH to send data to ABS then it send data directly to ABS, otherwise, it send data to its Routing CH [11]. In this work we consider only three levels of hierarchiy, i.e static sensor nodes, clusterheads and some routing clusters.

Reclustering. Topology maintenance is critical for a clustered network. At certain times an update is required in a cluster. Reclustering can happen in any of the following cases:

- A clustering update is required when a link is created between two nodes.
- When an existing link between two nodes is broken:
  - If one of the nodes is a clusterhead and the other is an ordinary sensory node then both the nodes trigger clustering again.
  - If both the nodes are ordinary nodes then both nodes announce their new weights to their clusterheads and all the neighbors. This would make the clusterheads update the neighbor list of their neighbors.
- When the residual energy of a clusterhead reaches a threshold value(an energy parameter), clusterhead delegates the responsibility to another node in the cluster who has high enough energy to become a clusterhead. If some of the neighbors of old clusterhead cannot communicate with the new clusterhead then either they retrigger clustering or join neighboring clusters.

3.3 Energy Consumption

Since the ABS are far from CHs, presumably the energy consumption follows the multi-path model. Therefore, energy per bit consumed in a cluster where there are N/k nodes in a cluster (N is the total number of nodes in the network and k is the number of clusters), is given by [17]:

$$E_{CH} = (\frac{N}{k} - 1)E_{circuit} + \frac{N}{k}E_{DA} + E_{circuit} + \epsilon_{mp}(d^*_{u,v})^4 \tag{1}$$

where $E_{circuit}$ is [5]:

$$M_t(P_{DAC} + P_{mix} + P_{filt}) + 2P_{syn} + M_r(P_{LNA} + P_{mix} + P_{IFA} + P_{filr} + P_{ADC}) \tag{2}$$

and $\epsilon_{mp}(d^*_{u,v})^4$ is given by:

$$(1+\alpha)(\gamma(M_t, M_r)N_o B N_f G_o M_l (d^*_{u,v})^4) \tag{3}$$

where $\gamma(M_t, M_r)$ is the required SNR at the receiver. In this case (where we have single input and single output at the transmiter and receiver ends) M_t and M_r are both one. P_* are power consumption values.E_{DA} is energy consumed for data aggregation ($d^*_{u,v}$, is the optimum distance between two communicating nodes u and v,. Rest of the notations are explained in table 1.

4 Network Model

Nodes have unique IDs and have some location coordinates namely x and y. Initially all the nodes have same amount of energy. Sensor nodes are deployed using a uniform random distribution. Two nodes are called neighbors if they are within the transmission range of each other. Mobile Airborne sinks vary from 1 to M. To test our framework, we have averaged our results on 100 different runs for varying distribution of nodes.

Table 1. System Parameters

Parameter	Meaning	Value
R_b	Bit Rate	1 Mbps
P_{DAC}	Digital-to-Analog converter	$15mW$
P_{ADC}	Analog-to-Digital converter	$15mW$
P_{mix}	Mixer	30.3 mW
P_{filt}	Active filters at transmitter	2.5 mW
P_{filr}	Active filters at receiver	2.5 mW
P_{syn}	Frequency synthesizer	50 mW
P_{LNA}	Low noise amplifier	20 mW
P_{IFA}	Intermediate frequency amplifier	2 mW
B	Bandwidth	10 KHz
N_0	PSD	-171 dBm/Hz
M_l, N_f	Link margin, Receiver noise figure	10 dB

For energy analysis we consider communication on Rayleigh fading channel. We analyze our scheme for a target BER of 10^{-3}. Signal to Noise Ratio (SNR) are the same as in [5]. All other values are assumed as mentioned in Table 1, otherwise stated. All nodes generate packets independently. We have used poisson distribution(λ) for packets generation. Packets are generated according to poisson distribution per 10 seconds. We assume basestation to be located outside the square field (sensor nodes deployment) unless stated otherwise. The BS is equipped with multiple antennas. ABS transmit data to the basestation. We implement two version of our framework, one for static network with a single basestation and the other version with mobile sinks. We compare our framework

with hop by hop MIMO called CMIMO [5], which is the most recent state-of-the-art work on cooperative communication in sensor networks, and traditional multihop SISO approach [4].

We choose these two schemes because of two reasons. Firstly, we show that our scheme is energy efficient as compared to SISO clustering approach. Secondly, we also show that hierarchical routing when used in conjunction with cooperative MIMO gives better results in terms of energy conservation as compared to existing multihop cooperative MIMO clustering/routing approaches in mobile airborne sensor networks.

5 Results and Discussion

5.1 Transmission Modes

Figure 2 shows a topology of a network with $k = 100$, (k is the number of level-one clusters). In this figure, at this particular instance the ABS/basestation lies at the center of square grid. The four different colored links show communication modes between clusters. Figure 2 shows various communication modes for routing in CMIMO.

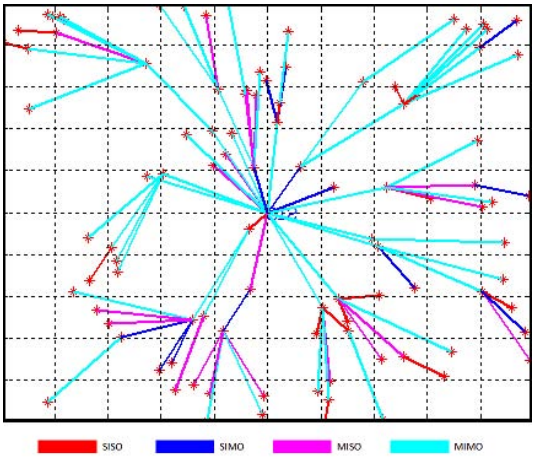


Fig. 2. Routing in CMIMO

In figure 3, packets are routed via HMIMO. In the previous approaches all the k number of clusters were involved but in our approach the number almost reduces to $log(k)$. Therefore, there are less number of hops in hierarchical cooperative communication as compared to the traditional CMIMO approach. The figure also shows that MIMO mode of communication is dominant in the network which is indicates less energy consumption.

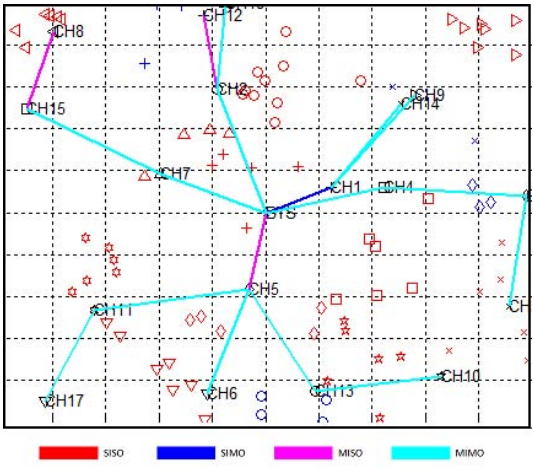


Fig. 3. Routing in HMIMO via backbone Clusterheads

5.2 Static Network with Hierarchical Cooperative Communication

In this experiment there are 100 nodes which are randomly deployed within a square region. Figure 4 shows a comparison between hierarchical MIMO with static nodes (HMIMO), CMIMO and traditional SISO. For this set of simulations we varied the size of the grid from 100×100 to 1000×1000 and Intercluster range is 33 % of the grid size. In HMIMO, data is forwarded via hierarchical routing. For the sake of simplicity, we have assumed a cost based link state routing for the other two algorithms. We alter the pathloss exponent between 2.7(semi-furnished rooms) and 3 (densely furnished buildings) [18]. For smaller grid sizes the performance of all the three algorithms is similar. This is because the intracluster distance between the clusterhead and sensory nodes and the intercluster distances between the communicating clusterheads is less. For shorter distances our algorithm and CMIMO uses SISO mode of communication. As we increase the grid size traditional multihop SISO continues to provide communication via SISO approach. However, the other two algorithms start switching to cooperative modes. In such scenarios, HMIMO outperforms CMIMO because number of hops for the data to reach basestation are less in HMIMO. The reason for decrease in number of hops is the hierarchical nature of our algorithm. As mentioned earlier, for small networks our scheme is using energy comparable to CMIMO. As we increase the network area, there is an upto 15-20% energy saving.

In state-of-the-art CMIMO (described in section 2), as the intra-cluster range increases, fewer clusters are formed. Thus, transmission distances between CHs become larger and circuit energy becomes less significant than transmission energy, which results in making other modes than SISO more favorable. A decrease in number of clusters increases number of nodes in each cluster which over

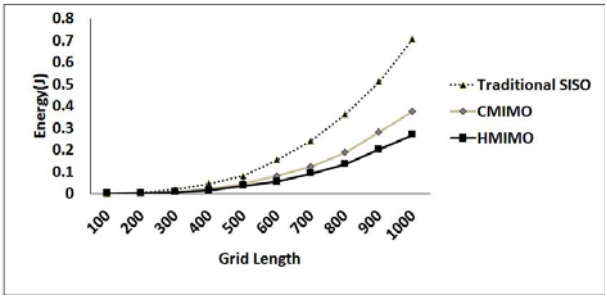


Fig. 4. Energy (J) consumption in HMIMO vs. Traditional approaches

burdens the clusterheads which are the communication centers. In our scheme, clustering parameters are not changed for the purpose of saving energy in routing. Distance dependent routing takes place by selecting appropriate nodes as routing clusterheads with the purpose of energy conservation. The advantage of our scheme is that routing and clustering jointly conserve power.

5.3 Energy Consumption with Mobile ABS

In this experiment, 100 nodes are randomly deployed within a square region. Figure 5 shows a comparison between HMIMO, hop by hop MIMO and traditional SISO. For this set of simulations we obtained results on a grid size of 1000×1000 meters. The Intercluster range for the first level clustering is taken as 33 % of the grid size. In HMIMO, data is forwarded via hierarchical routing. For the sake of simplicity, we have assumed a cost based link state routing for the other two algorithms.

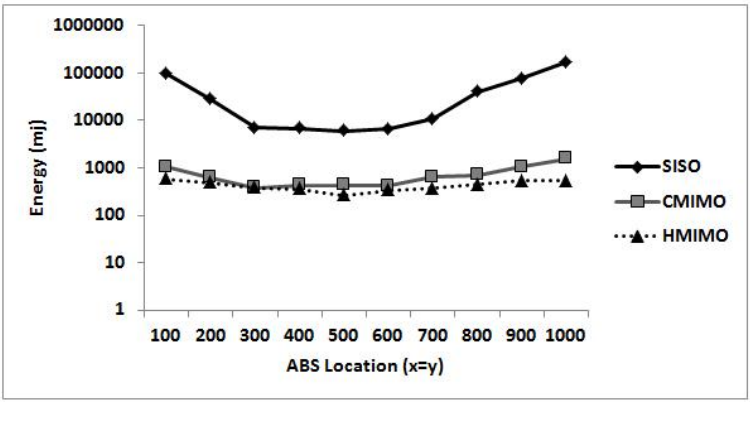


Fig. 5. Energy consumption with mobile ABS

ABS moves diagonally from one corner of the field to another. For shorter distances our algorithm and hop by hop MIMO uses SISO mode of communication. As the distance between ABS and majority of clusters decreases, i.e. when ABS is almost in the centre of the grid, HMIMO consumes less energy. The cooperative algorithms start switching to cooperative modes when distance between a clusterhead and ABS greater than the threshold value. In such scenarios, HMIMO outperforms hop by hop MIMO because number of hops for data to reach ABS are less in HMIMO. HMIMO utilized the distance dependant tradeoff. The reason for decrease in number of hops is the hierarchical nature of our algorithm. For small networks our scheme is using energy comparable to hop by hop MIMO. Aggregate energy consumption, when ABS moves from source to destination is the least in HMIMO. Results show approximately 10% energy conservation in HMIMO as compared to hop by hop MIMO. For same target BER Multihop SISO consume maximum energy because it requires maximum number of hops all using a single-input-single-output mode.

In another simulation we find energy expenditure by varying the number of airborne sensors 6. HMIMO outperforms tradtional approaches. When CHs are relatively equidistant from ABS, overall energy consumption in the network decreases.This is attributed to the fact that there is a decrease in number of hops and ABS are now in the ideal *distance-range* to cooperatively communicate with CHs. Even when there the number of ABS increases, HMIMO still outperforms tradtitional approaches because ABS send the data to base station via MIMO communication. This decreases the burden on static nodes deployed in the field.

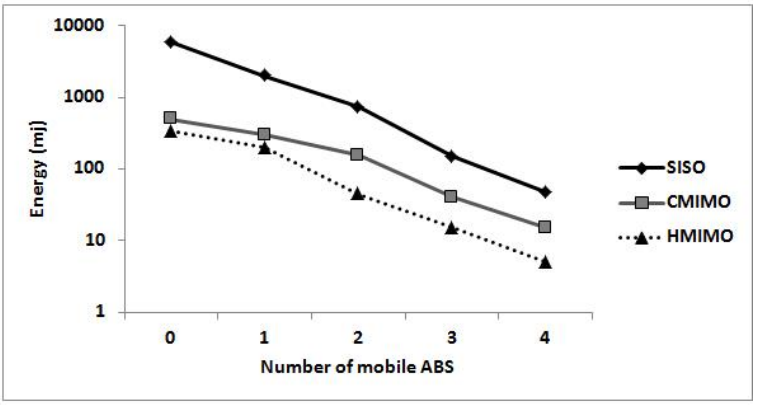


Fig. 6. Energy consumption with multiple ABS

5.4 Network Lifetime

In this experiment we measure network life time. We calculate network lifetime based on how much a node is used for communication. We analyze this performance metric on the basis of an energy parameter (an arbitrary threshold value).

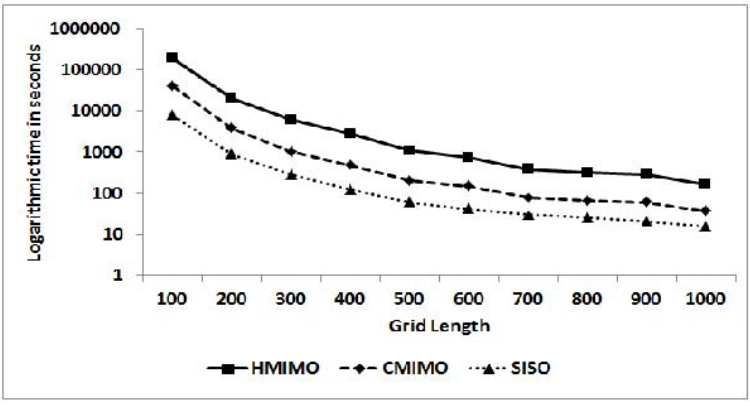


Fig. 7. Increase in network lifetime

We measure and compare the algorithms depending upon the time when 50% of nodes reach this threshold value. We introduce a mobile ABS in the network. Figure 7 shows a comparative analysis. Network employing Hierarchical cooperative clustering has a longer life time as compared to CMIMO and SISO. [1] An improvement in energy conservation at the communicating clusters elongated network lifetime manifolds. There is initially an increase in network lifetime with an increase in grid length because the intercluster distance increases which makes MIMO a more favorable communication mode. Nevertheless, when grid length is stretched beyond 500m, the total number of hops also increases, this results in a decrease in network lifetime for our algorithm as well as CMIMO.

6 Conclusion

In this work, we presented clustering/routing framework for mobile sinks (airborne sensors) in wireless sensor networks. Our framework uses cooperative MIMO techniques to communicate data to Airborne sensors. Each airborne sensor is equipped with two sensor nodes in order to carry out cooperative communication with clusterheads. Beacon nodes are deployed in the network for synchronization between clusterheads and Airborne sensors. We compared our results with a multihop MIMO and traditional multihop SISO techniques. For a target BER, there is energy conservation when data is routed via HMIMO. We tested our framework for fixed and random movements of ABS for a different number of ABS. Experimental results show approximately 15% energy gain as compared to CMIMO and more than 50% energy savings as compared to traditional SISO.

[1] When comparing the algorithms, distribution of nodes and packets generation is the same.

References

1. Hill, J., Culler, D.: A wireless embedded sensor architecture for system-level optimization (2001)
2. Barros, J., Servetto, S.: On the capacity of the reachback channel in wireless sensor networks. In: IEEE Workshop on Multimedia Signal Processing, pp. 408–411. IEEE, New York (2003)
3. Iwanicki, K., van Steen, M.: On hierarchical routing in wireless sensor networks. In: Proceedings of the Eighth ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN 2009), IP Track, San Francisco, CA, USA, April 2009, pp. 133–144 (2009), http://www.few.vu.nl/~iwanicki/publications/2009-04-IPSN
4. Basagni, S.: Distributed clustering for ad hoc networks, ispan, p. 310. IEEE Computer Society Press, Los Alamitos (1999)
5. Siam, M.Z., Krunz, M., Younis, O.: Energy-efficient clustering routing for cooperative mimo operation in sensor networks. In: INFOCOM 2009, April 2009, pp. 621–629 (2009)
6. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, p. 10 (2000)
7. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. Comput. Commun. 30(14-15), 2826–2841 (2007)
8. Cui, S., Goldsmith, A.J., Bahai, A.: Energy-efficiency of mimo and cooperative mimo techniques in sensor networks. IEEE Journal on selected areas in communication (2004)
9. Bravos, G., Kanatas, A.G.: Energy efficiency comparison of mimo-based and multihop sensor networks. EURASIP Journal on Wireless Communication and Networking, 2008, 1–13 (2008)
10. Ozgur, A., Leveque, O., Tse, D.N.C.: Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks. IEEE Transactions on Information Theory (2007)
11. Nasim, M., Qaisar, S.: Hierarchical mimo: A clustering approach for ad hoc wireless sensor networks. In: Proceedings of 44th Conference on Information Sciences and Systems, Princeton, USA (March 2010)
12. Cui, S., Goldsmith, A.J.: Cross-layer design of energy-constrained networks using cooperative mimo techniques. Signal Process 86(8), 1804–1814 (2006)
13. Allred, J., Hasan, A., Panichsakul, S., Pisano, W., Gray, P., Huang, J., Han, R., Lawrence, D., Mohseni, K.: SensorFlock: an airborne wireless sensor network of micro-air vehicles. In: Proceedings of the 5th international conference on Embedded networked sensor systems, pp. 117–129. ACM Press, New York (2007)
14. Purohit, A., Zhang, P.: SensorFly: a controlled-mobile aerial sensor network. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, pp. 327–328. ACM Press, New York (2009)

15. Luo, J., Panchard, J., Piórkowski, M., Grossglauser, M., Hubaux, J.: Mobiroute: Routing towards a mobile sink for improving lifetime in sensor networks. Distributed Computing in Sensor Systems, 480–497 (2006)
16. Zhang, Q., Cho, W., Sobelman, G., Yang, L., Voyles, R.: TwinsNet: A cooperative mimo mobile sensor network. Ubiquitous Intelligence and Computing, 508–516 (2006)
17. Kim, H., Kim, S.W., Lee, S., Son, B.: Estimation of the optimal number of cluster-heads in sensor network, pp. 87–94 (2005), http://dx.doi.org/10.1007/11553939_13
18. Rappaport, T.: Wireless communications. Prentice Hall PTR, New Jersey (2002)

Towards Fast and Energy-Efficient Dissemination via Opportunistic Broadcasting in Wireless Sensor Networks

Minjoon Ahn[1], Hao Wang[1], Mihui Kim[2], and Hyunseung Choo[1,⋆]

[1] School of Information and Communication Engineering
Sungkyunkwan University, Suwon, South Korea
ahn.m.j@skku.edu, wangh513@gmail.com, choo@ece.skku.ac.kr
[2] Department of Computer Engineering
Hankyong National University, Anseong, South Korea
mhkim@hknu.ac.kr

Abstract. Broadcasting in wireless sensor networks (WSNs) is a basic operation to control the entire network. However, traditional broadcasting is not appropriate for WSNs with a low-duty-cycle; it is devised for an energy-limited environment, where nodes stay asleep much of the time and wake up only for a short time. Duty-cycled broadcasting methods, such as the opportunistic flooding (OF) scheme, have been studied to reduce the flooding delay. However, OF suffers a problem of energy unbalanced consumption, incurring early network disconnection. In this paper, we modify OF to decrease the broadcast delay and prolong the network lifetime, through acquiring more candidates for senders and considering the remaining energy of nodes. Simulation shows our scheme achieves shorter delay and longer network lifetime than OF (i.e., decrease by up to about 60% and increase by up to about 100%, respectively).

Keywords: Opportunistic, Broadcasting, Energy-efficient, Lifetime, Remaining energy, Fast dissemination.

1 Introduction

Broadcasting facilitates sensor nodes to propagate messages across the entire network, as a fundamental service in wireless sensor networks (WSNs), whilst serving a wide range of high-level operations. This is critical to effectively implement a network-wide broadcast service for the overall optimized performance of WSNs [1]. Therefore, much effort has been made to improve broadcasting efficiency. Several researchers focus on reducing broadcast redundancy [2,3] or increasing broadcasting reliability [4]. Trickle [5], as a code propagation algorithm, maintains up-to-date information with a low maintenance overhead. DIP [6] extends Trickle to reduce the number of transmissions for a dense network.

However, such research is unsuitable under a duty-cycled environment. The sensor network should have a sufficiently long lifetime to fulfill the application requirements,

⋆ Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 173–184, 2011.

but most sensor nodes use a small battery. A sensor network has operate under a duty-cycle, in which a sensor node schedules itself to be active for only a very brief period of time and then stays dormant for a long time, to bridge the gap between limited energy supplies and application lifetimes. A sender may have to wait for a certain period of time (termed *sleep latency* [7]) until its receivers become active to deliver a packet.

Broadcasting methods under duty-cycles have been recently studied. Wang et al. [1] transformed this problem into a shortest-path problem, and addressed it as a solution via dynamic programming. However, the solution operates only with a centralized method. An opportunistic flooding (OF) scheme for low-duty-cycle networks with unreliable wireless links is proposed [8]. The key idea of OF is to make probabilistic forwarding decisions at a sender based on the delay distribution of next-hop nodes. From a probabilistic viewpoint, the transmission on an energy optimal tree (i.e., composed of links with best quality) could achieve the fastest dissemination. However, senders in OF get a chance to forward a packet using links outside the energy optimal tree when they receive a packet opportunistically early, to reduce flooding delay. Nevertheless, OF has a drawback (i.e., the energy consumption is concentrated on certain nodes on the energy optimal tree) due to the link-quality-based backoff during the sole sender selection phase (discussed in Section 2). Therefore, none of this research has provided good solutions for both small flooding delay and a long network lifetime.

In this paper, we propose an energy-efficient opportunistic broadcasting (EEOB) in a duty-cycled environment. EEOB uses links to the siblings to ensure more sender candidates, and thus increases the probability of energy balancing and fast dissemination. Moreover, EEOB applies a loop prevention scheme to prevent this side effect, by considering the sibling links. EEOB considers not only the link quality but also remaining energy of the nodes, when multiple senders compete for flooding. The nodes in the sender-candidate-set that have greater remaining energy and better link quality than the receiver get a higher chance to be the sole sender. We modify the OF scheme [8] to fit our scheme to provide reliable transmission (i.e., avoid the hidden terminal problem and reduce the collisions).Our contributions follow:

- We develop EEOB to decrease broadcasting delay and prolong network lifetime, through maintaining more candidates for senders and considering the remaining energy of nodes.
- Simulations show the broadcasting delay of EEOB decreases up to 60% and the network lifetime increases up to 100% compared to OF.

The remainder of the paper is organized as follows. Section 2 gives background information and related work. Section 3 details EEOB operations. Section 4 demonstrates the performances of the network lifetime and broadcasting delay through simulation. Section 5 concludes this paper.

2 Related Work

Much research into broadcasting has progressed in WSNs. The traditional flooding method and many improved schemes [3,4] have proven that they had good performance

on delivery ratio, delay and energy cost in an always-awake network. However, the performance of such flooding schemes under a duty-cycled environment will be seriously declined [8]. Moreover, if unreliable links and collisions in the wireless are considered, their delivery ratios worsen; simulations show a network with 2% duty-cycle delivers less than 5% of packets [8].

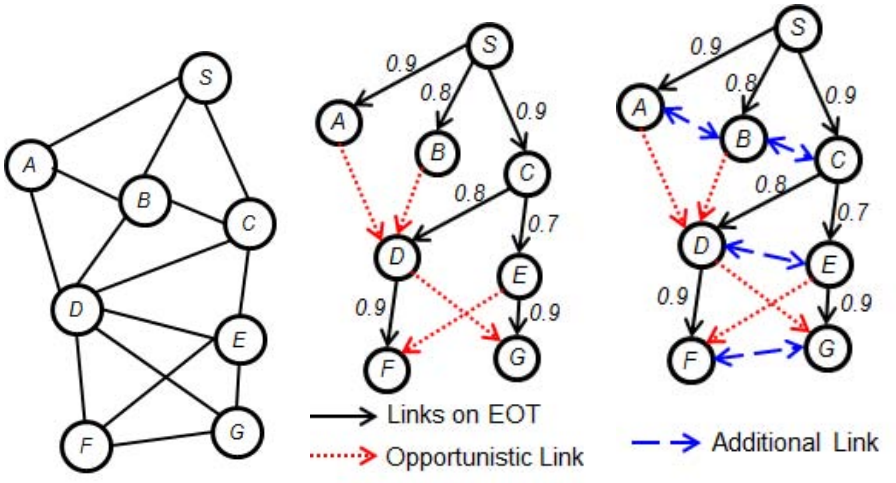
Guo et al. [8] proposed an opportunistic flooding (OF) that improves the reliability of traditional flooding for duty-cycled environments. The key point is the forwarding decision making, in which nodes forward a packet with a higher probability, if the packet arrives opportunistically earlier. This is achieved by comparing the delay of individual packets to the statistic packet delay distribution, i.e., *probability mass function (pmf)* at next-hop nodes. OF includes three steps:

- **Computing the *pmf*:** Each node in the networks computes the probability of receiving from its parents on the energy optimal tree (EOT) over time t, makes the *pmf* table using this information and shares it with its neighbors.
- **Composing the sender-candidate-set:** When the intermediate node receives a packet from its parents, it checks the *pmf* of receivers and checks if the expected receiving time is earlier than a threshold of *pmf*. If yes, the node is added to the sender-candidate-set.
- **Computing the *pmf*:** When there are multiple sender candidates, they compete to be the sole sender, and thus collision is avoided.

However, energy consumption in OF may be concentrated on certain links, as shown in Fig. 1. Note that node E has sole parent (node C). Node C has two children, both are on the EOT. When node C competes with A or B to be the sole sender of D, it backoffs based on the link quality to D, and thus C always gets the highest probability of transmitting the packet. Finally, node C may "die earlier" than other nodes. If node C is exhausted, node E will be isolated. Node E can no longer receive packets, even if it has sufficient energy to receive. Note that neighbor node D could deliver to node E, but there are no logical links. We can find out that the opportunistic links are only connected between the parents and the children in Fig. 1(b). In [8], the authors assume a link only between the parents and the children for loop prevention. This constraint decreases the chance of fast dissemination and energy balancing, since the size of the sender-candidate-set is limited. Thus, our scheme maintains more opportunistic links to the siblings, as shown in Fig. 1(c). These additional links also contribute to reduce the isolation problem. Our scheme uses additional information, such as a node ID in the packet header, for loop prevention.

3 Energy-Efficient Opportunistic Broadcasting

In EEOB, we consider not only the link quality but also remaining energy of nodes. The consideration for link quality is related to both fast broadcasting and the reduction of the total energy consumption in the network, since links with better quality can reduce retransmission. However, the frequent use of such links makes nodes of the links suffer extreme load, and finally die earlier. Consideration of remaining energy can prolong network lifetime by energy balancing. We also consider additional opportunistic links



(a) Network Topology (b) EOT (c) Modified EOT

Fig. 1. An Example of Opportunistic Flooding

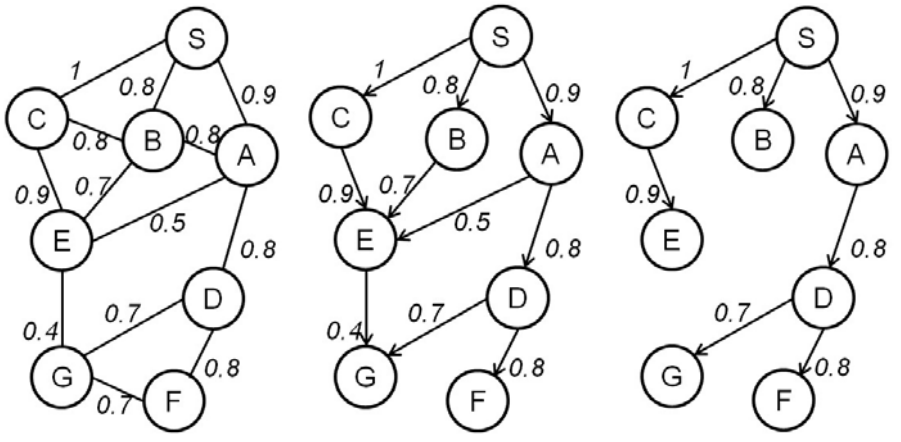
from each node to its siblings with the loop prevention scheme. Such additional opportunistic links may raise the possibility of both energy balancing and fast dissemination. In addition, we modify the OF to avoid the hidden terminal problem and reduce the collision for reliable transmission.

The assumptions in this paper are as follows. Each node sets up its working schedule and shares it with all its neighbors, as soon as it joins the network. This process is usually called low-duty-cycle rendezvous [9]. Each node knows its neighbors' working schedules after rendezvous. A node changes its working schedule, if its neighbors update schedules. We assume that unreliable links and collisions may occur in the network. In one communication range, if two or more ongoing transmissions occur simultaneously, none of them succeed. The measurement for link quality can be updated at a very low cost or by conventional low-cost piggybacking of data traffic. The network is locally synchronized and it can be achieved using the MAC-layer time stamping technique, as described by the flooding time synchronization protocol (FTSP) [10]. A node knows when it can send packets to the neighbors, given their working schedules.

3.1 Opportunistic Transmission

A node outside energy optimal tree (EOT) decides transmission to its neighbors by judging if the transmission is opportunistically early. This is the same as OF.

1) Constructing EOT and Computing pmf: Fig. 2 shows how to construct EOT. When an original network topology is constructed, as shown in Fig. 2(a), EEOB first reduces the sibling links and then specifies the flooding direction making a directed acyclic graph (DAG), as shown in Fig. 2(b). EEOB then selects a link that has the highest link quality from a parent to a child, as shown in Fig. 2(c).



(a) Network Topology (b) DAG for Flooding (c) Energy Optimal Structure

Fig. 2. Constructing Energy Optimal Tree

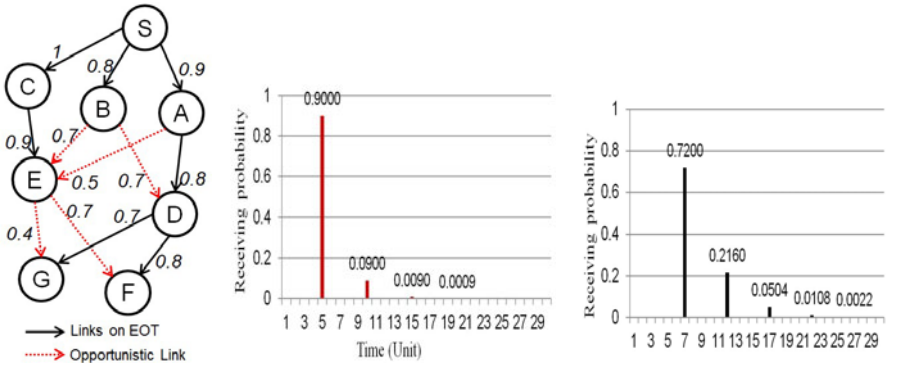
Computing the packet delay distribution via an EOT and comparing the delay along the EOT, a node can decide whether opportunistic forwarding via links outside of the EOT needs or not. This packet delay distribution is represented to *probability mass function (pmf)* and the computation process of *pmf* starts from a source node (level-0) and spreads throughout the network level by level. Initially, the source node always awakes and the probability that it receives the packet with delay 0 is 100%. In other words, the *pmf* of the source is (0,100%). Then, a level-1 node calculates its *pmf* based on its level-0 parent node's *pmf*. Similarly, a level-(*l*+1) node calculates its *pmf* based on its level-*l* parent's *pmf*. Given the *pmf* of level-*l* nodes (i.e., active time units t_l (*i*) and the probability of reception p_l (*i*) for any *i*) and t_{l+1} (*j*) for any j, each level-(*l*+1) node calculates the probability that it receives the flooding packet at its j^{th} active time unit as follows:

$$p_{l+1}(j) = \sum_{i:t_i(i)<t_{l+1}(j)} p_l(i)q(1-q)^{n_{ij}} \tag{1}$$

where q is the corresponding link quality satisfying $q \in (0, 1]$, n_{ij} is the number of the level-(*l*+1) nodes' active time units between $t_l(i)$ and $t_{l+1}(j)$. The term $p_l(i)q(1-q)^{n_{ij}}$ is the probability that the packet that arrives at the level-*j* node at its i_{th} active time unit is first delivered to the level-(*i*+1) node at its j_{th} time unit. Clearly, the *pmf* of a node can be derived from its parent's *pmf* with initial *pmf* (0,100%) at the source.

Fig. 3 shows an example of the *pmf* computation process. Assume that the nodes wake up periodically every 5 time units (20% duty-cycle) and node A and D first wake up at time 5 and at time 2, respectively. Node A computes its *pmf* first based on the link quality 0.9 and its own work schedule. The probability that node A receives the packet for the first time, at time 5, is 0.9. At time 10, the probability becomes $(1 - 0.9) \times 0.9 = 0.09$ and so on. Node D then computes its *pmf* based on the *pmf*

of A. For node D at time 7, the probability is the multiplication of the link quality and the probability that node A receives the packet at time 5, $0.9 \times 0.8 = 0.72$. For node D, at time 12, the probability is the sum of the probability that (i) node A receives the packet at time 5 and succeeds at the second transmission, and (ii) the probability that node A receives the packet at time 10 and succeeds in the first transmission, $0.9 \times (1 - 0.8) \times 0.8 + 0.09 \times 0.8 = 0.216$. Similarly, all the nodes within the network compute their *pmf*, as long as their parents' *pmf* becomes available.



(a) Network Topology (b) *pmf* of Node A (c) *pmf* of Node D

Fig. 3. Computing *pmf*

From the computed delay distribution, a node finds its p-quantile delay (denoted as D_p), as a threshold for delay, and shares this threshold with its previous-hop nodes. D_p is a threshold, such that if a flooding packet arrives at this node later than D_p, the probability that the node has already received this packet from its parent is greater than p. Then, for each new flooding packet and each next-hop node, a node computes the expected packet delay (EPD) and makes a forwarding decision based on the comparison between EPD and D_p. If EPD $\leq D_p$, the probability that the next-hop node has already received this flooding packet via the EOT is no greater than p. Thus, this packet is considered forwarded. If EPD$>D_p$, the next-hop node has more than p percentile of chance that it has already received this packet. Thus, this packet is considered redundant and will not be forwarded to the next-hop.

For example, in Fig. 3 we assume that node B wants to make an opportunistic link to node D. The expected number of transmissions for successful packet delivery is ceiling of 1/q, where q is the link quality between two nodes. Thus, the expected number from node B to node D is 2, since q is 0.7. Assume that p is 0.8, D_p is 12, because at time 12, node D will receive a packet successfully via an EOT with probability 0.936 (=0.72+0.216) more than p (=0.8). When node B receives a packet from S, if it can try to send at least twice before D_p (=12), then it decides to send a packet to D, else it gives up.

2) Using Sibling-links as Additional Opportunistic Links: We introduced the link limitation of OF in Section 2. We consider additional links to the siblings to prevent an isolation problem and maintain the larger sender-candidate-set. We can achieve both

load balancing and fast dissemination through the additional links. When a node receives a packet, it checks the possibility of making the opportunistic links to its children and siblings using the *pmf* of receivers. If EPD is less than D_p, this opportunistic link is included in the sender-candidate-set. We propose the loop prevention scheme by using sequence number (SEN) to prevent a broadcasting-loop in which each node floods a duplicated packet repeatedly. SEN indicates the packet sequence whose packet is newer. When a node receives a packet, it first checks the SEN in the packet header. If the packet has an older SEN than that received before, then the node discards it. Otherwise the node updates current SEN and it floods this packet. Using this scheme, each node does not send a duplicated packet that is has already received.

3) Considering the Residual Energy of Nodes: We know the energy balancing is still a major issue in WSNs. Thus, the remaining energy should be considered as a factor when choosing the sender. Our EEOB, via the factor, can improve energy balancing and enable the nodes to work longer. In EEOB, the node that has more energy and a better link quality has a higher probability to be the sole sender. When multiple senders are in the competing phase to send a packet to *j*, the weight of node *i* is computed using the following equation:

$$p_i = e_i^r \times q_{ij} \tag{2}$$

where e_i^r denotes the percentage of remaining energy of node *i*, and q_{ij} denotes link quality between *i* and *j*, p_i is the weight that node *i* is selected as the sole sender of *j*. The bigger P_i means node *i* has the high probability to be selected as the sender. In the early stage of the network, the nodes with higher link quality are selected as a sender more frequently, and thus their remaining energy decreases. As times go on, nodes with higher link quality are assigned a smaller weight to be senders, and thus energy consumption is balanced. Thus, energy balancing and fast dissemination can be achieved. In [11], Zhao et al. designed a residual energy scan that approximately depicts the remaining energy distribution within a sensor network. Their approach has good scalability to continuously extract the residual energy level individually from each node. We can use this approach to know the remaining energy of nodes in the network. After combining it with link quality, EEOB selects the optimal link to broadcast the packets.

3.2 Reliable Transmission

1) Preventing Hidden Terminal Problem: In wireless communication, a certain percentage of collisions are caused by the hidden terminal problem (HTP), where two nodes forward a packet to the same node without knowing each other. If this occurs, both keep sending but neither of them succeeds. OF scheme alleviated the HTP, using a link quality threshold l_{th} [8]. We apply the scheme with some modifications to fit our method. At Eq.2, we showed an equation for the sole sender selection. When there are multiple opportunistic links to a receiver, the links are added in the sender-candidate-set. The nodes in the sender-candidate-set must have higher link quality than l_{th} with all other nodes in the set. If there are some mutual links with bad quality in the sender set, some nodes may not overhear others' transmission, even if the packet has already arrived at the receiver.

2) Reducing the Collision: We should resolve the collisions problem after the sender set is constructed. Ideally, a node with the highest value P (in Eq.2) has the highest priority to use the channel and start a transmission with no collision. Selecting the best link always means the least number of transmissions is expected, so that both the expected next-hop delay and energy cost are minimized. When nodes intend to start a transmission, they first do backoff for random times. The duration of the backoff depends on the *P* value of the node. A node that has a higher P gets a higher probability to get a shorter backoff duration. When multiple nodes within communication range make their decisions to send towards the same node, they backoff first before transmission, and the one with the shortest backoff time starts first.

The probability that more than two nodes get the same backoff duration is very low. Nevertheless, if more than two nodes get the same backoff duration, they backoff again with a longer duration than before. After starting backoff, the nodes listen to the channel, and they can catch the ongoing transmission. The nodes that overhear other transmissions abort their transmissions. Using this random backoff method, EEOB can reduce collisions and decrease the chance that a packet is forwarded via a very weak link, since the winner must have a relatively good link quality and more energy to start early.

4 Performance Evaluation

We proposed two main ideas to achieve fast dissemination and prolong network lifetime. First, use additional links to the siblings for broadcasting. Second, consider the remaining energy of nodes when there are multiple senders. In this section, we compare three schemes, OF, OF+AL (i.e., OF using additional links), and EEOB (i.e., using additional links and considering the remaining energy of each node) in the aspects and *network lifetime* and *broadcasting delay*.

4.1 Simulation Setup

We implement a simulator with C# based on real topology with 54 sensors, as shown in Fig. 4 [12], to evaluate our design. We select only the links with quality higher than 0.3, based on measured link information and quality on the topology, since the links with quality below 0.1 or 0.3 may not be considered as valid links in a real environment [13]. We follow the energy consumption model of CC2420 for a more sophisticated simulation. ChipCon CC2420 radio [14] draws 19.7 mA when receiving or idle listening, larger than the 17.4 mA used in transmitting. We convert the unit from mA to energy unit, and assume that each node has 100,000 energy units initially to apply the energy model to the simulation.

We measure the network lifetime, until the first node in the network dies. When a source node broadcasts a packet, all nodes alive have to receive it. However, when the first node is dead, the network could be partitioned, and thus some nodes may not receive the packet, even if they are alive. Thus, we measure the network lifetime until the first node dies. We measure the lifetime and delay with time units. A time unit is

a period to complete 1-hop transmission. We assume that node #1 is a unique source node, and broadcasting delay is measured until all nodes receive a packet successfully. Nodes periodically wake up with duty-cycle, and the initial wake up time is randomly set. For example, when a network has 10% duty-cycle, the nodes in the network first wake up from 0 to 9 time units randomly, and they wake up every 10 time units after their first wake up time.

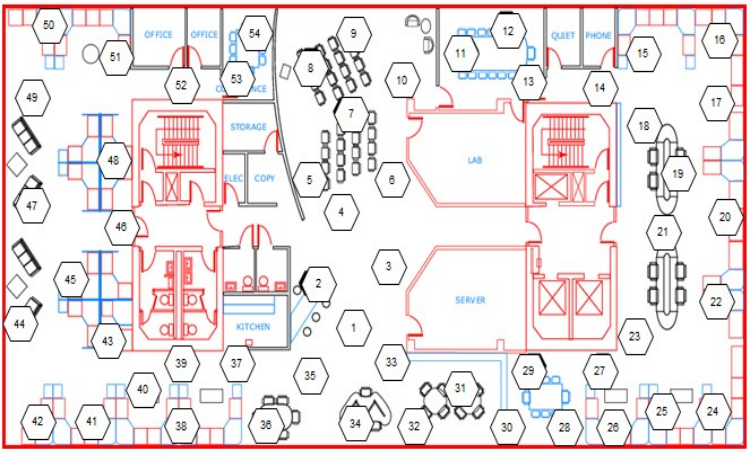


Fig. 4. Real-topology in the Intel Berkeley Research Lab

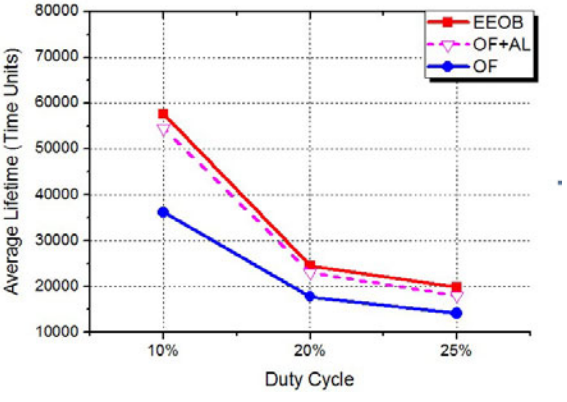
4.2 Simulation Results

In OF, energy consumption is concentrated in some nodes on the EOT due to lack of sender candidates and link-quality based backoff. OF+AL considers additional sibling links to acquire more sender candidates. Thus, it can have a greater chance to achieve faster dissemination and prolong network lifetime. EEOB considers both using sibling links and residual energy of nodes for more sophisticated energy balancing.

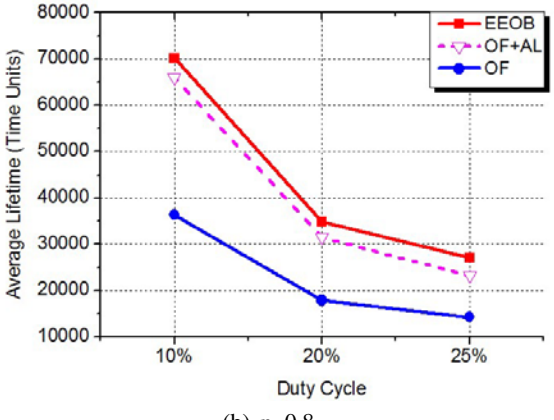
The results show the average lifetime and broadcasting delay when a quantile probability p is 0.6 and 0.8. We study the impact of p, the threshold to decide if a packet is opportunistically early. As p increases, more opportunistic links are considered, and thus energy balancing can be more sophisticated. In this simulation, a link quality threshold l_{th} is fixed to 0.5, while the duty-cycle is set with 10%, 20%, and 25%, respectively. The higher duty-cycle can achieve faster dissemination, but it reduces network lifetime.

Fig. 5 shows the average network lifetime when p is 0.6 and 0.8, and l_{th} is fixed to 0.5. The results show that the network lifetime is increased when a duty-cycle is lower and p is higher, as per our expectation. EEOB considers the remaining energy of nodes, and thus the network is maintained for longer.

Fig. 6 shows the average broadcasting delay with the same parameter values as in Fig. 5. OF+AL achieve a slightly better performance from the viewpoint of delay than EEOB, because it only considers the link quality in the competing phase of multiple senders, and thus it could minimize retransmissions. Despite increasing p, the results of delay are similar. Commonsensically, as p increases, the waiting delay to obtain more candidates increases. However, as p increases, more links can be considered, and thus a better link could be chosen. Therefore, the delay is not increased against expectations. In conclusion, EEOB outperforms other schemes in synthesizing the results of two performance metrics.

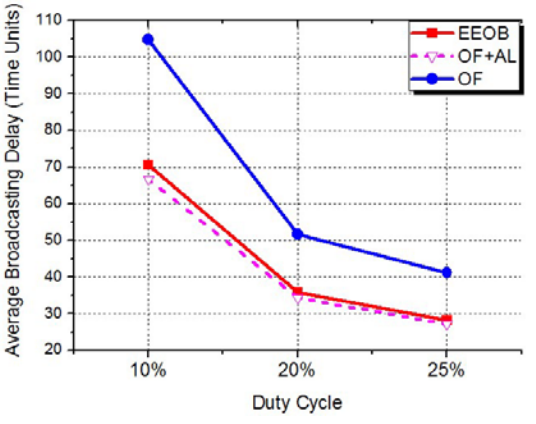


(a) p=0.6

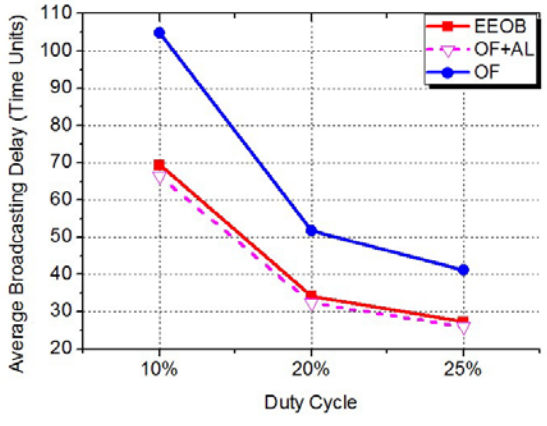


(b) p=0.8

Fig. 5. Average Network Lifetime



(a) p=0.6



(b) p=0.8

Fig. 6. Average Broadcasting Delay

5 Conclusion

One of the important issues in WSN is to extend the lifetime of the sensor nodes. Thus, the design of an energy-efficient flooding scheme with balanced consumption is essential, because flooding schemes have significant impact on the overall energy consumption of sensor networks. The existing opportunistic flooding scheme gives a chance to become the opportunistic link to only the link towards the children nodes, but energy consumption tends to be concentrated in certain nodes on the EOT. In our design, each node makes probabilistic forwarding decisions based on the delay distribution of next-hop nodes and its own remaining energy. The node that has higher energy and better link quality has a greater probability to be the sender. We built a qualified sender set to alleviate the hidden terminal problem, and in the same sender set, we use the random backoff method based on a link quality and remaining energy to resolve simultaneous

forwarding operations. We used the sibling links for broadcasting and considered the remaining energy of the nodes at the candidate selection for forwarding. We also prevented the loop occurrence by using node ID. Simulations showed our scheme significantly contributed to prolong network lifetime and reduce broadcasting delay.

Acknowledgment

This research was supported in part by MKE and MEST, Korean government, under ITRC NIPA-2011-(C1090-1121-0008), WCU NRF (No. R31-2010-000-10062-0) and PRCP(2010-0020210) through NRF, respectively.

References

1. Wang, F., Liu, J.: Duty-cycle-aware Broadcast in Wireless Sensor Networks. In: IEEE INFOCOM (2009)
2. Lou, W., Wu, J.: On Reducing Broadcast Redundancy in Mobile Ad Hoc Networks. IEEE Transactions on Mobile Computing 1(2), 111–123 (2002)
3. Kyasanur, P., Choudhury, R.R., Gupta, I.: Smart Gossip: An Adaptive Gossip-based Broadcasting Service for Sensor Networks. In: IEEE MASS (2006)
4. Stann, F., Heidemann, J., Shroff, R., Murtaza, M.Z.: RBP: Robust Broadcast Propagation in Wireless Networks. In: ACM SenSys. (2006)
5. Levis, P., Patel, N., Culler, D., Shenker, S.: Trickle: A self-regulating algorithm for code propagation and maintenance in wireless sensor networks. In: ACM NSDI (2004)
6. Lin, K., Philip, L.: Data Discovery and Dissemination with DIP. In: ACM IPSN, pp. 433–444 (2008)
7. Gu, Y., He, T.: Data Forwarding in Extremely Low Duty-Cycle Sensor Networks with Unreliable Communication Links. In: ACM SenSys. (2007)
8. Guo, S., Gu, Y., Jiang, B., He, T.: Opportunistic Flooding in Low-Duty-Cycle Wireless Sensor Networks with Unreliable Links. In: ACM MobiCom. (2009)
9. Dutta, P., Culler, D.: Practical Asynchronous Neighbor Discovery and Rendezvous for Mobile Sensing Applications. In: ACM SenSys. (2008)
10. Maroti, M., Kusy, B., Ledeczi, A.: The Flooding Time Synchronization Protocol. In: ACM SenSys. (2004)
11. Zhao, Y.J., Govindan, R., Estrin, D.: Residual Energy Scan for Monitoring Sensor Networks. In: IEEE WCNC (2002)
12. Real-topology in the Intel Berkeley Research Lab, http://db.lcs.mit.edu/labdata/labdata.html
13. Zamalloa, M., Seada, K., Krishnamachari, B., Helmy, A.: Efficient Geographic Routing over Lossy Links in Wireless Sensor Networks. In: ACM TOSN, vol. 4(3) (2008)
14. Texas Intruments. 2.4 GHz IEEE 802.15.4 / ZigBee-Ready RF Transceiver (Rev.B), http://focus.ti.com/docs/prod/folders/print/cc2420.html

A Dynamic Multiagent-Based Local Update Strategy for Mobile Sinks in Wireless Sensor Networks

Jinkeun Yu[1], Euihoon Jeong[1], Gwangil Jeon[1], Dae-Young Seo[1], and Kwangjin Park[2,*]

[1] Department of Computer Engineering
Korea Polytechnic University, 429-703, Korea
{yjk760,ehjeong,gijeon,seody}@kpu.ac.kr
[2] Division of Electrical Electronic and Information Engineering
Wonkwang University, Iksan, 570-749, Korea
kjpark@wku.ac.kr

Abstract. Recently, several geographic routing strategies considering a mobile sink node have been proposed for wireless sensor networks. A mobile sink node should frequently update its location information in source nodes for successfully receiving data from the latter. However, frequent location updates from the mobile sink node may result in the consumption of too many network resources. In this paper, we propose an efficient multiagent-based location update strategy for a mobile sink node in wireless sensor networks. Agent nodes that are located on the path between the source node and the mobile sink node contain the location information of the mobile sink node. In addition, the agent nodes are changed dynamically to maintain the approximate shortest path between the source node and the mobile sink nodes. We analyze the performance of the proposed scheme by performing simulation using Qualnet 5.0.

Keywords: Wireless Sensor Networks, Mobile Sink, Location Management, Multiagent.

1 Introduction

Wireless sensor networks have been researched extensively and applied in a number of fields such as battlefield monitoring, residence monitoring, traffic congestion monitoring, and security. Sensor nodes are small and simple and use limited memories and batteries; hence, it is necessary to design a protocol for minimizing unnecessary transmission in wireless sensor networks [1]. A number of researches have been performed on networks with a fixed sensor node and a sink node, and recently, studies have been carried out on networks with mobile sink nodes as well. A mobile sink node is constantly in motion; therefore, changing the global network topology is inevitable. Changing the global network topology results in excessive energy consumption in large-scale sensor networks, and hence, geographic routing is frequently used [2-3].

* Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 185–196, 2011.

Geographic routing uses the location information of each node to transmit a data packet, and the location information of neighbor nodes can be confirmed through a beacon message. When transmitting a data packet, the source node includes the location information of the sink node in the packet. In addition, each intermediate node uses the location information of the sink node of a data packet to forward the packet to the next closest node. Since a mobile sink node is constantly in motion, its location information must be repeatedly updated in the source node [2-4]. Further, since a number of overheads are created during the path when updating the location information of a sink node, energy consumption increases. Therefore, it is necessary to come up with a method for effectively updating the location information of a mobile sink node. Some of these methods are grid-based methods such as TTDD (two-tier data dissemination) [5], local flooding methods such as ALURP (Adaptive Location Updates Routing Protocol) [6], and Elastic routing [7]. TTDD is a method in which a grid is used to transmit data. A mobile sink node within a local grid cell performs flooding of query messages to transmit them to a dissemination node, which informs the source node of the sink's location. However, a number of overheads occur when the size of a grid cell increases. In ALURP, the sink's location information is updated through flooding within an adaptive area. A source node transmits a data packet to a virtual center, and when this data packet encounters a node within the adaptive area, it receives the latest location information about the sink for transmission. Elastic routing involves updating a sink's location information along the same path but in the direction opposite to that of data transmission. In previous studies, the number of overheads in the location update of a mobile sink node was effectively decreased but not to a significant extent. Therefore, in this paper, we plan to decrease the number of overheads required when updating the location information of a sink node, by setting up an agent between a source node and a mobile sink node using the Elastic routing method.

In this paper, Dynamic MultiAgent-based Local Update Protocol (DMALUP) supporting a mobile sink is proposed. DMALUP is designed to set up one agent per constant hop and transmit data through this agent. Agents can be changed depending on the sink's location so as to support the approximate shortest path. Agent changing occurs when a value greater than the constant critical value is assigned to the straight line connecting the source node and the sink node. At this stage, when the node preceding an agent node has a value greater than the critical value, detection of new agents between the mobile sink nodes is initiated. Since DMALUP transmits the location information of a mobile sink node to the last agent, the number of overheads created by an update decrease. Thus, DMALUP is considered an efficient location information update process. The proposed method is capable of reducing the cost associated when updating a sink's location information and helps identify the approximate shortest path as the agent's location is changed.

The rest of this paper is organized as follows. Relevant past researches are introduced in section 2, and DMALUP is described in detail in section 3. In section 4, performance evaluation using the Qualnet 5.0 simulator is described, and in section 5, conclusions are presented.

2 Related Work

Some protocols propose a method for decreasing the number of overheads required for updating the location information of a mobile sink in wireless sensor networks. Initially, TTDD [5] uses a grid-based routing protocol to support the supply of queries and data to a mobile sink. When an event interested by a source node occurs, a global-grid structure is created in the overall network on the basis of the source node, and a dissemination node is assigned as the intersection point of the grid. A sink node performs flooding of query messages within a grid cell, and a close dissemination node uses the grid to transmit information about the location of the sink to a source node. The data are transmitted by the source node via the same path used to transmit the query messages, but in the reverse direction. TTDD only performs flooding of query messages within a local cell. It is an efficient method, but when the cell size increases, a number of overheads are created because of the aforementioned flooding. In addition, a number of overheads are created during the formation and maintenance of a global grid.

ALURP [6] sets the location of the initial mobile sink node as the virtual center, establishes a purpose boundary, and transmits the location of the virtual center to a source node. A source node transmits data to the virtual center, and a sink node performs flooding of its own location information to an adaptive area within the purpose boundary. Then, when a data packet transmitted by a source node arrives at a node within the adaptive area, the node uses the location information of the updated mobile sink node to transmit data. If a sink node moves out of the purpose boundary, the initial process is repeated. Since ALURP updates the location information of a mobile sink within the adaptive area, the overheads required for transmitting the location information of a mobile sink can be reduced. However, when the adaptive area enlarges, a number of overheads are created.

In Elastic routing [7], a source node uses greed forward to transmit data to a mobile sink node, and the updated location information of this node is transmitted in the reverse direction along the same path used for data transmission. Data transmission in this case is very efficient since the data are forwarded to the location of the new mobile sink node when the location information is encountered in the data transmission path. The service is executed in the order C-B-A, as shown in Fig. 1(a). However, when the sink node approaches node B, as shown in Fig. 1(b), the service is executed in the order C-B-Sink, and when the sink node escapes A's transmission boundary, as shown in Fig. 1(c), the sink node greed-forwards location information to A. Then, a register stores sink's new location information and transmits it to a source node in the order A-B-C. In Fig. 1(d–f), when the sink is reset to a new location, a data packet is greed-forwarded to the sink's new route. Elastic routing always maintains the shortest route to reduce data transmission delay. However, since Elastic routing transmits the sink node's location information to a source node every time the mobile sink node moves, a number of overheads are created. If the number of paths in the data transmission route is increased, more overheads are created.

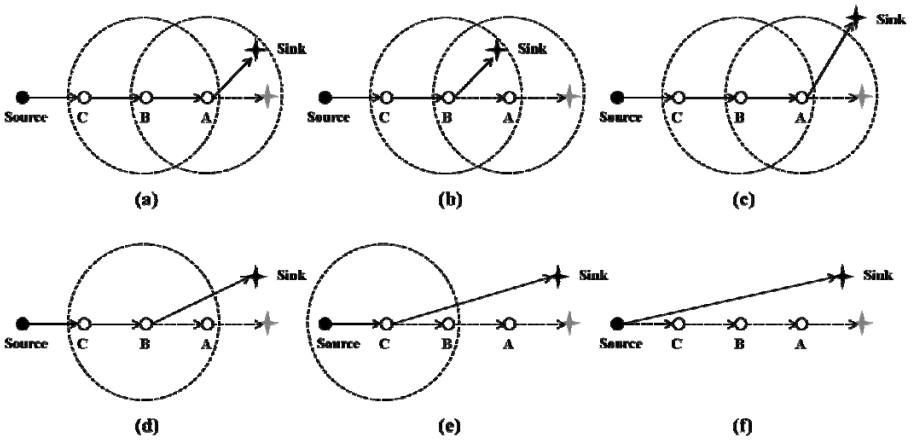


Fig. 1. Location propagation of a mobile sink

3 The Proposed Scheme

In this section, the proposed DMALUP is described in detail. DMALUP works under the following basic network assumptions:

1. The common node and the source node are fixed, and the mobile sink node is free to move very slowly.
2. All the nodes obtain their own location information using the GPS (Global Positioning System) or other location systems. They also obtain the location information of the neighbor nodes through a beacon message.
3. Initially, a source node is aware of the location of a sink node.
4. All the nodes are closely spaced.
5. Designing is executed under the assumption that there is no void area.

DMALUP sets up one static agent per constant hop and uses it to transmit data in order to decrease the number of control messages in the location information service of a constantly moving sink, while effectively transmitting a data packet. The number of control messages can be reduced since the location information of the sink is updated in the last agent.

Now, the path for setting an agent and transmitting data and the strategy for maintaining an agent and updating the location information of a sink node are presented.

3.1 Agents Setting and Data Packet Transmission

DMALUP uses agents to transmit data packets, and these agents assign an agent per constant hop interval between the transmitting source node and the sink node. Then, each agent memorizes the location information of the next agent and transmits the data packet along with this information to the next agent. The last agent then transmits the data packet to the sink node.

The procedures for setting up an agent between the source node and sink node are as follows.

1. A source node, as shown in Fig. 2(a), includes its own ID and location information in the agent list and transmits the sink ID, location information of the sink, hop intervals, and an agent list in the direction of the mobile sink node by using the greed-forward method.
2. The intermediate nodes that receive the message increase the hop and check if it counts as a hop interval. If it does not count as a hop interval, the message is greed-forwarded toward a sink node. If it is counted as a hop interval, the agent list is registered. Then, it adds its own ID and location information to the agent list and greed-forwards the message to a sink node.
3. When a sink node receives the message, it registers the agent list and uses the last agent, as shown in Fig. 2(b), to include its own ID and location information in the next ID and location, respectively, for a transmission.
4. An agent registers the next ID and location information received from the previous agent or sink node and updates its own ID and location information for transmission to the previous agent.
5. Lastly, when a source node receives a message, as shown in Fig. 2(b), it is registered in the next ID and location, thus ending the agent-setting course.

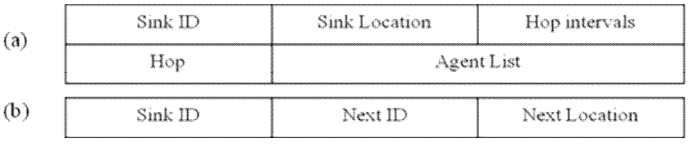


Fig. 2. Control message structure

When a data packet is transmitted, the location information of the next agent is included in the packet being transmitted. Data transmission between agents involves greed forwarding. Lastly, an agent uses the latest location information of a sink node to transmit data to it. If a particular agent lacks energy or does not perform its role, the surrounding nodes act on its behalf.

3.2 Sink Location Propagation and Agent Change

In DMALUP, since the location information of a sink node is updated not in a source node but in the last agent, the number of overheads can be reduced. When a data packet arrives at the final agent, the latest location of a mobile sink node can be confirmed. Such an agent can be dynamically changed. An agent changes in the following three cases: 1) when a sink node grows apart from the last agent, 2) when the sink node is closer to the source node than to the last agent node, and 3) when the agents grow apart from a straight line between the source node and the sink node.

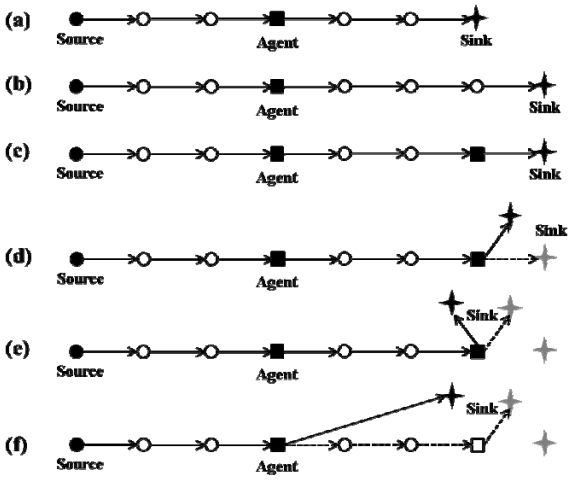


Fig. 3. When the sink node grows apart from the last agent node and the sink node is closer to the source node than to the agent

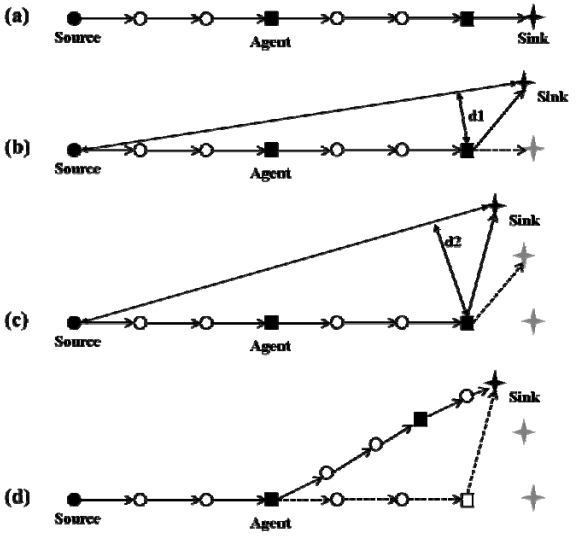


Fig. 4. When the agent node grows apart from a straight line connecting the source node and the sink node

In the initial case, the sink, as shown in Fig. 3(a), provides a service to an agent assigned to every three hops if the sink node grows apart from the last agent and the number of hops increases, as shown in Fig. 3(b). Then, the sink node transmits an agent update message to the final agent. As shown in Fig. 3(c), the last agent sets a new agent before sink node.

In the second case, the sink node is closer to the source node than to the last agent, and the node provides a service, as shown in Fig. 3(d). When the sink node is closer to the source node than to the last agent, as shown in Fig. 3(e), it transmits an agent update message to the previous agent, thus forming a new route from the previous agent.

The third case is when the agent grows apart from the straight line connecting the source node and the sink node. In this case, the purpose is to maintain the approximate shortest path by positioning an agent close to this straight line, even though this is not the shortest complete path. If the agent is configured and data are transmitted, as shown in Fig. 4(a), the sink node moves as shown in Fig. 4(b). In this case, the sink node on the path between the current location and the location of the source node does not change the agent if the distance to each agent, d1, is smaller than the threshold. If d2 is greater than the threshold (Fig. 4(c)), an update message is transmitted to the previous agent. At this stage, the message is transmitted in the direction opposite to that of data transmission. The agent receiving this message forwards it through a path for creating a new agent in the direction of the sink node.

4 Performance Evaluations

We use the Qualnet 5.0 network simulator [8] to evaluate the performance of the proposed DMALUP compared with Elastic routing. Then, we evaluate the performance depending on the maximum distance between the agents and the straight line connecting the source node and the sink node in DMALUP. The system parameters used in this simulation are presented in Table 1. For the sensor node, as shown in Fig. 6, 250 nodes are randomly distributed in a network of area 250 m × 250 m. The radio propagation range is set to 25 m. The IEEE 802.11b radio is used, and MicaZ is used as the energy model. The simulation is performed for 503 seconds. Three seconds of the simulation time is used to set the initial neighbor nodes.

As shows by the arrows in Fig. 5, the routes to the sink node are set to allow diagonal movement. This is to allow for frequent change of the agents. All the nodes transmitted a beacon message every second, CBR data at the rate of 1 *packet/s*, and a total of 500 data packets. Location information update for the mobile sink node is created every time this node moved 1 m away from its previous location.

Table 1. System parameters

Network Area	250m × 250m
Energy Model	MicaZ
Radio Type	IEEE 802.11b Radio
Size of Data Packet	512bytes
Radio Propagation Range	25m
Simulation Time	503s

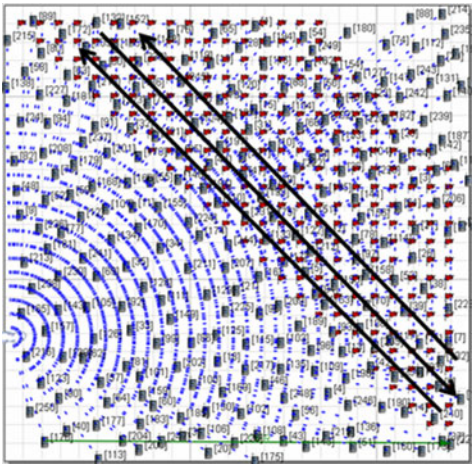


Fig. 5. Placement of nodes in the simulation environment

4.1 Simulation Results for Different Moving Velocities

We evaluate the performance of DMALUP and Elastic routing. The moving velocity of the mobile sink node is changed from 1 *m/s* to 10 *m/s*. DMALUP set an agent every three hops along the transmitting route between the source node and the sink node and the maximum distance of an agent from the straight line connecting the source node and the sink node is set in the range 30 *m*. The data transmission rate is found to be 100%.

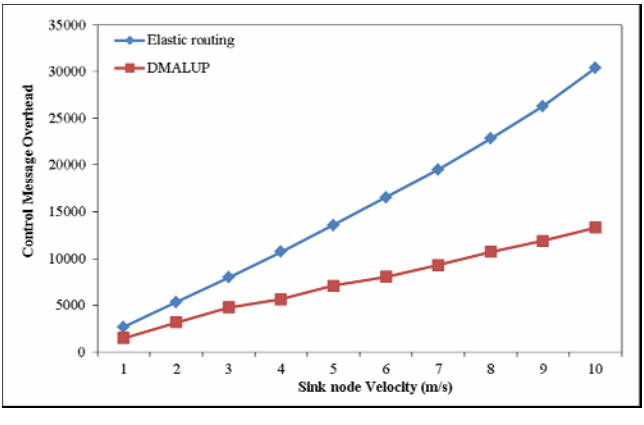


Fig. 6. Comparison of control message overhead

Fig. 6 shows the number of control message overheads in Elastic routing and DMALUP depending on the moving velocity of the mobile sink node. When the moving velocity is 1 *m/s*, the number of control message overheads in Elastic routing and DMALUP is 2680 and 1477, respectively. When the moving velocity is 5 *m/s*, the

number of control message overheads in Elastic routing and DMALUP changed to 13570 and 7086, respectively. When the moving velocity is increased to 10 *m/s*, there are 30385 and 13299 control message overheads in Elastic routing and DMALUP, respectively. On an average, there is a 48% decrease in the overhead in the case of DMALUP. It is found that the location information of the mobile sink node is updated more efficiently in the case of DMALUP.

As shown in Fig. 7, the data transmission delays in the case of DMALUP and Elastic routing are approximately the same. In Elastic routing, the location information of the sink node is transmitted via the data transmission path but in the reverse direction, and message collision causes transmission delay. In DMALUP, the location information of the sink node is transmitted up to the last agent node, thereby reducing message collision. In Elastic routing, the transmission delay shows a zigzag pattern because a greater number of message collisions occur when the moving velocity of the sink node is an even number than when the moving velocity is an odd number, thus causing an increase in the data transmission delay.

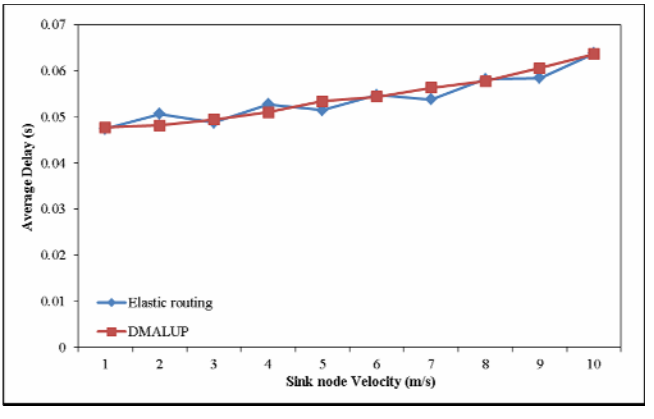


Fig. 7. Comparison of average delay

Fig. 8 shows the energy consume during a message transmission. The energy consume at this stage of transmission includes that consume during the transmission of the beacon message, data messages, and control messages. The overall energy consume during the beacon message transmission is approximately 17.67 *mJ*. When the cost associate with the transmission of a beacon message is excluded, the energy consumption in Elastic routing and DMALUP is 3,813 *mJ* and 3.70 *mJ*, respectively, for a moving velocity of 1 *m/s*. When the moving velocity is 5 *m/s*, the energy consume in Elastic routing and DMALUP is 6.52 *mJ* and 5.3 *mJ*, respectively. When the moving velocity is 10 *m/s*, the energy consumption becomes 11.10 *mJ* and 7.51 *mJ* in Elastic routing and DMALUP, respectively. Thus, we can see that the energy consumption in DMALUP is 2%, 18%, and 32% lower than that in elastic routing for moving speeds of 1 m/s, 5 m/s, and 10 m/s, respectively. No significant difference in energy consumption is seen when the moving velocity of the mobile sink node is low.

However, a significant difference is observed as the moving velocity increases. The reason for the difference in the results depicted in Fig. 6 and 8 is the overhead associated with agent maintenance in DMALUP and the fact that the data transmission route for DMALUP is slightly longer than that for elastic routing.

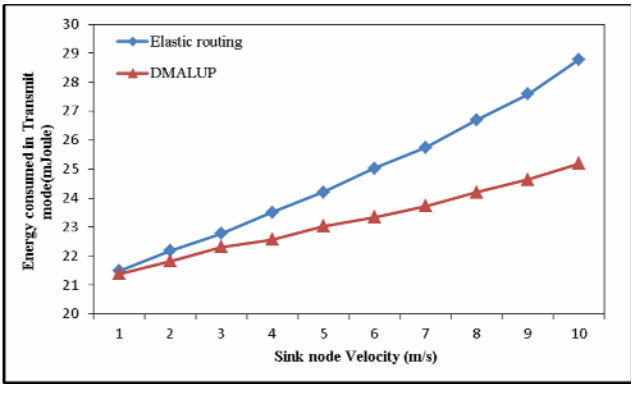


Fig. 8. Comparison of energy consumption during transmission

4.2 Simulation Results Depending on the Maximum Distance of the Agents from Straight Line Connecting the Source Node and the Sink Node

For DMALUP, the maximum distance of an agent from the straight line connecting the source node and the sink node is set in the range 10 *m* to 60 *m*. As shown in Fig. 9, the number of control message overheads is 13769 when the maximum distance is set to 10 m. This number decreases as the distance increases: 6332 at 40 *m*, 6350 at 50 *m*, and 6399 at 60 *m*. This is because the distance to the agent to be updated increases when the agent is updated. As shown in Fig. 10, the reason for the increase in the transmission delay with the maximum distance is the increase in length of the data transmission route. Therefore, the energy consumed during transmission is as shown in Fig. 11.

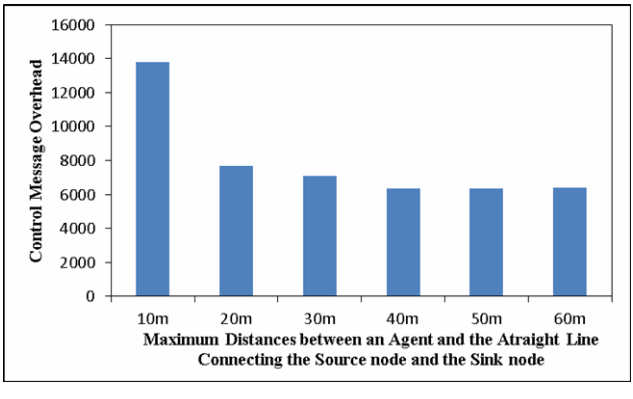


Fig. 9. Control message overhead as a function of the maximum distance between an agent and the straight line connecting the source node and the sink node

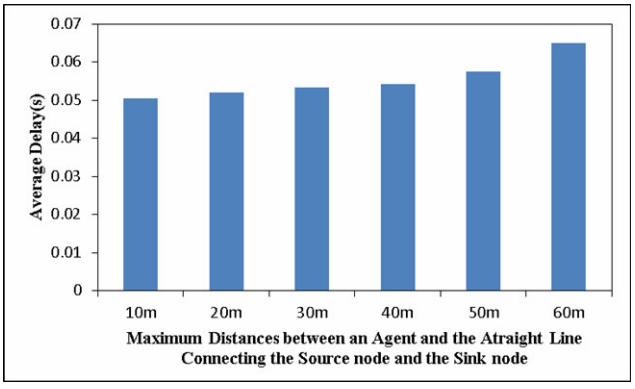


Fig. 10. Comparison of the average transmission delays for various maximum distances between an agent and the straight line connecting the source node and the sink node

From this result, it can be conclude that frequent agent updates caused by a small maximum distance between an agent node and the straight line connecting the source node and the sink node result in high cost. In addition, a very large maximum distance increases the route through which data are transmitted as well as the agent update route, thus increasing energy consumption.

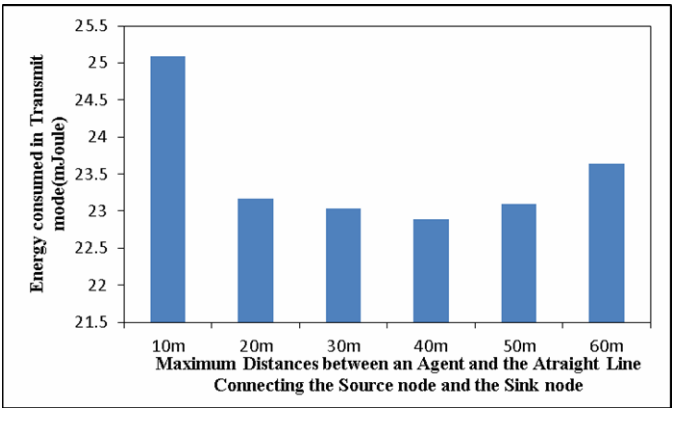


Fig. 11. Comparison of transmission energy consumption for various maximum distances between an agent and the straight line connecting the source node and the sink node

5 Conclusion

We proposed a method in which dynamic agents are used in wireless sensor networks for updating the location information of a sink. In DMALUP, agents are dynamically assigned between the source node and the sink node for data packet transmission, and the last agent is used to update the location information of the sink node. In addition, an agent can undergo dynamic change to create the proximate shortest path, even

though it would not be the shortest complete path. In DMALUP, there is a short path for updating the location information of the sink node, and hence, the number of overheads is decreased. Further, the agent is dynamically changed and the shortest proximate route is created, and hence, the delay in data packet transmission is decreased.

The performance of DMALUP was compared with that of elastic routing using the Qualnet 5.0 simulator. In DMALUP, the maximum distance between the agent and the straight line connecting the source node and the sink node was evaluated as well. DMALUP has reduced the control message overhead than elastic routing up to approximately 48%, and has reduced energy consumption up to 32% when velocity is 10*m/s*. Therefore, DMALUP was concluded to be better than elastic routing in terms of control overhead and energy consumption. In DMALUP, when the maximum distance between the agent and the straight line connecting the source node and the sink node was very small, frequent agent changing was required, and hence, a large number of overheads was observed. On the other hand, when the maximum distance was very large, the DMALUP performance decreased owing to the extension of the data packet transmission route.

Acknowledgments

This research is supported by basic science research program through the National Research Foundation of Korea funded by the ministry of education, science and technology (grant No. 2010-0007311). The corresponding author is Professor K. Park.

References

1. Akyildiz, I.F., et al.: Wireless sensor network: A survey. Comp. Networks J. 38(4), 393–422 (2002)
2. Li, J., Jannotti, J., Couto, D., Karger, D., Morris, R.: A scalable location service for geographic ad hoc routing. In: ACM MOBICOM, pp. 120–130 (2000)
3. Mauve, M., Widmer, J., Hartenstein, H.: A survey on position-based routing in mobile ad hoc networks. In: IEEE Network, pp. 30–39 (2001)
4. Karp, B., Kung, H.T.: GPSR: Greedy perimeter stateless routing for wireless networks. In: Proc. of 6th Annual Int'l Conf. on Mobile Computing and Networking, Boston, MA, pp. 243–254 (2000)
5. Ye, F., Luo, H., Cheng, J., Lu, S., Zhang, L.: A two-tier data dissemination model for large-scale wireless sensor networks. In: Proc. of ACM MOBICOM, pp. 148–159 (2002)
6. Wang, G., Wang, T., Jia, W., Guo, M., Li, J.: Adaptive location updates for mobile sinks in wireless sensor networks. J. Supercomputing 47(2), 127–145 (2009)
7. Yu, F., Park, S., Lee, E., Kim, S.: Elastic routing: A novel geographic routing for mobile sinks in wireless sensor networks. IET Communications 4(6), 716–727 (2010)
8. Scalable Network Technologies website, `http://www.scalable-networks.com`

Multipath-Based Reliable Routing Protocol for Periodic Messages on Wireless Sensor Networks

Hoai Phong Ngo and Myung-Kyun Kim

School of Computer Engineering and Information Technology, University of Ulsan, Ulsan, Republic of Korea
mkkim@ulsan.ac.kr

Abstract. Reliable transmission is one of the most critical requirements in industrial distributed control systems. This paper proposes a multipath-based reliable routing protocol that can guarantee a specified end-to-end target packet reception rate. In the route discovery phase, the path has highest end-to-end packet reception rate is set up and chosen to transmit data packets. To maintain the target packet reception rate, the destination node monitors periodically its actual packet reception rate and sends a feedback control message to the source if its reception rate drops below the target rate. If the source node receives the feedback message, it adds more paths while maintaining the existing paths to keep the end-to-end packet reception rate above the specified target rate. The performance of the proposed protocol has been analyzed by simulation using QualNet simulator. The simulation result has shown that the proposed protocol has a better packet reception rate than the previous routing protocols and maintains a target packet reception rate in a network where the amount of traffic changes dynamically.

Keywords: Wireless Sensor Networks; Reliable, Multipath Routing Protocol.

1 Introduction

Recently, there have been many researches on wireless sensor networks (WSNs) to be used in industrial distributed control systems. In these systems, a sensor periodically senses the physical environment and transmits the collected data to a controller, which perform a control operation based on the value received from the sensor. For stable control operation, those messages from sensors to actuators have to be transmitted reliably to the destination, and if the packet reception rate of the actuator becomes lower than a certain threshold, the performance of the control system degrades greatly [1]. However, it is difficult to provide a reliable message transmission in WSNs because higher error rate and wireless channel characteristics change dynamically over time [10]. There have been many approaches to provide reliable message transmissions on WSNs [2, 3, 5, 6], but they do not support periodic messages explicitly and do not guarantee a certain level of end-to-end packet reliability.

This paper proposes a multipath-based reliable message routing protocol for periodic messages on a WSN which guarantees a specified end-to-end target packet reception rate. For reliable message transmission for periodic traffic, the proposed

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 197–208, 2011.

protocol first finds a path has highest end-to-end packet reception rate and is higher than the target packet reception rate, and transmits data packets periodically through this path. To maintain the target packet reception rate, the destination node monitors periodically its actual packet reception rate and transmits a feedback control message to the source if its reception rate drops below the target value. If the source node receives the feedback message, it adds additional paths such that the sum of end-to-end packet reception rates of all the paths become greater than the target reception rate. The performance of the proposed protocol has been analyzed by simulation using QualNet simulator. The simulation result has shown that the proposed protocol has a better packet reception rate than the previous routing protocols and maintains a target packet reception rate in a network where the amount of traffic changes dynamically.

The organization of the paper is as follows. Section 2 describes related works on reliable message transmission in WSNs. Section 3 describes the protocol proposed in this paper. Section 4 describes a performance evaluation of the proposed protocol and the conclusion of the paper is described in section 5.

2 Related Works

Multipath-based reliable communication on WSNs has been paid much attention to be applied to WSNs in industrial environment [1, 7]. But, wireless links tend to be unreliable compared to wired links due to factors such as interference, attenuation, and fading, and this is more severe in wireless networks using low power like WSNs [8, 10]. Zhang et al. [9] proposed Multipath Source Routing (MSR) protocol which is a multipath-based routing algorithm based on DSR (Dynamic Source Routing). MSR inherits the advantages of DSR and uses the detection mechanism to obtain the route information and update the invalid path. In MSR, the multipaths are assumed to be completely independent when the network load is distributed, however, in route discovery phase, it cannot always find the completely independent paths. Instead, it just attempts to find the maximum node-disjoint paths, and it increases the computational complexity. Mainaud et al. [2] proposed MAODV-SIM which is based on AODV and uses multiple routes called the “emergency paths” from a source to a destination during the control message exchange. In MAODV-SIM, each link measures and maintains the Signal Intensity Metric (SIM) of the link. MAODV-SIM finds the smallest SIM value among all the links in each of the multiple paths, and chooses the path of which the smallest SIM is the highest. When the path is broken while transmitting data, another path is chosen immediately among the emergency paths, so the overhead and latency due to the link failure can be decreased efficiently. The problem of MAODV-SIM is that SIM is not a good metric to measure the reliability of the paths and the path with the smallest SIM value being the highest does not mean that the most reliable path. MP-MAODV [11] uses two node-disjoint routes to improve network efficiency and balance the network loads by distributing traffic. This approach decreases the number of route discovery and reduces routing overhead. However, because MP-MAODV protocol chooses the shortest paths for real-time communication, and so the paths are not guaranteed to be stable. K. Guan et al. [12] proposed a novel energy-efficient multipath routing protocol for WSNs. In their approach, the source and the destination broadcast RREQ packets concurrently to build trees rooted at source and destination. Multiple intermediate nodes, which

belong to and are shared by the two trees, are found during this flooding. Then, each shared node transmits control packet to the source and destination to establish the route. This is a novel way to establish multipath from the source to destination. However, the drawback of this approach is that the established routes are not guaranteed to be stable and also disjoint.

3 Multipath-Based Reliable Message Routing of Periodic Messages

This section describes a multipath-based reliable message routing protocol proposed in this paper. The proposed protocol uses Packet Reception Rate (*PRR*) as the link cost metric, evaluates the end-to-end packet reception rate for each path through flooding mechanism. The proposed protocol also uses an end-to-end feedback control mechanism to maintain a specified target end-to-end *PRR*.

3.1 Reliable Message Transmission Based on End-to-End Packet Reception Rate

The PRR of a link between node *x* and *y*, *PRR(x, y)*, is defined as follows:

$$PRR(x,y) = \frac{NumberOfSuccessfulPacketTransmissionsOver(x,y)}{TotalNumerOfPacketTransmissionsOver(x,y)}$$

Each node in the network collects the PRR value of each link to its neighbors and periodically updates the PRR values according to the change of the traffic in the network. Given a network with PRR in the links, the end-to-end PRR of a path from a source node *S* to a destination node *D*, *e2e_PPR(S,D)*, is defined as follows:

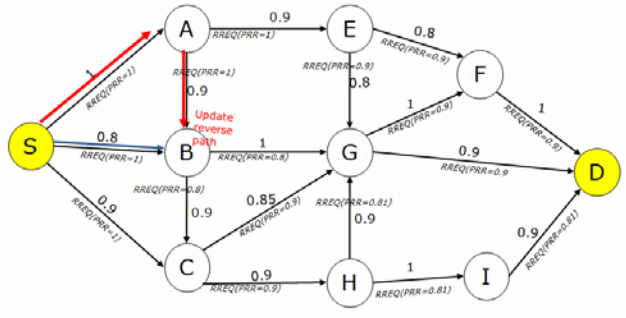
$$e2e\_PRR(S,D) = \prod_{(x,y)\in path(S,D)} PRR(x,y)$$

path(S,D) denotes a set of successive links in the path from node *S* to *D* such as:

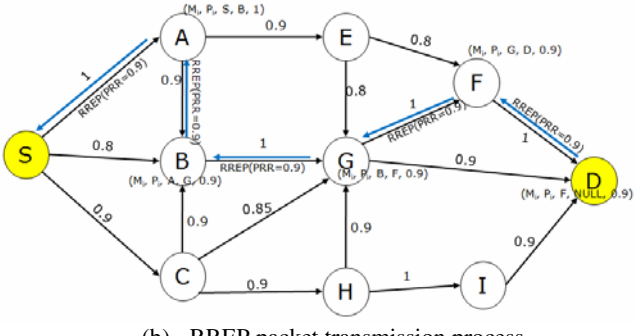
$$path(S,D) = \{(S, X_1), (X_1, X_2), \ldots, (X_{k-1}, X_k), (X_k, D)\}.$$

A reliable path from S to D is a path with a high *e2e_PRR* value. To find a reliable path for periodic message traffic, the source transmits a *RREQ*(M_i, P_i, *PRR, T_PRR, ttl)* packet to its destination by flooding where M_i, P_i, *PRR, T_PRR, TTL* is the periodic, period, *PRR,* required target reception rate (*T_PRR)*, and time-to-live (*ttl)* fields, respectively. The *PRR* field contains the product of the PRRs of the links over which the *RREQ* packet has been transmitted. During the flooding of *RREQ* packets, when an intermediate node *y* receives *RREQ* packet from node *x*, it decreases *ttl* value by one and drops the packet if it becomes 0. Otherwise, node *y* updates the *PRR* value of the *RREQ* packet by multiplying *PRR(x,y)*, and searches its message queue if the message M_i is in the queue. If the message M_i does not exist in the queue, it stores the message M_i in its message queue and broadcasts *RREQ* packet. If the message M_i is in the queue, node *y* compares the *PRR* in the *RREQ* packet with the PRR of M_i in the message queue. If the *PRR* in the *RREQ* packet is greater than the PRR of M_i in the message queue, then node *y* replaces the message M_i in the message queue with the received *RREQ* packet and broadcasts the *RREQ* packet. If the *PRR* in the *RREQ* packet is smaller than or equal to the PRR of M_i in the message queue, node *y* drops

the received *RREQ* packet. If *RREQ* packets arrive at the destination according to this flooding process, the PRR fields of the packets have the product of the PRRs of the links over which the *RREQ* packets have been transmitted. Instead of immediately transmitting RREP packet when receiving a *RREQ* packet, the destination waits for a period of time ($\Delta waitRREQ$) to collect multiple *RREQ* packets. When $\Delta waitRREQ$ is expired, the destination chooses the most reliable path by selecting the *RREQ* packet with the highest end-to-end PRR value and replying with a RREP packet.



(a) *RREQ* packet transmission process



(b) RREP packet transmission process

Fig. 1. An example of route discovery

Fig. 1(a) shows an example of route discovery by exchanging *RREQ* and *RREP* packets between a source *S* and a destination *D*. In the figure, the value on a link is the PRR value of the link. Source *S* floods $RREQ(M_i, P_i, PRR=1, T\_PRR=0.8, ttl=10)$ packet toward destination *D*. Node *A*, *B*, and *C* receive this *RREQ* packet first time then decrease *ttl* by one and update the *PRR* field in the packets, store the packets in their message queues and transmit $RREQ(M_i, P_i, 1, 0.8, 9)$, $RREQ(M_i, P_i, 0.8, 0.8, 9)$ and $RREQ(M_i, P_i, 0.9, 0.8, 9)$ packets by broadcasting, respectively. These nodes also set up the PRR values up to the source *S* in their message queues and the backward paths for forwarding RREP response packets for the *RREQ* packets. After that, if node *A* receives $RREQ(M_i, P_i, 0.8, 0.8, 9)$ packet broadcasted by node *B*, it decreases *ttl* by one and updates the PRR of the *RREQ* packet with 0.72 (=0.8*0.9), and compares it with the PRR value of message M_i maintained in its own message queue, which is 1.

Node *A* drops the *RREQ* packet because 0.72<1. On the other hand, if node *B* receives *RREQ(M_i, P_i, 1, 0.8, 9)* packet transmitted by node *A*, it updates the PRR of the *RREQ* packet with 0.9 (=1*0.9) and compares it with the PRR value of message M_i in the message queue, which is 0.8. Node *B* broadcasts the *RREQ(M_i, P_i, 0.9, 0.8, 8)* packet again and replaces the message M_i in the message queue with *RREQ(M_i, P_i, 0.9, 0.8, 8)* packet because 0.9>0.8. Likewise, node *C* drops the *RREQ(M_i, P_i, 0.8, 0.8, 9)* packet broadcasted by node *B*. In this way, if the intermediate nodes receive a *RREQ* packet which has higher PRR value than the PRR value maintained in the message queue, the nodes broadcast the *RREQ* packet again and update the PRR value and the backward path toward the source in the message queue. If the *RREQ* packets are transmitted this way, the destination node will receive *RREQ* packets through multiple paths and their PRR fields have the product of the PRRs of the links over which the *RREQ* packets have been transmitted, that is, the *e2e_PRR(S, D)*. In the example of Fig. 1, the destination node *D* can receive the following *RREQ* packets:

- Path1: *S→A→ E)→F→D*: *RREQ(M_i, P_i, 0.72, 6)*
- Path2: *S→A→ B→G→F→D*: *RREQ(M_i, P_i, 0.9, 5)*
- Path3: *S→B→G→D*: *RREQ(M_i, P_i, 0.72, 7)*
- Path4: *S→C→H→I→D*: *RREQ(M_i, P_i, 0.729, 6)*

Node D chooses among those *RREQ* packets, *RREQ(M_i, P_i, 0.9, 5)* packet which has the highest *e2e_PRR* value and transmits a RREP*(PRR=0.9)* packet to the source where 0.9 is the *e2e_PRR* of the selected route. After receiving the RREP packet, the source node starts to transmit the data packets for the periodic message M_i regularly at the interval of the period P_i. While setting up the path by exchanging *RREQ* and RREP packets, each node in the path maintains an entry *(M_i, P_i, BP_i, FP_i, PRR_i)* for the periodic message M_i in its own message queue, where BP_i and FP_i denote a backward node and a forward node in the path and PRR_i denotes the *e2e_PRR(S, x)* between the source S and itself. Fig. 1(b) shows the transmission of RREP packet and a message queue entry maintained by each node in the established path.

3.2 Maintaining Packet Reception Rate Using Feedback Control Mechanism

During the packet transmission, the PRR value of links can be changed due to interference, links broken…etc which in turn affect the *e2e_PRR* of the established path. To reflect dynamic changing of PRR value, each node measures and updates the PRRs to its neighbors at a regular interval. The destination node maintains a target packet reception rate (*T_PRR*) for each periodic message flow, and while receiving data packets from the source, it periodically measures the actual packet reception rate (*A_PRR*) for the flow, and transmits a feedback message, *IncPRR(A_PRR) (Increase PRR)* packet which contains the current actual packet reception rate, if *A_PRR* falls down below *T_PRR*. If the source receives the *IncPRR(A_PRR)* packet, it tries to set up an additional path by transmitting *AddRREQ(M_i, P_i, 1, N_PRR, ttl)* packet by flooding to maintain the required target packet reception rate. *N_PRR* denotes the required additional packet reception rate, which is equal to *T_PRR – A_PRR*. The *AddRREQ* packet is transmitted in the same way as the *RREQ* packet except that only the nodes which are not contained in the existing path participate in the flooding.

Fig. 2 shows the example of *IncPRR* packet transmission and *AddRREQ* packet flooding. Assumes that T_*PRR* is 0.8 (the same with example in Fig. 1) and A_PRR

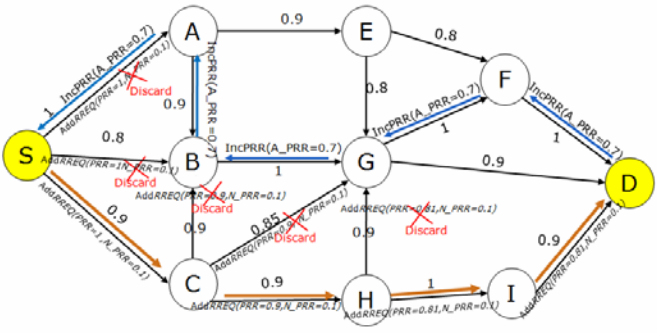


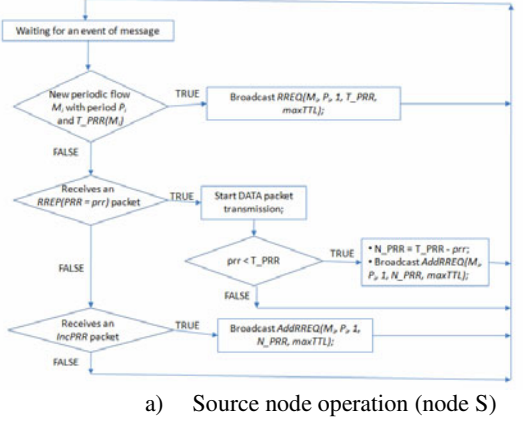
Fig. 2. Packet transmission when the actual PRR below the target value

of current flow falls down to 0.7. In this case, the destination sends *IncPRR*(*A_PRR* = 0.7) back to the source. If the source receives this packet then broadcasts *AddRREQ*(M_i, P_i, *1, N_PRR, ttl) with N_PRR* = 0.8 - 0.7 = 0.1. If the neighbors of the source *S*, *A*, *B*, and *C* receive *AddRREQ*(M_i, P_i, *1, 0.1, ttl)* packet broadcasted by S, *A* and *B* will drop the packet because they are already in the existing path between *A* and *D*. But, *C* receives the *AddRREQ*(M_i, P_i, *1, 0.1, ttl)* packet, it broadcasts the packet again and stores the packet in its message queue. In the example of Fig. 2, the *AddRREQ* packet is transmitted to the destination through the following path:

- Path1: *S*→*C*→*H*→*I*→*D*: *AddRREQ*(M_i, P_i, *0.729, 0.1, 6)*

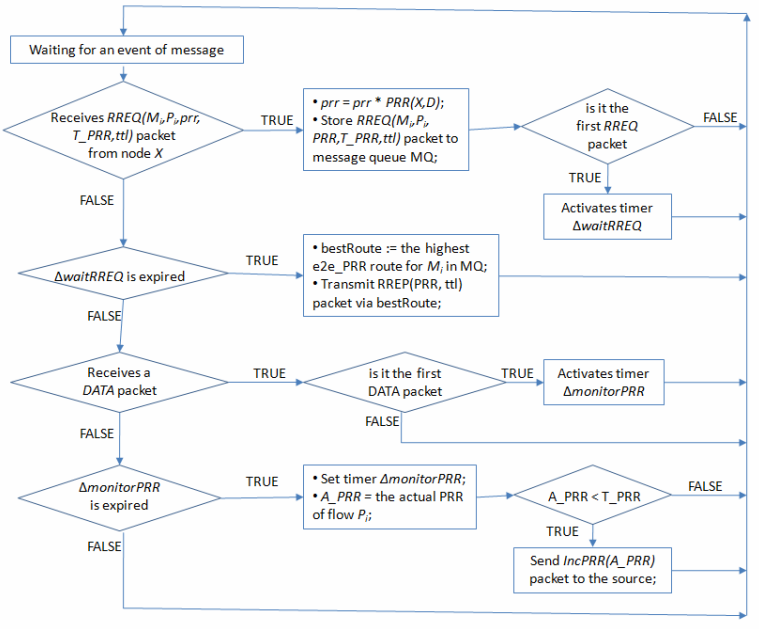
If the destination node receives multiple *AddRREQ* packets, it selects the path which has the highest *e2e_PRR* value and greater than *N_PRR* value (0.729 > 0.1), and transmits a *RREP* packet through the reverse path toward the source. The chosen path by flooding *AddRREQ* packet is a node disjoint path from the existing path, which can minimize the interference between them. If the source receives the *RREP* packet for the *AddRREQ* packet, it transmits its data packet through the two disjoint paths simultaneously, which can enhance the *e2e_PRR* for the data packets.

Fig. 3 shows the operation in each node, the source, destination, and intermediate node, for the proposed protocol.

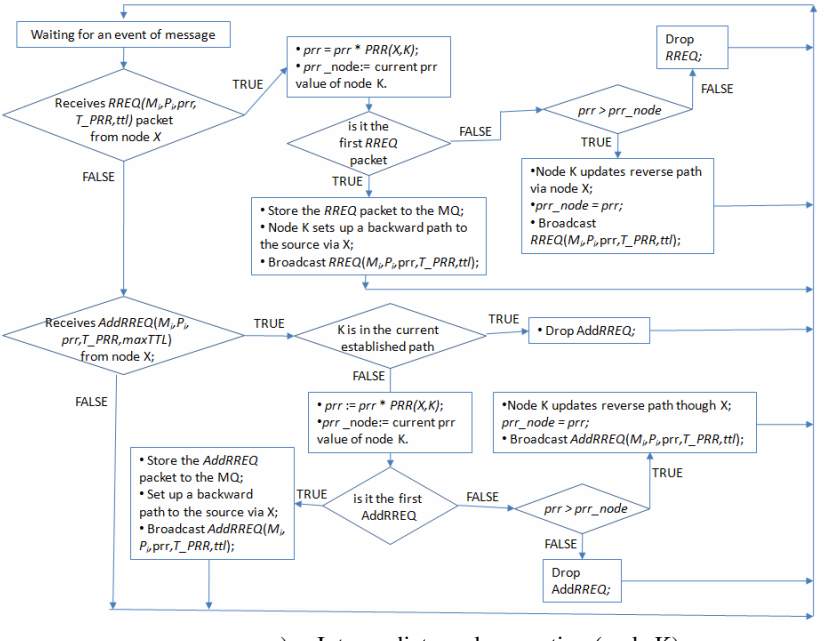


a) Source node operation (node S)

Fig. 3. The protocol operation in each node



b) Destination node operation (node D)



c) Intermediate node operation (node K)

Fig. 3. (*continued*)

4 Performance Evaluation

The performance of the proposed protocol has been evaluated through simulation using Qualnet [13]. The following table shows some parameters of our simulation:

Table 1. Simulation Parameters

Simulation time	30 minutes
Dimension	600m x 600m
Transmission range	100m
Number of data packets transmitted	500
Packet size	512 byte
Data packet interval	400ms – 2s
Waiting time for RREQs (*ΔwaitRREQ*)	250ms
PRR monitoring interval *(ΔmonitorPRR)*	10s
T_PRR	0.9
MAC protocol	802.11 DCF

In our simulation, each link in the network is assigned an initial PRR value, and after that, the PRR value of the link is monitored and updated every when a data packet is transmitted in this link. To initialize the PRR of each link, we conducted the following experiment: 108 nodes are placed at a regular distance of 5m in 4 lines topology in an area 300m x 300m. Node 1 at the central broadcasts 500 packets and the other nodes counted the number of packets successfully received. Then the average PRR value according to the Received Signal Strength (RSSI) value of the link is recorded. Base on this experiment, we initialized the PRR of a link *(x, y)* of which RSSI is *rssi* as *PRR(rssi)*.

The performance of proposed protocol, AODV [4] and MAODV-SIM [2] have been evaluated in terms of the average packet reception ratio at the destination, the average end-to-end delay of packets, the average delay jitter of packets, and total number of transmitted RREQ packets by the source due to the link failure. Two scenarios have been deployed: when there is no contention (1-flow) and when the contention is very high (5-flows), the interval of each periodic flow is set to 400ms. The target e2e_PRR (*T_PRR*) value for the proposed protocol is set to 0.9. We have excuted the simulation while increasing the number of nodes in the network from 75 to 200.

Fig. 4 shows the packet delivery ratio in terms of the node density. When there is no contention (1-flow), the packet delivery ratio of the proposed protocol was almost 100% invariably as the node density increases while the AODV and the MAODV-SIM protocol decrease a little as the node density increases. In the case of 5 flows, the proposed protocol still has maintained a high average e2e_PRR compared to MAODV-SIM and AODV protocols. This result shows that the proposed protocol transmits data packet through a very reliable path. In the case of MAODV-SIM, there is a small variation of the packet delivery ratio in terms of node density due to the several weak links in the chosen path of the protocol.

Fig. 5 shows the performance result of the average end-to-end delay. In this figure, the proposed protocol shows a little larger delay than AODV and MAODV-SIM. The reason is that the proposed protocol chooses a roundabout path with a little more hops to select the most reliable path.

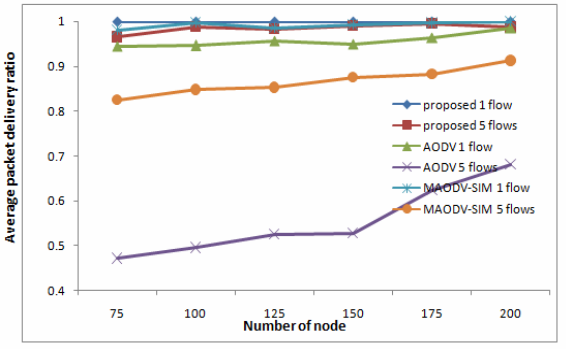


Fig. 4. Packet delivery ratio in terms of node density

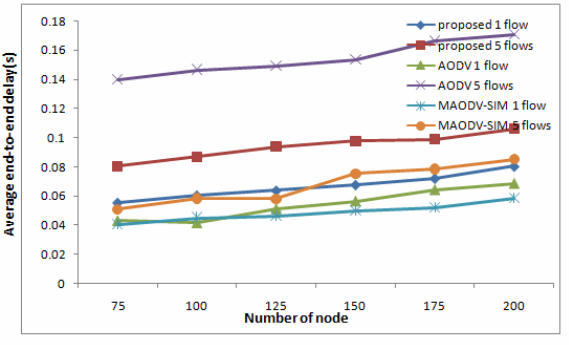


Fig. 5. End-to-end delay in terms of node density

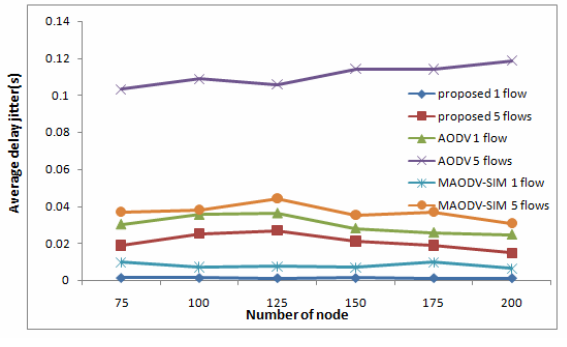


Fig. 6. Delay jitter vs. node density

Fig. 6 presents the average delay jitters of the three protocols. Contrary to the average delay, the proposed protocol showed a very little delay jitter compared with other protocols as the node density increases. In the case of 1 flow, the average delay

jitter of the proposed protocol was almost 0 independent of the number of nodes in the network. In the case of 5 flows, the proposed protocol showed smaller delay jitter compared with other protocols. This result shows that the proposed protocol selects a very reliable and stable path regardless of the node density.

Fig. 7 shows the total number of transmitted RREQ packets by the source due to the link failure. As the figure shows, the proposed protocol has the smallest number of RREQ packet retransmissions. This result shows that the proposed protocol transmits data packets using more reliable and stable paths than the other protocols.

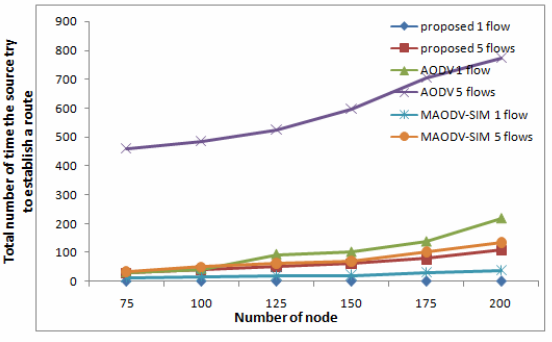


Fig. 7. Number of *RREQ* packets by the source due to link failure

The following simulation shows the capability of maintaining the target packet reception rate of the proposed protocol. After setting up the path for the first flow, we have added additional flows at the interval of 20s until the *A_PRR's* become lower than *T_PRR's* in some flows. The number of nodes in the network is 200. Fig. 8 shows the simulation result when *T_PRR* is set to 0.95 for each flow and the data packet intervals (Δt) of the flows are set to 300ms, 400ms, and 500ms depending on the flow. We can see in the figure that as flows are added more, the *A_PRR* of each flow decreases gradually. As the figure shows, the *A_PRR* of each flow recovers its target e2e_PRR after setting up additional paths using the feedback mechanism.

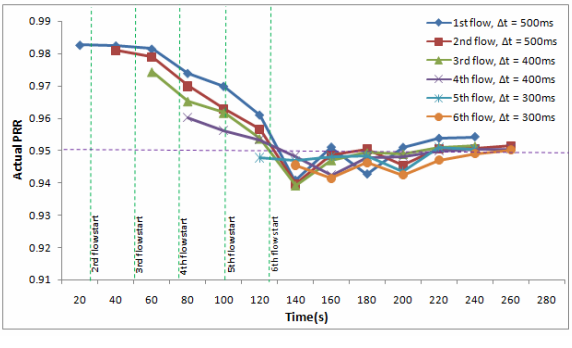


Fig. 8. An example showing the feedback mechanism: T_PRR = 0.95 and data packet interval = 300ms, 400ms, and 500ms.

5 Conclusion

In this paper, we have proposed a multipath-based reliable routing protocol for periodic messages which guarantees a certain level of packet reception rate. For reliable message transmission for periodic traffic, the path with the highest end-to-end packet reception rate is chosen to transmit data packets periodically. To maintain the target packet reception rate, the destination node monitors periodically its actual packet reception rate and transmits a feedback control message to the source if its reception rate drops below the target value. If the source node receives the feedback message, it tries to set up additional paths to the destination to use multi-paths for data transmission. The performance of the proposed protocol has been analyzed and been compared with AODV and MADDV-SIM protocols by simulation using QualNet. The simulation result has shown that the proposed protocol has a better packet reception rate than AODV and MADDV-SIM protocols and maintains a target packet reception rate in a network where the amount of traffic changes dynamically.

Acknowledgments. This work was supported by 2011 University of Ulsan, School of Excellence in Electrical Engineering.

References

1. Gungor, C., Hancke, G.P.: Industrial Wireless Sensor Networks: Challenges Design Principles and Technical Approaches. IEEE Tr. on Industrial Electronics 56(10) (2009)
2. Mainaud, Zekri, M., Afifi, H.: Improving routing reliability on wireless sensors network with emergency paths. In: The 28th International Conference on Distributed Computing Systems Workshops, pp. 545–550 (2008)
3. Nasipuri, Das, S. R.: On-Demand Multipath Routing for Mobile Ad Hoc Networks. In: proceedings of the 8th Int. Conf. On Computer Communications and Networks (IC3N), October 1999, pp. 64–70 (1999)
4. Perkins, Belding-Royer, Das, S.E.: Ad hoc on-demand distance vector (AODV) routing. In: Internet experimental RFC 3561 (July 2003)
5. Sankarasubramaniam, Akan, O. B., Akylidiz, I. F.: ESRT, Event-to-sink reliable transport in wireless sensor networks. In: Mobihoc (2003)
6. Stann, Heidemann, J.: RMST, Reliable Data Transport in Sensor Networks. In: SNPA 2003 (2003); Willig: Recent and Emerging Topics in Wireless Industrial Communications: A Selection. IEEE Tr. on Industrial Informatics 4(2) (May 2008)
7. Willig: Recent and Emerging Topics in Wireless Industrial Communications: A Selection. IEEE Tr. on Industrial Informatics, vol. 4(2) (May 2008)
8. Woo, T.T., Culler, D.: Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks. In: Sensys. (2003)
9. Zhang, Zhao, Z., Shu, Y., Wang, L.: Load balancing of multipath source routing in ad hoc Networks. In: IEEE Int'l Conf. on Communications (ICC 2002), pp. 3197–3201 (2002)

10. Zhao, Govindan, R.: Understanding Packet Delivery Performance in Dense Wireless Sensor Networks. In: Sensys. 2003 (2003)
11. Tang, H., Xue, F., Huang, P.: MP-MAODV: A MAODV-Based Multipath Routing Algorithm. In: IFIP International Conference on Network and Parallel Computing, pp. 296–301 (2008)
12. Guan, K., He, L.M.: A Novel Energy-Efficient Multi-Path Routing Protocol for WSNs. In: Communications and Mobile Computing (CMC), pp. 214–218 (2010)
13. Qualnet simulator, http://www.scalable-networks.com

A Multi-hop Based Media Access Control Protocol Using Magnetic Fields in Wireless Sensor Networks

EuiHoon Jeong[1], YunJae Won[2], SunHee Kim[2], SeungOk Lim[2], and Young-Cheol Bang[1]

[1] Korea Polytechnic University, Jungwang-Dong, Siheung-City, Kyonggi-Do, Korea
[2] Wireless Network Research Center Korea Electronics Technology Institute, Seoul, Korea
ehjeong@kpu.ac.kr, {yjwon,elecsunny,solim}@keti.re.kr, ybang@kpu.ac.kr

Abstract. Recently wireless communication technology using magnetic fields has been proposed as an alternative communication technology for wireless sensor network in unfavorable underwater or underground environment. However, the previous works have limit on communication distance between nodes since they only consider direct communication between nodes. In this paper, we propose a multi-hop based media access control protocol for magnetic field communication to extend communication distance between nodes. The proposed scheme provides a relay node that can relay packets from source nodes to multi-hop distance destination node. We analyze the performance of the proposed scheme by simulation study with qualnet simulator.

Keywords: MAC protocol, magnetic field communication, wireless sensor network.

1 Introduction

Wireless sensor networks are one of the active areas of research. There are a variety of existing and potential applications such as environmental monitoring, infrastructure monitoring, location determination, and border patrol and security monitoring [1]. A lot of applications require underground sensors for monitoring soil conditions, such as water and mineral content, and soil properties. However, there are some limits in applying typical electromagnetic waves to the wireless communication in underground environment [2]. The electromagnetic waves encounter much higher attenuation in soil environment comparing to air environment, which results in poor communication quality. The path loss is determined by both the frequency of the electromagnetic wave and the properties of the soil or rock through which it propagates [1]. Low frequencies show less attenuation than high frequencies over a given distance and soil condition for the electromagnetic waves. The path loss is also dependent on soil type and water contents. As the sizes of soil particles are reduced and the water contents in soil increase, signal power attenuation increases.

Magnetic field communication may be an attractive alternative to the typical electromagnetic waves in the underground environment [3]. Unlike the electromagnetic signals, magnetic field signals are affected mainly by the permeability of medium. Since

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 209–224, 2011.

the permeability of dense medium in the underground, such as soil and water, is similar to air, the channel conditions in the underground environment for the magnetic field signals remain constant. Additionally, since a magnetic dipole loop antenna is used for the magnetic field communication, the antenna can be maintained significantly smaller than that of the electromagnetic waves even with the low frequencies.

Recently, a communication system using magnetic fields has been proposed for the underground environment [3][4][5]. There is also an international standard activity for the magnetic field communication in the name of Magnetic Field Area Network(MFAN) in ISO/IEC JTC1 SC6 WG1 [6][7]. The critical technologies for the wireless communications system using magnetic fields include a physical layer and a media access control layer. Here, we focus on the media access control protocol. The previous media access control protocol using magnetic fields only considers single-hop direct communication. Therefore, the range of communication is bounded to the distance that the signal from a transmitter reaches. Considering that the underground environment enforces comparatively high path loss, it is reasonable to extend the communication distance in a manner of relaying packets with many short-distance hops.

In this paper, we enhance the media access control protocol for the magnetic field communication system to extend the communication distance by adopting a relay node. The relay node can relay packets to communicate with a node in multi-hop distance. The relay node utilizes a relay table to forward packets to the next node and filter unnecessary packet propagation. The remainder of the paper is organized as follows. In Section 2, we provide an overview of the original media access control protocol for the magnetic field communication. Section 3 describes the proposed media access control protocol. In section 4, we analyze the performance of the proposed protocol by a simulation study. We conclude the paper in Section 5.

2 Media Access Control Protocol for Magnetic Field Communication

In this section, we briefly describe the previous media access control protocol proposed for the magnetic field communication [5][6].

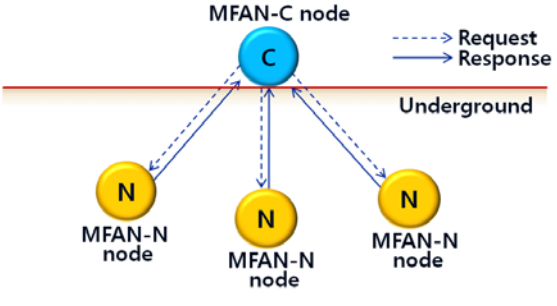


Fig. 1. Magnetic field area network structure

2.1 Magnetic Field Area Network Structure

As shown in figure 1, Magnetic Field Area Network (MFAN) is composed of MFAN-C(Magnetic Field Area Network Coordinator) node and MFAN-N(Magnetic Field Area Network Node) nodes. MFAN-C node is on duty of constructing a super-frame and configuring a MFAN. The super-frame provides synchronization information, network control, and management information such as network join/disjoin request, data request, join status request and group address setup. Any node can be MFAN-C node. However, there must be only one MFAN-C node in a MFAN. All nodes in the MFAN except the MFAN-C node correspond to MFAN-N nodes. A MFAN can have 65,519 MFAN-N nodes maximally.

2.2 Super-Frame Structure

Figure 2 shows the super-frame structure for communication between MFAN-C node and MFAN-N nodes. The super-frame consists of a request period, a response period, and an inactive period. The response period is subdivided into several time slots. The sizes of the three periods in a super-frame are all variable according to the sizes of the packets that are carried on the request period and the response period. MFAN-C node determines both the contents of a super-frame and transmission timing. The request period starts as MFAN-C node transmits a request packet. A request packet includes the command code field that represents request types, such as join request, disjoin request, join state request, data request, and group address setup request.

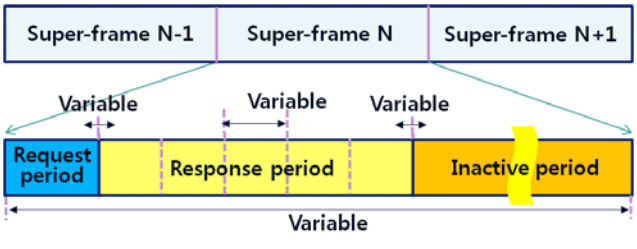


Fig. 2. Super-frame structure

In response to a request packet from MFAN-C node, MFAN-N nodes transmit response packets during the response period of current super-frame. When a MFAN-N node receive a join request packet that asks MFAN-N nodes to join MFAN, the MFAN-N node tries to transmit a join response packet containing it's own unique address in a contention manner. In this paper, we assume that a join response packet is transmitted by a probability based packet transmission scheme. The inactive period starts when there are no MFAN-N nodes that try to transmit packets. The inactive period continues until a new request period starts. During the inactive period, MFAN-N nodes may transmit data packets without the data request from MFAN-C node.

2.3 Protocol Operation

2.3.1 Node Operation

MFAN-C node starts a new request period by transmitting a request packet. MFAN-C node determines the command code in a request packet according to the scheduling strategy that is implementation dependant. After MFAN-C node transmits a request packet, MFAN-C node starts a response period and waits response packets from MFAN-N nodes in MFAN. When MFAN-C node receives a response packet successfully, MFAN-C node interprets the response packet and can transmit an acknowledgement packet to the MFAN-N node according to the acknowledgement option in the response packet. MFAN-C node can start an inactive period if there is no response packet from MFAN-N nodes during the predefined time duration. During the inactive period, MFAN-C node can be in an energy saving mode.

MFAN-N nodes wait a request packet from MFAN-C node. When a MFAN-N node receives a request packet, it interprets the command code in the request packet. The MFAN-N node does the corresponding action according to the command code in the request packet. If the command code is join request and the MFAN-N node has not been joined to MFAN, the MFAN-N node tries to transmit a join response packet in the response period of current super-frame. Here, we assume that each MFAN-N node tries to transmit a join response packet at the start point of each time slot with the transmission probability P_r. At the beginning of each time slot, a MFAN-N node generates a random value between 0 to 1 and compares the random value with P_r. If the random value is less than P_r, the MFAN-N node transmits a join response packet in the current time slot. Otherwise, the MFAN-N node tries to transmit the join response packet in the following time slot by doing the same transmission procedure. After a MFAN-N node transmits any response packet, the MFAN-N node waits an acknowledgement packet for the response packet if the MFAN-N node set the acknowledgement option to receive an acknowledgement packet. Subsequently, the MFAN-N node goes to inactive state until it receives a new request packet from MFAN-C node.

2.3.2 MFAN Join Procedure

MFAN-N nodes are required to be joined to MFAN before they transmit data packets to MFAN-C node. MFAN-C node gives chances of joining MFAN to MFAN-N nodes by a join procedure. MFAN-C node starts the join procedure by transmitting a join request packet during a request period. When a MFAN-N node receives a join request packet, the MFAN-N node tries to send a join response packet at the start point of each time slot of the response period with the transmission probability P_r.

Figure 3 shows an example of MFAN join procedure. In the figure 3, coincidentally, MFAN-N node 1 and MFAN-N node 2 have chances to transmit join response packets in the first time slot of the response period. There is a collision between the two join response packets and MFAN-C node cannot decode any join response packet. In the next time slot, only MFAN-N node 2 has chance to transmit a join response packet and the join response packet is delivered to MFAN-C node successfully. The MFAN-N node that succeeds in transmitting a join response packet is assigned an exclusive time slot in the response period to transmit data packets

without contention with other MFAN-N nodes. Subsequently, MFAN-C node transmits the acknowledgement packet containing the assigned time slot information to the MFAN-N node.

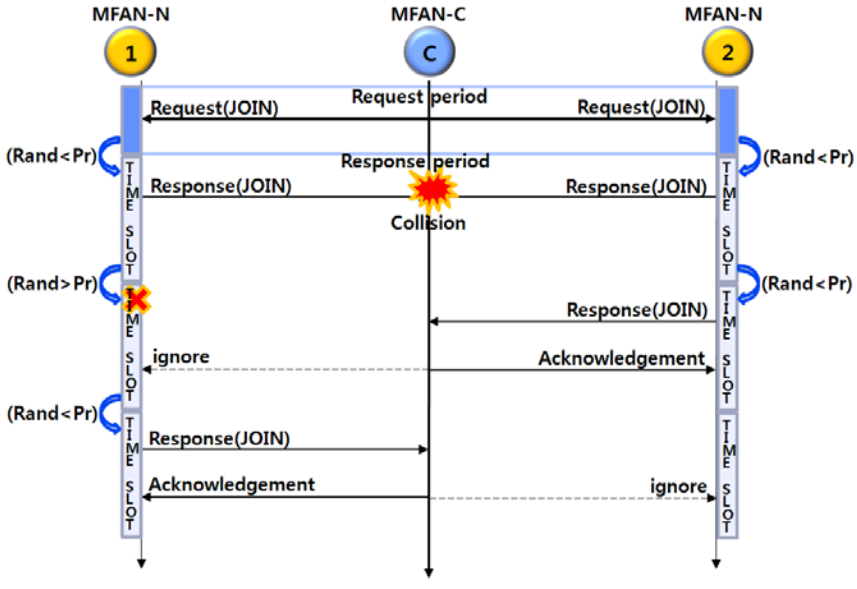


Fig. 3. An example of MFAN join procedure

2.3.3 Data Transmission Procedure

MFAN-C node can transmit a data request packet to ask MFAN-N nodes, already joined to MFAN, to transmit data response packets. The data request packet carries a set of request blocks in the packet payload. Each request block consists of a MFAN-N node address field and a time slot number field. The MFAN-N node address field represents the node address of the MFAN-N node that will be asked to transmit a data response packet. The time slot field represents the time slot number reserved for the MFAN-N node addressed by the MFAN-N node address field. When a MFAN-N node receives a data request packet during a request period, the MFAN-N node parses the data request packet and then checks whether the request packet includes the node's own address. If the MFAN-N node finds it's own address in the data request packet, the MFAN-N node transmits a data response packet in the assigned time slot of the response period. On the other hand, if the MFAN-N node doesn't find it's own address in the request packet, the MFAN-N node waits the next data request packet. Figure 4 shows an example of data transmission procedure. As shown in figure 4, MFAN-N node 1 and 2 have been respectively assigned the second and the first time slot in the response period and respectively transmit data response packets in the second and the first time slot.

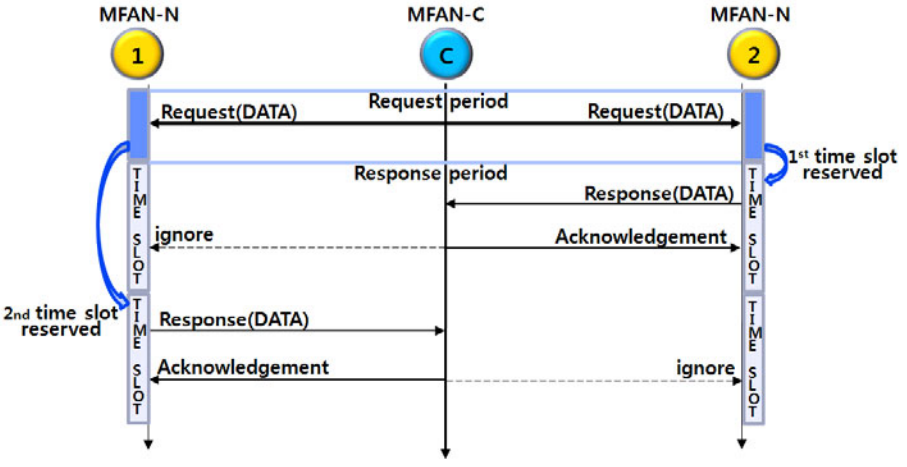


Fig. 4. An example of data transmission procedure

3 Multi-hop Based Media Access Control for Magnetic Field Communication

In this section, we describe the proposed media access control protocol supporting packet relaying through multi-hop MFAN. The packet relaying function is implemented in a new MFAN relay node (MFAN-R).

3.1 Multi-hop MFAN Structure with Relay Nodes

We assume that MFAN is multi-hop network if there is at least one MFAN-N node that should communicate with MFAN-C node via one or more additional node. Here, we propose a relay node (MFAN-R) to extend communication distance. The MFAN-R node has a function of relaying packets to the nodes in the next hop.

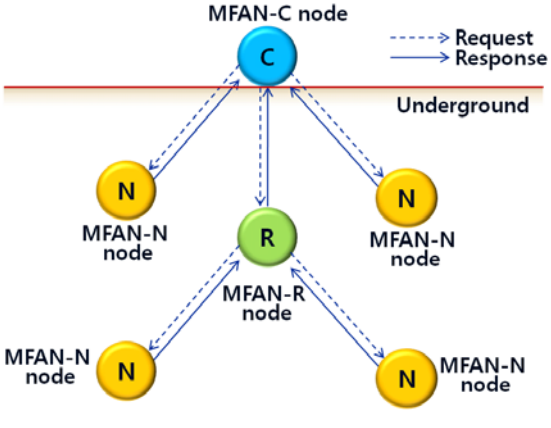


Fig. 5. MFAN structure with relay nodes

Figure 5 shows the structure of multi-hop MFAN. The MFAN in figure 5 is exactly 2 hop network. There is a MFAN-C node in the multi-hop MFAN and the MFAN-C node is responsible for constructing a super-frame and configuring MFAN. MFAN-N nodes just process the request packets received from MFAN-C node and transmit response packets. On the other hand, MFAN-R node relays request packets from MFAN-C node and relays response packets from MFAN-N nodes.

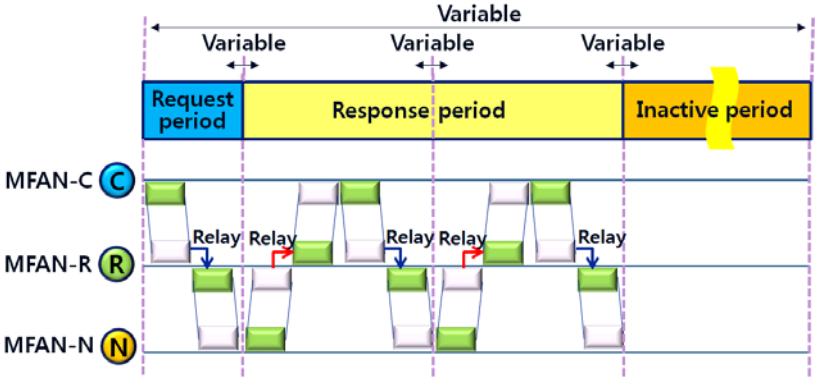


Fig. 6. Super-frame structure for multi-hop communication

3.2 Super-Frame Structure for Multi-hop Communication

Figure 6 shows the super-frame structure for the proposed media access control protocol. The basic structure of the super-frame is similar to that of the previous media access control protocol except the sizes of the request period and the response period. The sizes of the request period and a time slot of the response period in a super-frame expand in proportion to the maximum hop count of MFAN. In case of two-hop MFAN, the sizes of the request period and the response period are two times longer than those of a single-hop MFAN as shown in figure 6.

The durations of the request period (P_{req_join} or P_{req_data}) and the response period (P_{resp_join} or P_{resp_data}) can be calculated from the following equations. Here, we just consider two request types, i.e. join request and data request.

$$P_{req\_join} = D_{req\_join} \times H_{max} . \quad (1)$$

$$P_{req\_data} = D_{req\_data} \times H_{max} . \quad (2)$$

$$P_{resp\_join} = \left((D_{resp\_join} + D_{ack}) \times H_{max}\right) \times N_{total} . \quad (3)$$

$$P_{resp\_data} = \left((D_{resp\_data} + D_{ack}) \times H_{max}\right) \times N_{total} . \quad (4)$$

where, D_{req_join}, D_{req_data}, D_{resp_join}, D_{resp_data}, and D_{ack} represents the sizes of a join request packet, a data request packet, a join response packet, a data response packet, and an acknowledgement packet respectively. The H_{max} represents the maximum hop count in MFAN. The N_{total} means the total number of MFAN-N nodes that have been joined to MFAN.

3.3 Relay Table Management

MFAN-R node performs packet relaying by using a relay table. Each MFAN-R node keeps a relay table and the entries of the relay table are assumed to be set up in a static manner by a network operator. An entry of the relay table consists of two fields, i.e. <MFAN-N node address> and <previous node address>. The <MFAN-N node address> field represents the node address of the MFAN-N node that MFAN-C node wants to communicate with. The <previous node address> field represents the node address of any node to which the current MFAN-R node should relay the response packets originated from the MFAN-N node addressed by the <MFAN-N node address> field.

3.4 Packet Relay Procedure in Multi-hop Environment

3.4.1 Packet Relay Procedure from MFAN-C Node to MFAN-N Nodes

MFAN-C node broadcasts request packets and acknowledgement packets to all nodes in MFAN. MFAN-R nodes should relay request packets and acknowledgement packets to support multi-hop communication. When packets are broadcasted in a wireless network, nodes can receive the same packet several times. Therefore, we need a proper packet filtering strategy to avoid the duplicate packet reception. In this point of view, MFAN-R node relays only the request packets or the acknowledgement packets received from the nodes listed in the <previous node address> fields of the relay table.

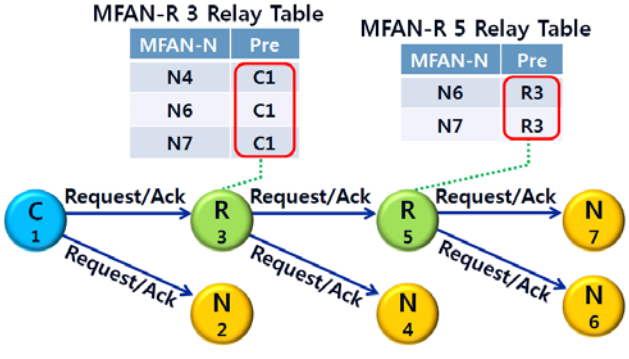


Fig. 7. An example of packet relay from MFAN-C node to MFAN-N nodes

Figure 7 shows an example of packet relay operation in a sample MFAN topology. As shown in figure 7, when MFAN-R node 3 receives a request(ack) packet from MFAN-C node, addressed as C1, the MFAN-R node 3 searches it's own relay table for an entry including C1 address in the <previous node address> field. If the MFAN-R node 3 finds an entry including C1 address, the MFAN-R node 3 relays the request (ack) packet. Otherwise, the MFAN-R node 3 stops relaying the received packet. The request (ack) packet arrived at the MFAN-R node 5 will be processed in a similar manner.

3.4.2 Packet Relay Procedure from MFAN-N Nodes to MFAN-C Node

MFAN-N nodes transmit response packets in response to a request packet from MFAN-C node. The response packets head to MFAN-C node in an uncast manner.

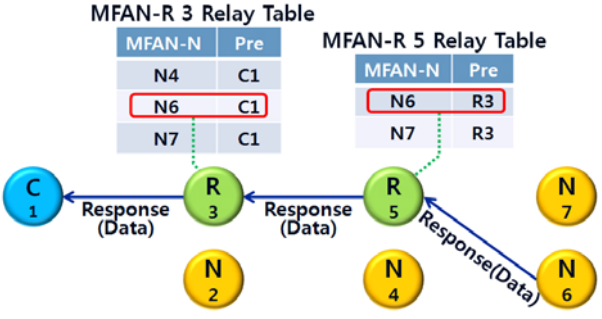


Fig. 8. An example of packet relay from MFAN-N node 6 to MFAN-C node

Figure 8 shows that a response packet from MFAN-N node 6 traverses to the MFAN-C node via MFAN-R node 5 and MFAN-R node 3. When MFAN-R 5 node receives a response packet, MFAN-R 5 node retrieves the address of the MFAN-N node originating the response packet from the relay table. If MFAN-R node 5 finds the matching entry in the relay table, MFAN-R node 5 relays the response packet to the next node addressed by <previous node address> field of the found entry. The response packet arrived at MFAN-R node 3 will be processed in a similar manner. Finally, the response packet will arrive at MFAN-C node.

4 Performance Evaluation

In this section, we analyze the performance of the proposed media access control protocol through the simulation study using qualnet simulator.

4.1 Simulation Environment

Figure 9 shows two types of MFAN topologies assumed for this simulation study. We consider the two types of MFAN topologies to see the influences of both the maximum hop count and the number of MFAN-N nodes in a MFAN on the protocol performances. Figure 9 (a) shows the MFAN topology with varying hop counts. With this topology, we examine the performances of the proposed protocol by varying the hop count from 1 to 6. In this topology, we assume only three nodes in each hop to reduce the effect of the number of MFAN-N nodes. Each hop consists of a MFAN-R node and two MFAN-N nodes except the last hop. The last hop includes three MFAN-N nodes because the nodes in the last hop are not required to relay packets. The MFAN-R node in each hop is responsible to relay packets between the previous hop and the next hop. In addition, we consider another MFAN topology with varying number of MFAN-N nodes as shown in figure 9 (b). We examine the performance of

the proposed protocol by varying the number of MFAN-N nodes from 5 to 15. In this case, we consider a MFAN with just 2 hops to reduce the effect of the hop counts. For each MFAN topology, there is only one MFAN-C node addressed as 1.

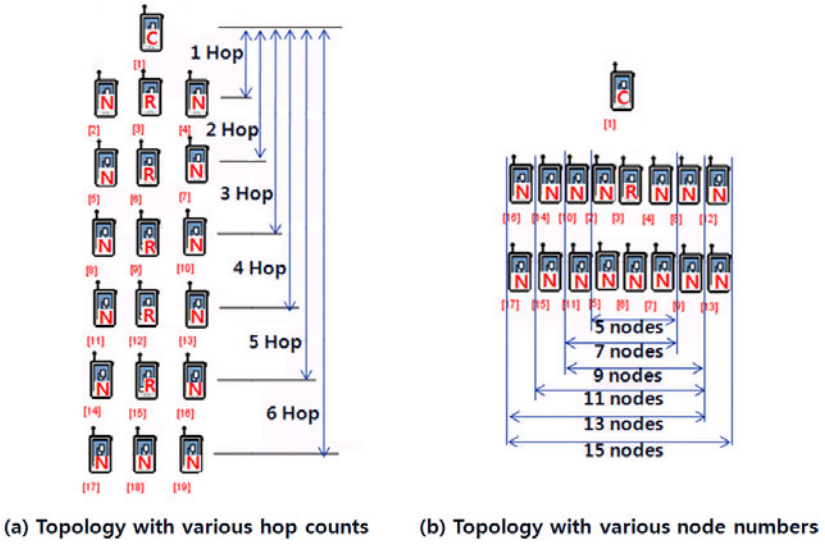


Fig. 9. Two MFAN topologies

System parameters for the simulation study are summarized in Table 1. It is assumed that all MFAN-N nodes generate CBR traffic. The CBR traffic generates a packet every 20 seconds and the size of a packet is 32 bytes. The CBR traffic generation starts at 20 second simulation time and ends at 500 second simulation time. We also assume that MFAN-C node is scheduled to transmit a data request packets to MFAN-N nodes after all MFAN-N nodes in the MFAN have been joined.

Table 1. System parameters

Items	Values	
	Topology (a)	Topology (b)
Super-frame size (sec)	0.7 ~ 16	2.2 ~ 6.2
Data request packet size (D_{req_data}) (msec)	90 ~ 160 / 3 ~ 15 MFAN nodes	
Join request packet size (D_{req_join}) (msec)	80	
Acknowledgment packet size (D_{ack}) (msec)	70	
Join response packet size (D_{resp_join}) (msec)	80	
Data response packet size (D_{resp_data}) (msec)	125	
CBR Traffic parameters	32Bytes, Start:20sec, End:500sec, Interval: 20sec	
Channel	Bandwidth: 5Kbps, Frequency: 300KHz	

We assume that a super-frame consists of a request period and a response period without an explicit inactive period. Therefore, the size of the super-frame can be calculated as ($P_{req\_join} + P_{resp\_join}$) for the join request procedure and ($P_{req\_data} + P_{resp\_data}$) for the data request procedure from the equations (1)~(4). Given the packet sizes of D_{req_join}, D_{req_data}, D_{resp_join}, D_{resp_data}, and D_{ack}, the size of the super-frame can be determined by both the maximum hop count and the number of MFAN-N nodes. As the maximum hop count increases, both the request period and the response period increases in proportion to the maximum hop count. As the number of MFAN-N nodes increase, the size of the data request packet (D_{req_data}) and the number of time slots in the response period increase to accommodate the MFAN-N nodes. The channel frequency and data bandwidth are assumed as 300KHz and 5Kbps respectively according to the specification for MFAN communication system [6].

As mentioned previously, MFAN-N nodes are required to be joined to a MFAN before they start transmitting data packets. Therefore, we examine the join procedure and the data transmission procedure separately to find characteristics of the two procedures. We use join completion time and average packet transmission delay as performance measures. The join completion time means the total time elapsed until all MFAN-N nodes succeed in joining to a MFAN. The average packet transmission delay is defined as the average elapsed time between the generation of a packet at a MFAN-N node and the arrival of the packet at MFAN-C node.

4.2 Join Completion Time

We examine join completion time performance for two MFAN topologies. We also examine the effects of various transmission probability P_r on the join completion time since the join completion time is directly influenced by the transmission strategy for join response packets. We are expected to find an optimal P_r, showing the best join completion time performance, for each MFAN topology.

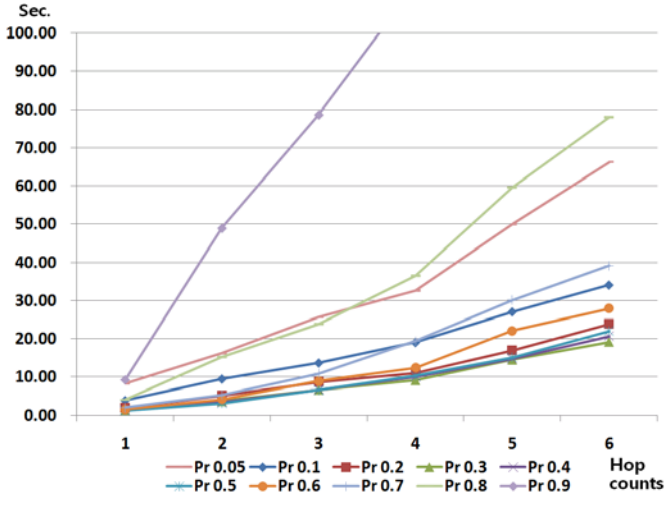


Fig. 10. Join completion time versus the maximum hop count for various P_r

Figure 10 shows the simulation results with the MFAN topology of varying the maximum hop count. Simulation has been performed as the maximum hop count increases from 1 to 6. According to the results, the join completion time increases gradually within the P_r range of (0.1, 0.7) as the maximum hop count increases. The gradual performance degradation is mainly due to the growth of super-frame size rather than the collisions among the join response packets from contending MFAN-N nodes. The sizes of the request and response periods in a super-frame increase in proportion to the maximum hop count as shown in equations (1)~(4). On the other hand, with the P_r of 0.8 and 0.9, the join completion time performance degrades rapidly as the maximum hop count increases. For this case, the rapid performance degradation is mainly due to severe collisions among the join request packets from contending MFAN-N nodes. In addition, the result with P_r = 0.05 also shows comparatively rapid performance degradation. This is because MFAN-N nodes are allowed little chances to transmit join request packets when P_r goes below 0.05.

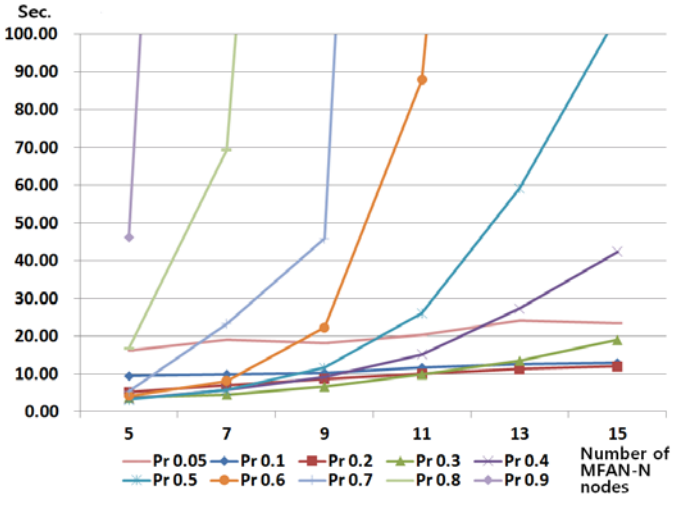


Fig. 11. Join completion time versus the number of MFAN-N nodes for various $\mathbf{P_r}$

Figure 11 shows the simulation results with the MFAN topology of varying number of MFAN-N nodes. Simulation has been performed as the number of MFAN-N nodes varies from 5 to 15 with various P_r. Since we assume that the maximum hop count is fixed as 2, the join completion time performance mainly influenced by the collisions among the join response packets from contending MFAN-N nodes. According to the results, the join completion time maintains very stable state within P_r range of (0.1, 0.3). On the other hand, the join completion time performance deteriorates rapidly due to the severe collisions among the join response packets from contending MFAN-N nodes as P_r approaches 0.4. In case of P_r = 0.05, the join completion time maintains a little high over the whole range of the number of MFAN-N nodes due to the rare chance of transmitting join response packets by MFAN-N nodes. From the above simulation results, we respectively select 0.3 and 0.2 as

optimal P_r values for the MFAN topologies with various maximum hop counts and with various numbers of MFAN-N nodes. The optimal P_r values will be used in evaluating the average packet transmission delay performance.

4.3 Average Packet Transmission Delay

We analyze the impacts of both the maximum hop count and the number of MFAN-N nodes on the average packet transmission delay. We also examine the effect of packet generation interval on the average packet transmission delay. The packet generation interval means the time interval between two CBR packet generations.

Figure 12 shows the simulation results of average packet transmission delay for different maximum hop counts. According to the results, the average packet transmission delay increases as the maximum hop count increases. In general, when a new CBR packet is generated during a super-frame, the packet can be delivered during the response period of the next super-frame. Therefore, the average packet transmission delay increases as the duration of a super-frame increases. The duration of a super-frame is enlarged in proportion to the maximum hop count (H_{max}) as mentioned before. In addition, the number of MFAN-N nodes joined to MFAN (N_{total}) also increases as the maximum hop count increases as shown in figure 9 (a). The increased number of MFAN-N nodes also contributes to enlarge the super-frame duration and degrade the average packet transmission delay performance.

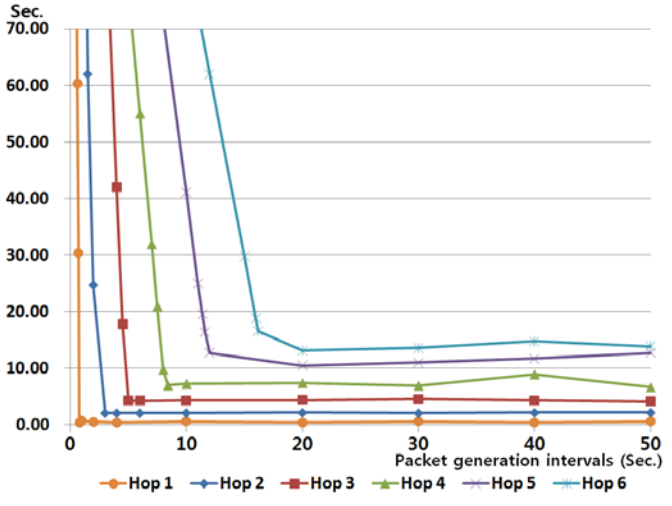


Fig. 12. Average packet transmission delay versus packet generation interval for various hop counts

Considering the impact of the packet generation interval on the average packet transmission delay, we can also find that when packet generation interval is less than the duration of a super-frame, the average packet transmission delay performance deteriorates dramatically. As the packet generation interval becomes less than the super-frame duration, the rate of packet generation becomes greater than the rate of

packet transmission and packets start to be stored in a transmission queue. The increased length of the transmission queue results in drastic performance degradation in terms of the average packet transmission delay.

Figure 13 shows the simulation results of average packet transmission delay for different numbers of MFAN-N nodes. According to the results, the average packet transmission delay increases as the number of MFAN-N nodes increases mainly due to the increase of super-frame duration. According to the simulation results of figure 12 and 13, we can find that the average packet transmission delay is more sensitive to the maximum hop count rather than the MFAN-N node numbers. We can also find that the average packet transmission delay performance degrades rapidly as packet generation interval becomes less than the duration of a super-frame.

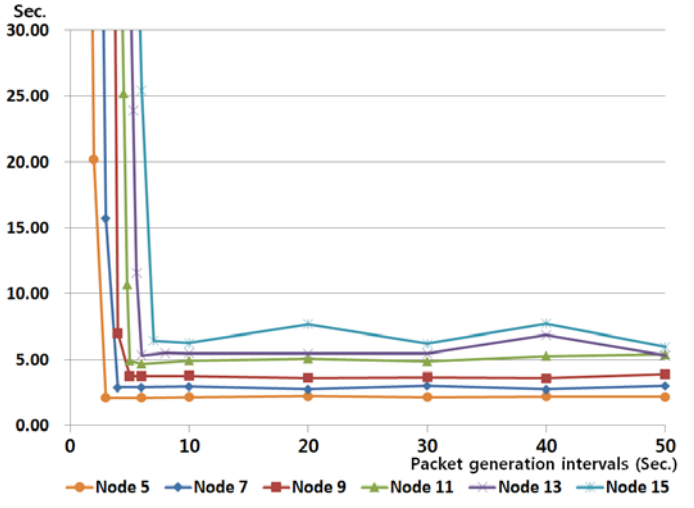


Fig. 13. Average packet transmission delay versus packet generation interval for various MFAN-N node numbers

4.4 Node-by-Node Pattern of Average Packet Transmission Delay

We also examine the node-by-node pattern of the average packet transmission delay performance. Here, we assume the MFAN topology of figure 9 (b). It is also assumed that the number of MFAN-N nodes is 12 and the maximum hop count is 2 and the other system parameters are based on those of Table 1.

Figure 14 shows that the patterns of average packet transmission delay performance are not uniform. This non-uniform performance patterns are caused by the positions of the time slots, assigned for MFAN-N nodes, in the response period. As the position of the assigned time slot is closer to the start point of the response period, the average packet transmission delay performance is better. From this fact, MFAN-C node can schedule the positions of time slots considering the traffic characteristics of each MFAN-N node.

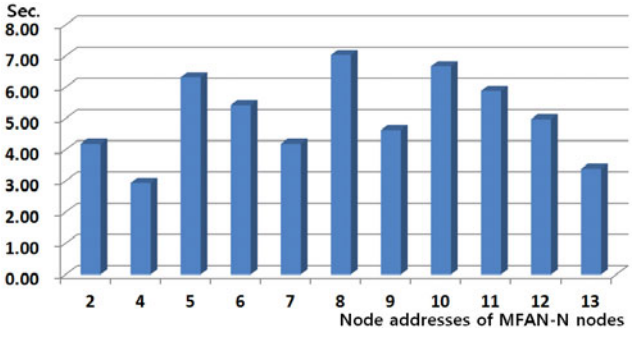


Fig. 14. MFAN-N node specific average packet transmission delay performance

5 Conclusion

In this paper, we enhanced the media access control protocol for magnetic field communication to extend communication distance by introducing packet relay functions. The proposed protocol allows MFAN-C node to communicate with MFAN-N nodes residing in multi-hop distances. We analyzed the proposed protocol in terms of join completion time and average packet transmission delay through simulation study. According to the results, the join completion time increases as the maximum hop count and the number of MFAN-N nodes increase. In steady state, the join completion time is more sensitive to the maximum hop count than the number of MFAN-N nodes. In addition, the join completion time performance degrades rapidly mainly due to the severe collisions among the join response packets from contending MFAN-N nodes as the number of MFAN-N nodes increases. Therefore, it is important to find an optimal transmission probability when a probability based transmission scheme is used for transmitting join response packets. The average packet transmission delay also increases as the maximum hop count and the number of MFAN-N nodes increase. The performance degradation mainly caused by the increase of super-frame duration. The maximum hop count has more effects on the increase of super-frame duration rather than the number of MFAN-N nodes. It is also memorable that packet generation interval should be less than the super-frame duration. Since the node-by-node pattern of the average packet transmission delay is non-uniform according to the positions of the time slots assigned for MFAN-N nodes, a scheduling mechanism considering traffic characteristics can be explorable.

Acknowledgments. This work was supported by the IT R&D program of MKE/KEIT. [10033359, Development of Underground Magnetic Field Communication].

References

1. Akyildiz, I.F., Stuntebeck, E.P.: Wireless Underground Sensor Networks: Research Challenges. Ad Hoc Networks Journal 4, 669–686 (2006)
2. Sun, Z., Akyildiz, I.F.: Underground wireless communication using magnetic induction. In: IEEE International Conference on Communications (ICC2009), pp. 1–5. IEEE Press, Germany (2009)
3. Won, Y.J., Kang, S.J., Kim, S.H., Choi, D., Lim, S.O.: A Communication System Using Magnetic Fields. In: Wireless (VITAE 2009), pp. 265–269. IEEE Press, Denmark (2009)
4. Kim, S.H., Won, Y.J., Lim, Y.S., Seo, K.H., Lim, S.O.: Design of Physical Layer for Magnetic Field Area Network. In: 4th International Conference on Ubiquitous Information Technologies & Applications (ICUT), pp. 293–296 (2009)
5. Jeong, E.H., Won, Y.J., Lim, S.O.: Media Access Control Protocol for Magnetic Field Communication. In: Korea Information and Communication Society Winter Workshops, Korea, pp. 205–205 (2010)
6. Information Technology – Magnetic Field Area Network – Low Frequency Band – Part 2: MAC Layer Requirement, KS X 4651-2 (2009)
7. Lim, S.O., Kang, S.J.: Magnetic Field communication Technology and Standardization Trends. TTA Journal 127, 83–88 (2010)

A Study on Hierarchical Policy Model for Managing Heterogeneous Security Systems

DongYoung Lee[1], Sung-Soo Ahn[1], and Minsoo Kim[2]

[1] Dept. of Information and Communication, Myong-ji College
356-1 Hongeun3-Dong, Seodaemun-Gu, Seoul 120-776, Korea
{dylee,ssan}@mjc.ac.kr
[2] Agency for Defense Development 3rd System Development Center, MEP System Development, Daejeon Korea
mskim1019@paran.com

Abstract. Enterprise security management system proposed to properly manage heterogeneous security products is the security management infrastructure designed to avoid needless duplications of management tasks and inter-operate those security products effectively. In this paper, we defined the hierarchical policy model and the detection algorithm of policy conflict for managing heterogeneous firewall systems. It is designed to help security management build invulnerable security policies that can unify various existing management infrastructures of security policies. Its goal is not only to improve security strength and increase the management efficiency and convenience but also to make it possible to include different security management infrastructures while building security policies. With the process of the detection and resolution for policy conflict, it is possible to integrate heterogeneous security policies and guarantee the integrity of them by avoiding conflicts or duplications among security policies. And further, it provides convenience to manage many security products existing in large network.

1 Introduction

Network management issues include performance management, fault management, accounting management, configuration management and security management. The importance of efficient management of information technology resources grows as systems, networks, and applications become more complex. Management of heterogeneous distributed resources calls for a new methodology to replace explicit control a reactive management with effective, automated, and proactive management. To relieve system and network managers, the level of abstraction needs to be raised in order to hide system and network specifics. Policies which define the desired behavior of resources have been recognized as a concept to support this complex management task by specifying means that enable to enforce this behavior[1-3]. Policies are found at every level of a corporation forming a policy hierarchy, starting from conceptual level through to end system units such as Firewalls, IDS, and other security systems. At all levels they specify the desired behavior of the underlying resources.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 225–234, 2011.

In this paper, we present the hierarchical policy model of ISMS(Integrated Security Management System) consists of five levels which are conceptual/strategic policies level, goal-oriented policies level, target policies level, process policies level and low-level policies. A high-level policy of conceptual/strategic policies level may be used as the basis from which multiple lower-level policies are derived. The derivation can be by refining the goals, partitioning the targets, or delegating the responsibility to other managers. Also, we presented the detection algorithm of policy conflict for managing heterogeneous firewall systems.

The rest of this paper is in four sections. Chapter 2 explains the policy hierarchy and transform for the ISMS. Chapter 3 describes the hierarchy policy model of the ISMS and the algorithm of detection and resolution to policy conflicts. Finally, chapter 4 summarizes our work and discusses possible future work.

2 Related Works

2.1 Policy Hierarchy

When analyzing catalogues of polices from various network and system service providers, it becomes apparent how different policies can be. Security policies specifying the precise format of the allowed password structure or the IP addresses of systems to be protected by firewalls are mixed with abstract polices describing the required availability and accessibility of the computing resources or polices documenting the precautions to be taken when using a specific management tool. It is also essential to structure these polices to guarantee that all policies are applied to their targets(MOs: Managed Objects – which are not in conflict with each other). Therefore, a policy hierarchy is a way of splitting the many policies into smaller groups of different levels of abstraction, which can be further processed in distinct steps and transformed into applicable low-level polices.

What is central to the MSME(Multidimensional Security Policy Management for Dynamic Coalitions) system is the Security Abstraction Layer(SAL). The MSME SAL is based in part on the ISO security architecture(ISO 7498-2)[4] . It includes additional services and mechanisms not defined in ISO 7498-2, for example a transitive communications service between coalition members and a steganography mechanism. The SAL enables policy administrators to think and plan in terms of higher security services that are independent of the specific security implementations. Figure 1 shows the abstraction levels.

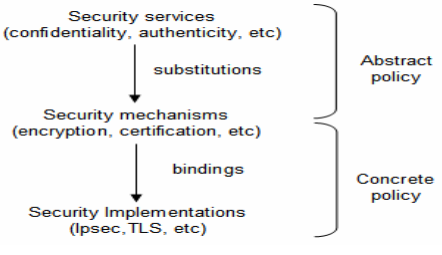


Fig. 1. Conceptual architecture of the ISMS

Thus, although we use security services as the top layer of our abstraction model, it should be noted that most of the resolution techniques and issues that we describe below would be equally relevant to any alternative set of abstract services. That is, certain polices can be assigned to exactly one level of the hierarchy, yet other policies may be assigned to different levels and thus must be split into separate policies before the transformation process can be applied.

2.2 Policy Transformation Process

As we showed Fig. 1, following the definition of polices using the policy classification and each characteristic property can be further detailed to allow a stepwise refinement of the policy. That is, the lower level of abstraction policy will transform into the more precise and detailed policies. After the transformation ends, it is possible for the reached degree of detail cannot be refined further or when a mapping between the value(object, action, etc.) to MOs(managed objects) or management functions of the management systems is possible. It is a process of merging the results from a top-down approach(the polices refinement) with the results from a bottom-up approach(the analysis of available management functionality). For example, if the derived targets or monitor objects can be related to existing MOs or if the management actions to be performed can be mapped to management functions or services, the process of refinements will end. This process will follow the concept of the Fig.1.

3 Policy Modeling of ISMSF

3.1 Policy Hierarchy of ISMS

The ISMS(Integrated Security Management System)[5-7] has an integrated management facility to manage various security systems and consists of three modules – security management client, security management engine, and intelligent agents. The security management client displays conceptual management policies of security services to security managers and sends monitoring or control requests to security management engine. The security management engine receives and processes policies from clients, and stores status information of security systems from intelligent agents. The intelligent agents collect information from security products - Firewall, IDS, VPN, etc) which control MOs(managed objects)in domains and directly control upon the requests from the security management engine. Also, the process of policy application for the ISMS is that at first step towards this was the definition of the term policy. Policies are found at every level of a abstract policy from abstact/conceptual level through to detailed MO(Managed Object)s units. At all levels policies specify the desired behavior of the underlying resources. At the abstract/conceptual policy level, polices are pre dominantly subjective and guided by conceptualism, whereas policies for security systems and network management are technology oriented. The level of abstract/conceptual policy in terms of the desired behavior of distributed heterogeneous security systems, applications, and networks, depends on the degree of detail contained in the policy definition and ratio of business related aspects to technological aspects within the policy. Figure 2 shows the hierarchical policy architectural of the ISMS.

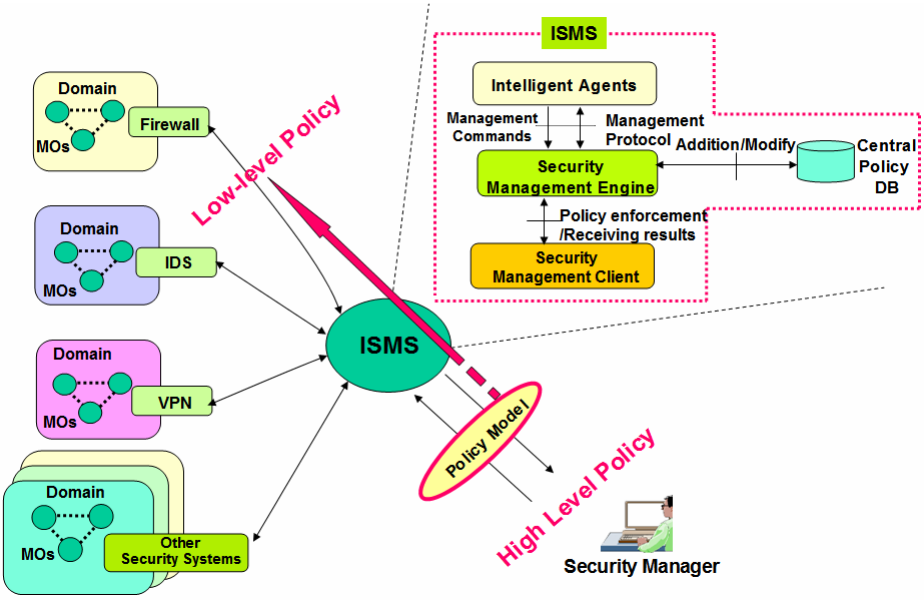


Fig. 2. Hierarchical policy architectural of the ISMS

3.2 Policy Transformation

Policy hierarchies reflect the different available means by which management policies can be specified and realized within a distributed and heterogeneous management environment. Therefore, policy hierarchies in terms of integrated security management are characterized by distributed resource and resources and delegated responsibilities. So that, we present the hierarchical policy architecture of managing the ISMS and Figure 3 describes the hierarchical policy architecture of processing from conceptual/High-level policies to low-level policies.

- Abstract/Conceptual Policies:
  These are directly derived from conceptual goals and thus embody aspects of strategic management rather than aspects of technology oriented management.

- Goal-oriented/Target Polices:
  These are directly enforced and defined the usage of security management functions(e.g Permission, Deny, Monitoring, Authorization, Authentication) for the ISMS.

  - Target Polices:
    These are fragmented expression of policies which is associated with some structured language.(e.g, pseudo-code, macros, Z , other formal language etc)

- Process Polices:
  These are encoded process for agent support(e.g, SNMP, CMIP, Message Format, etc)

• Low-level policies:
Policies are written in languages that MOs(Managed Objects-security systems)understand and replaced with encoded polices that are executable. (e.g, agent software which support security systems)

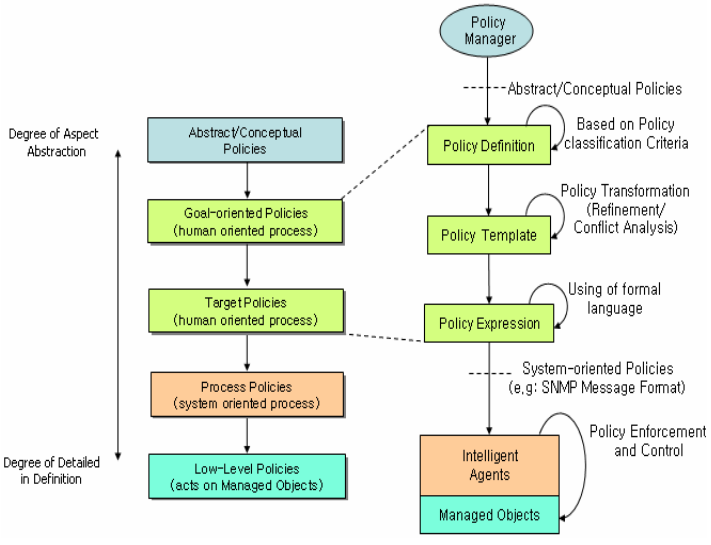


Fig. 3. Hierarchical Policy Architecture and Transformation Process

3.3 Policy Conflicts: Classification and Detection

Whenever several policies are applied to the same target there is potential conflict between them. A preliminary work on conflict classification is reported in[8-11] where different types of conflict are distinguished according to the overlaps among the subject, action and target scopes. In order to guarantee the integrity of policies among policies on each MO, the ISMS engine processes the integrity checking at each time when a user requests the policy manipulation - addition, deletion, modification of policy - against whole policies that ISMS engine currently stores in the database. The ISMS engine performs the integrity check procedure under the following three conditions. We say that there is a policy conflict when any of three conditions are satisfied. The primary function of the central policy database is detecting and resolving policy conflicts. The policy of ISMSF $P(x)$ is defined by the existing *policy(old)* and the newly requested *policy(new)*. A policy $P(x)$ consists of policy manager $M(x)$, Target $T(x)$ and the action of policy $A(x)$. That is, $P(new)$ and $P(old)$ is defined as follows:

$$P(new) = \{M(new), T(new), A(new)\}, P(old) = \{M(old), T(old), A(old)\}$$

Condition 1. Equivalence; $P(old) = P(new)$
An equivalent policy conflict occurs when two policies have the same values of policy manager $M(x)$, Target $T(x)$ and the action of policy $A(x)$. This can be resolved by

the user levels – network manager(NM), general security manager(GSM) and top-level security manager(*TSM*).

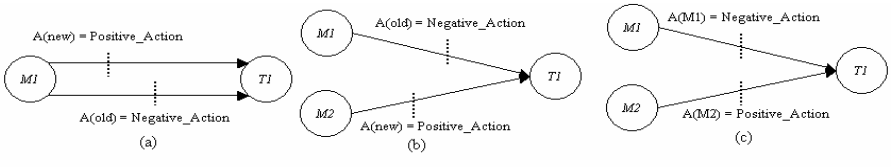
- *Network Manager(NM) : can view the policy, statistical data, log data and other user/host information of his/her network but cannot perform control operations.*
- *General security manager (GSM) : is responsible for the security management of the network. The permission of the GSM to configure the security policy is limited to his/her management scope.*
- *Top-level security manager(TSM) : can manipulate the whole policy of the network organization. The roles of the TSM are harmonizing policies, not to break policy integrity or organizational security policy requirements.*

Condition 2. Contradiction; *P(old) ↔ P(new)*

A contradictable policy conflict occurs when both positive and negative policies of the same kind (i.e, permit policies or deny policies) exist. This conflict is very serious because there are no means of deciding whether the action is to be permitted or denied. A contradictable policy conflict can be resolved by user level and the priority of policy. The classification of contradiction conflict is shown in Table 1 and Fig.4 shows the example of contradictable policy conflict.

Table 1. Classification of contradiction conflicts

M(x); Manager	T(x); Target	A(x); Action
M(old) = M(new)	T(old) = T(new)	A(new) = Positive_Action A(old) = Negative_Action
		A(old) = Negative_Action A(old) = Positive_Action
M(old) ≠ M(new)		A(new) = Positive_Action A(old) = Negative_Action
		A(new) = Negative_Action A(old) = Positive_Action



(a) M(old) = M(new), T(old) = T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_Action
(b) M(old) ≠ M(new), T(old) = T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_Action
(c) M(old) ≠ M(new), T(old) = T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_Action at the same time.

Fig. 4. Examples of the contradictable policy conflict

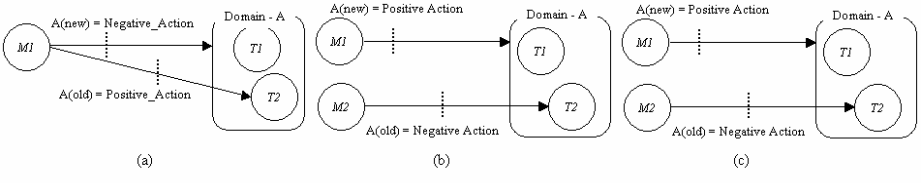
Condition 3. Inclusion; {P(old)⊃P(new) ∨ P(old)⊂P(new)} ∧ {p(old) ↔ P(new)}

Examples of inclusive policy conflict are shown in Figure 5. This policy conflict occurs when the inclusive relationship with contradictable relationship between

existing policy *P(old)* and newly requested policy *P(new)* exist. This problem can also be resolved by user level and the priority of policy. The classification of inclusive policy conflict with contradictable relationship is shown in Table 2 and Fig. 5 shows the example of inclusive relationship policy conflict.

Table 2. Classification of inclusive policy conflicts

M(x); Manager	T(x); Target	A(x); Action
M(old) = M(new)	T(new) ⊂ T(old)	A(new) = Positive_Action A(old) = Negative_Action
		A(new) = Negative_Action A(old) = Positive_Action
	T(new) ⊃ T(old)	A(new) = Positive_Action A(old) = Negative_Action
		A(new) = Negative_Action A(old) = Positive_Action
M(old) ≠ M(new)	T(new) ⊂ T(old)	A(new) = Positive_Action A(old) = Negative_Action
		A(new) = Negative_Action A(old) = Positive_Action
	T(new) ⊃ T(old)	A(new) = Positive_Action A(old) = Netative_Action
		A(new) = Negative_Action A(old) = Positive_Action



(a) M(old) = M(new), T(old) ⊂ T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_Action
(b) M(old) ≠ M(new), T(old) ⊂ T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_Action
(c) M(old) ≠ M(new), T(old) ⊂ T(new) and contradictable relationship between A(new) = Negative_Action and A(old) = Positive_ Action at the same time.

Fig. 5. Examples of inclusive policy conflicts

3.4 Algorithm of Policy Conflicts Detection and Resolution

In this paper, we present the algorithm of policy conflicts detection & resolution for firewalls. The modeling environments are consist of three domain and FW_1(Firewall_1) in Domain1 and FW_2(Firewall_2) in Domain2 are established respectively. Also, the algorithm we presented was limited to firewall and assume a firewall uses packet-filtering routers and Fig. 6 and Table 3 shows to the modeling environment and the policy table of the ISMS respectively.

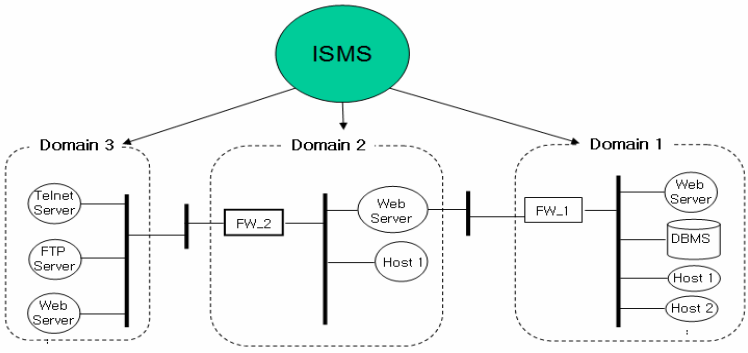


Fig. 6. Modeling environment

(1) Resolution Algorithm of Policy conflict on Condition 1 and Condition 2

As we mentioned in Condtion1 and Condition2, the equivalent policy conflict occurs when two policies have the same values of policy manager M(x), Target T(x) and the action of policy A(x). and the contradictable policy conflict occurs when both positive and negative policies of the same kind (i.e, permit policies or deny policies) exist. And, Fig. 7 shows the algorithm of detecting the contradictable policy conflict and resolving process.

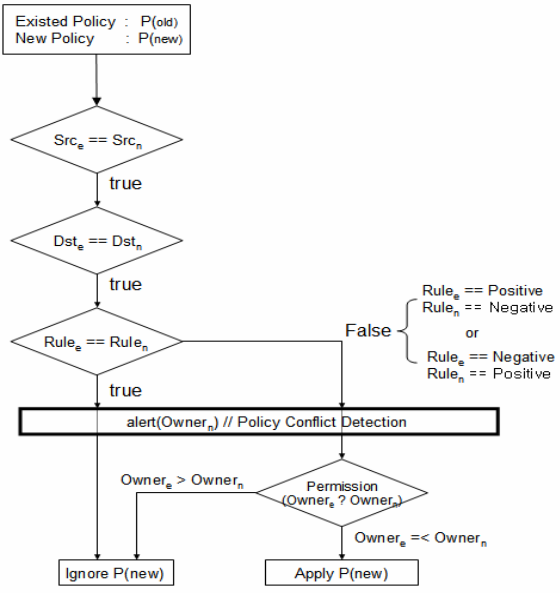


Fig. 7. Algorithm of detecting and resolution for Condition 1 & Condition 2

(2) Resolution Algorithm of Policy conflict on Condition 3

We assume that the Table3 is the ISMS policy table in the modeling environment of Fig. 6. The contradictable policy conflict which is both positive and negative policies of the same kind (i.e, permit policies or deny policies) occurs in the case of No 2, No4, No5 and No6 in the ISMS policy table. And, Fig. 8 shows the state diagram of detecting the inclusive policy conflict and resolving process.

Table 3. ISMS Policy Table

No	Firewall	Direction	Src	Dst	Rule	Policy Conflict
1	FW_1	Outbound	Domain 1 Host 1	Domain 2	Deny	-
2	FW_1	Outbound	Domain 1 Host	Domain2 Host _1	Permit	Collision
3	FW_1	Inbound	Domain 2 Host 1	Domain 1 Web Server	Permit	-
4	FW_1	Inbound	Domain 2	Domain 1	Deny	Collision
5	FW_1	Inbound	Domain 1 Host 2	Domain 1	Permit	Logical Collision
6	FW_2	Outbound	Domain 1 Host 2	Domain 3 Telnet Server	Permit	Logical Collision

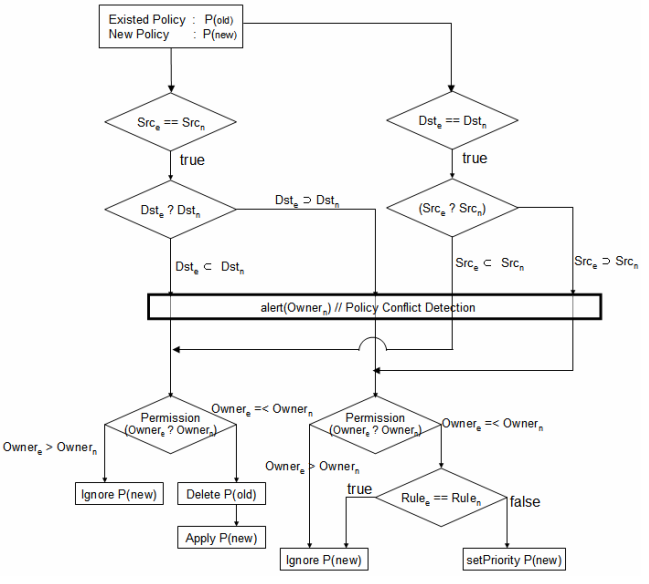


Fig. 8. Algorithm of detecting and resolution for Condition 3

4 Conclusion and Future Works

In this paper, we described the system architecture of central policy management for numbers of firewalls and its merits and illustrated the tree-tired system composed of client, engine and agent. Also, we defined the hierarchical policy model and the detection algorithm of policy conflict for managing heterogeneous firewall systems. It is designed to help security management build invulnerable security policies that can unify various existing management infrastructures of security policies. Its goal is not only to improve security strength and increase the management efficiency and convenience but also to make it possible to include different security management infrastructures while building security policies. With the process of the detection and resolution for policy conflict, it is possible to integrate heterogeneous security policies and guarantee the integrity of them by avoiding conflicts or duplications among security policies. And further, it provides convenience to manage many security products existing in large network.

References

1. An Introduction to Computer Security: The NIST Handbook, NIST Special Publication 800-12 (January 1)
2. A Study on the Development of Countermeasure Technologies against Hacking and Intrusion in Computer Network Systems, KISA final development report (January 1999)
3. Cheswick, W.R., Bellovin, S.M.: Firewalls and Internet Security: repelling the willy hacker. Addison-Wesley, Reading (1994)
4. ISO 7498-2, Information processing systems-Open Systems Interconnection – Basic Reference Model—Part 2: Security Architecture
5. Lee, D.Y., Kim, D.S., Pang, K.H., Kim, H.S., Chung, T.M.: Web-based integrated security management system using SNMP. KNOM Review 2(1), 1167–1171 (1999)
6. Lee, D.Y., Kim, D.S., Pang, K.H., Kim, H.S., Chung, T.M.: A Design of Scalable SNMP Agent for Managing Heterogeneous Security Systems. In: NOMS, April 10-15 (2000)
7. Kim, K.H., Kim, D.S., Chung, T.M.: The Firewall Selection Algorithm for Integrated Security Management. In: Proceedings of The International Conference on Information Networking(ICOIN), February 2003, vol. II, pp. 940–949 (2003)
8. Kim, Y., Song, E.: Privacy-aware Role Based Access Control Model: Revisited for Multi-Policy Conflict Detection (2010)
9. Moffett, J., Sloman, M.S.: Policy Conflict Analysis in Distributed System Management. Journal of Organizational Computing 4(1), 1–22 (1994)
10. Lupu, E.C., Sloman, M.: Conflicts in Policy-Based Distributed Systems Management. Journal of IEEE Transaction on Software Engineering 25(6), 852–869 (1999)
11. Lupu, E., Sloman, M.: Conflict Analysis for Management Policies. In: International Symposium on Integrated Network Management IM 1997, pp. 430–443 (1997)

Hashing-Based Lookup Service with Multiple Anchor Cluster Distribution System in MANETs

Jongpil Jeong*

School of Information and Communication Engineering,
Sungkyunkwan University, Suwon, Korea
jpjeong@skku.edu

Abstract. This paper proposes a novel cluster-based lookup service in mobile ad hoc networks. By exploiting a multi-hop clustering topology, we avoid the empty region problem of grid system and eliminate the need of costly and energy-consuming GPS device. First, we introduce a new size-based hash function, rather than the uniform hash function which is used by most of previous home region based lookup services. Then, we allow multiple bindings between one node and lookup servers, namely anchor cluster-heads which are scattered evenly in the networks by the multiple anchor cluster distribution algorithm. Our extensive performance evaluation confirms that these procedures result in an extremely low communication overhead and a significant reduction of querying delay time. By varying different simulation parameters, we show that the proposed scheme is scalable and adaptive to many networks scenarios.

1 Introduction

In networks, lookup service is a very general and important problem in which a source node wants to query desired information of a specific node, event or service. A significant research has been conducted in peer-to-peer system. However, there is a limited success of this topic in mobile ad hoc networks (MANETs). The challenge is due to the infrastructure-less characteristic and mobility. Unlike normal IP networks which consist of many subnets, MANETs do not impose any hierarchical system by default. Furthermore, all nodes freely move around and constantly change the network topology. These characteristics affect all higher layer protocols including the lookup service.

Lookup service can support any type of information. In routing problem, a source node queries the path to a specific destination. Renowned protocols such as AODV, DSR [1] handle the issue by implementing a flooding-based route discovery. In another case of routing, there is a focus on looking up the location information of mobile hosts [4, 2, 3]. Then, a location-aided protocol is used to improve the performance and scalability of routing. GPS-based system uses geographical coordination while clustering and grid system may use the cluster or grid ID as location information. Mapping between a node ID, or IP address

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 235–247, 2011.

and a hostname is also a specific application of lookup service. Furthermore, in sensor networks, it is desired to obtain sensing value of a specific node or event by using lookup mechanism. Most of the previous work focus on location service and each node has its own location information. However, the mechanism can be used for any type of lookup information.

There are two major types of lookup service, reactive [5,6] and proactive protocols [7]. Reactive lookup service is one major type thanks to its simplicity and extremely low maintenance overhead. It is used in the route discovery process of well known routing protocols such as AODV, DSR [1]. Reactive lookup service does not rely on any hierarchical system and hence has no overhead of such system. However, because reactive scheme is based on flooding, the querying overhead is significant especially when the querying rate increases. For the reactive lookup service, as its cost highly depends on the flooding mechanism, many researches have been conducted to improve the performance of flooding. One of such improvements is Expanding Ring Search [1] in which the flooding scope is expanded gradually to reduce the communication overhead.

In the second category, proactive protocol, the destination information is stored in one or more servers and the source queries the nearest server to obtain the information. KCLS [7], k-hop cluster-based location service, is a proactive scheme which exploits a single level multi-hop cluster structure. In this work, the lookup information is the location of mobile node, the current cluster ID of that node. KCLS [7] is the first paper that exploits a self-adaptive self-organizing clustering without relying on GPS system. KCLS does not use global hashing but stores each information in all available cluster-heads. Each source only needs to query its own cluster-head to obtain the desired information. Because the distance between cluster-head and members is very short, the querying delay of KCLS is thus extremely small. However, this benefit comes with a cost, tremendous update overhead, as each change of information requires update in all cluster-heads.

A special case of proactive scheme is home region lookup service in which each mobile node is associated with a home region by a hash function [4,3]. This mechanism satisfies the scalability requirement of MANET and reduces the information update overhead with the small cost of querying delay time. Nevertheless, most of the researches [4,2,3] assume the existence of a global hash function without any specific procedure. Moreover, these home region protocols rely on the predefined rectangle or cycle system which is calculated from the information of GPS device [4,3]. Unfortunately, many prerequisites have to be met for proper GPS function. The antenna must have a clear view of the sky, making it difficult for using indoors or in urban canyons. The power consumption of such devices greatly shortens the lifetime of the mobile nodes, and greatly increases the cost of each node. Furthermore, using geographical location could be problematic. Empty home regions may exist due to geographic condition. Nodes which are supposed to store information in these home regions have to find alternative regions for that purpose. For the consistency of home region mechanism, every other node also needs to know about the empty home

regions and about the alternative regions of all destinations. Another issue is the dependency of hash function on the topology or grid system. As the hash function takes the backbone topology as input, any change of the topology affects the result of hashing and may require a complete reconstruction. This effect is significant in MANET due to the node mobility. Existing home region lookup services do not concern about these above issues.

Motivated by the challenge of lookup service, we propose a cluster-based hashing lookup service, CHLS, in this paper. Like KCLS [7], our method does not rely on GPS system due to disadvantages such as cost, power consumption, geographical condition, unclear location signal, etc. In contrast, we use the self-adaptive and self-organized multi-hop cluster structure [8,9] on lookup service. Taking the cluster size as input of hashing, we introduce a distributed size-based hash function. Compared with uniform hash function used by most of previous home region lookup services, our hash function results in smaller querying delay time. Besides, unlike previous work which binds each node with only one home region, our scheme allows multiple bindings between a mobile node and clusters. More than that, the multiple anchor-cluster distribution algorithm evenly scatters the bound clusters of one node in the networks so that the information of that node can be obtained with very small querying delay time.

The rest of this paper is organized as follows. In the next section, we propose our lookup service. This part presents about our assumptions, the anchor cluster system, the formula of querying latency and cluster priority. Then, we show the procedure of size-based hash function and the multiple anchor cluster distribution algorithm. Next, in section 3, we discuss the performance evaluation of our scheme. Finally, in section 4, we conclude the paper.

2 Lookup Service

This section presents our lookup service, CHLS, which is based on a self-adaptive self-organizing clustering backbone. In this paper, we use a multi-hop clustering which is similar to 3hBAC scheme [9]. We do not go into details of clustering algorithm here because it is out of scope for this paper.

Generally, in lookup scheme, each information is associated with a key. A source wants to query an information with a specific key. The domain of information key is arbitrary and dependent on the type of lookup information. For example, in DNS service, the information, hostname of the node, is associated with the key, the IP address of that node. Or in sensor networks, each key value can represent a type of specific event and the source may want to query the related information of that event. In this paper, like other work, we assume that each node has its own specific information. It means each information is associated with a node ID. This key domain is mostly used location service in which each node location is the information of interest [4,3]. We present our lookup service with that assumption. However, in the case of general information with a random key domain, we just need to substitute the node ID with the key value for all the procedures of our scheme.

2.1 Anchor Cluster System

We present the procedure and system of our lookup service here. For each mobile node with its own specific ID, there is a corresponding group of clusters which is obtained by performing the hash function on that node ID. This group is defined as the anchor group for a specific node. Clusters of that group are called anchor clusters. Using the cluster topology information and node ID as input on the global hash function, each node obtains its anchor group of clusters. At one time, the global hash function must return the same result by the same input even though it may be performed by different nodes. After getting the hashed anchor group, the node sends UPDATE message which contains a pair of values, its node ID and the attached information, to its anchor group. Let the cluster-head of a specific anchor cluster be anchor cluster-head and make it store all lookup information. In Fig. 1, assume that performing the hash function on node 12 return the result 2. It means the cluster 2 is the anchor cluster of node 12. By the basic mechanism of lookup service, node 12 updates its information to cluster-head 2 which then replies the information to querying sources.

The same process occurs when a source needs to query desired information of a destination. The source first uses the hashing on the destination ID to obtain destination's anchor group. The uniqueness characteristic of hash function ensures the correct result of destination anchor group. Then, the source sends the QUERY message to the nearest destination's anchor cluster and obtains the information of interest by REPLY message.

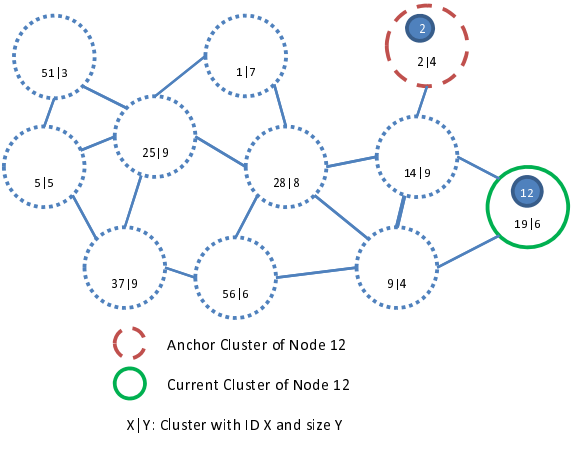


Fig. 1. Lookup Service with Single Binding

Hash functions depend on the cluster topology and this may lead to some issues. The hashing result of every node is changed whenever a cluster is disbanded or formed. Notice that this issue also happens with home region lookup service with GPS device as the home region of one mobile node cannot be accessed

anymore due to connectivity or empty home region problem. To handle this problem, we introduce the interval parameter Δ. Time is divided into periods with length Δ. The lookup service only updates new clusters at the beginning of each interval. Each node recomputes its new anchor group with the new cluster topology and updates its information to anchor cluster-headds. If a cluster is formed in the middle of an interval, it cannot become an anchor cluster of any mobile nodes until the next interval. Furthermore, when an anchor cluster j is disbanded in the middle of the time interval, its associated information is transferred to the next available cluster in the cluster list $L_{cluster}$. For example, in the case of multiple bindings, node 12 may choose clusters 5 and 56 as its anchor clusters by using a hash function. In maintenance phase, assume that cluster 5 is disbanded. Then, cluster-head 5 transfers all its lookup information to the next available cluster-head in the order of cluster ID, cluster-head 9. Being aware that cluster 5 is disbanded, any node that performs hash function and obtains cluster 5 as the result alternates cluster 5 with the next available cluster, cluster 9.

2.2 Cluster Priority

Most of previous home region lookup services use uniform basic hash function on grid system. However, the uniform hash function equally assigns the same number of mobile node to every cluster. The effect of hash function to the performance of lookup service is not evaluated yet. Hence, in this section, we show the relationship between cluster-size and the querying delay time of lookup service. Based on the relationship, we define a priority metric for anchor cluster and anchor group. Our scheme then stores an amount of information one anchor group proportional to the priority value of that group.

Assume that node i is currently in cluster $current\_cluster(i)$ and stores its information in anchor cluster $anchor\_cluster(i)$ $(anchor\_cluster(i) \neq current\_cluster(i))$ by any single binding hash function. There is a query m which is sent to cluster $anchor\_cluster(i)$ for the information of node i. Let $cluster\_size(j)$ be the number of members in cluster j and n be the total number of mobile nodes in the networks. Let $ring\_size(j, u)$ be the total number of members in clusters which are u cluster-hops away from cluster j $(ring\_size(j, 0) = cluster\_size(j))$. Then, the probability that query m is generated by a node which is u cluster-hops away from cluster $anchor\_cluster(i)$ is then $P_u(i)$.

$$P_u(i) = \frac{ring\_size(anchor\_cluster(i), u)}{n - cluster\_size(current\_cluster(i))}$$

Let $T_{msgDelay}(j, u)$ be the mean delay time for a node in a cluster which is u cluster-hops away from cluster j to send a message to cluster-head j $(T_{msgDelay}(j, u + 1) > T_{msgDelay}(j, u)))$. The expected querying delay time $T_{queryDelay}(i)$ for querying information of node i is as follows.

$$T_{queryDelay}(i) = \sum_{u} P_u(i).T_{msgDelay}(anchor\_cluster(i), u)$$

Generally, we want to store the information of node i in the anchor cluster $anchor\_cluster(i)$ with small latency. In the above formula, we see that the querying delay time is low if the probability $P_u(i)$ for small value of u is high. This is due to the low latency $T_{msgDelay}(j, u)$ for small value of u. In practice, the component $cluster\_size(current\_cluster(i))$ is unknown so that the exact value of $P_u(i)$ cannot be measured. However, as $P_u(i)$ is proportional with $ring\_size(anchor\_cluster(i), u)$, we can assign the priority to one cluster by its own size and neighbor-cluster size. Clusters with high priority are also ones with low querying delay time. So, $cluster\_priority(j)$ is calculated for each cluster j. For simplicity, the value $cluster\_priority(j)$ is calculated as the weight sum of its own size and the total size of its one-hop neighbor-clusters.

$$cluster\_priority(j) = \gamma_0.cluster\_size(j) + \gamma_1.ring\_size(j, 1)$$

For a group of clusters G, the priority of the group is the mean of cluster priority

$$group\_priority(G) = \frac{1}{|G|} \sum cluster\_priority(j) \ (with\ every\ cluster\ j\ \in G)$$

2.3 Size-Based Hash Function

Taking the cluster-size into account, we propose a size-based hash function in this section. The hash function consists of two phases. The first phase is used to construct available anchor group and calculate group priority. At the beginning of each time interval, when cluster-heads receive the information of cluster backbone as in section 2.1, cluster-heads perform the following procedure.

Let the cluster list which is already sorted by the order of increasing cluster ID be $L_{cluster}$ with its m^{th} element, $L_{cluster}[m]$ $(0 \leq m < |L_{cluster}|)$. We calculate the priority value for each cluster and obtain the priority list, $Pr_{cluster}$, according to the previous formula of cluster priority.

$$Pr_{cluster}[m] = cluster\_priority(L_{cluster}[m])$$

We perform a group selection algorithm $group\_selection$ for each cluster in $L_{cluster}$. The procedure of this algorithm is delayed until later. For now, we just understand that this algorithm takes a cluster as an input and return a group of clusters which are evenly scattered in the networks. The number of clusters in the result group is predefined by parameter S_{AG}, size of anchor group. The cluster input is also included in the result group. Each group in the list acts as information server for some particular nodes. As the clusters in each anchor group are evenly distributed, a source can query the nearest anchor cluster with low querying delay time. So, the result of group selection algorithm for every cluster is put in the list L_{group} as:

$$L_{group}[m] = group\_selection(L_{cluster}[m])$$

Then, we obtain the list of group priority, Pr_{group}, the accumulated table of priority, Tbl_{group}, as follows:

$$Pr_{group}[m] = group\_priority(L_{group}[m])$$

$$Tbl_{group}[m] = \begin{cases} Pr_{group}[0], & (m = 0) \\ Tbl_{group}[m-1] + Pr_{group}[m], & (m > 0) \end{cases}$$

Next, each cluster-head sends the cluster group list, L_{group}, and the accumulated table, Tbl_{group} to its members. Based on the list and table, each node i then calculates its priority point, $\delta(i)$:

$$\delta(i) = \frac{i}{n} Tbl_{group}[|Tbl_{group}| - 1]$$

Then the node checks that which interval of the accumulated table that its priority point belongs to. Based on that, the anchor group of that node is found as below. Each node thus sends the information to its anchor clusters.

$$anchor\_group(i) = L_{group}[m], with\ Tbl_{group}[m-1] < \delta(i) \leq Tbl_{group}[m]$$

The second phase occurs during the time interval. When a source wants to query the information of a destination, the source also calculates the destination's priority point. Then, the source uses the cluster group list and accumulated table to look for the destination's anchor group. Finally, the source sends the QUERY message to the nearest destination's anchor cluster and obtains the result by the REPLY message. By using this mechanism, we can store an amount of information to a cluster group proportional to that group priority. The higher the group priority, the lower querying delay time. As a result, the average querying delay time is reduced. We present an illustration for the cluster topology in Fig. 1. Following all above steps, we can get $Pr_{cluster}$, L_{group}, Pr_{group}, Tbl_{group} as in Table 1. The L_{group} is obtained by the group selection algorithm which is shown later. The parameters n, S_{AG}, γ_0, γ_1 are set to $70, 2, 2, 1$ respectively in this illustration. Suppose that we wants to get the anchor group of node 12. Its priority point is $\delta(12) = \frac{12}{70}394 = 67.54$ ($\delta(12) \in (63; 95] = (Tbl_{group}[1], Tbl_{group}[2])$). So, the anchor group of node 12 is the group of clusters {5,56}.

2.4 Multiple Anchor Cluster Distribution

Previous home region based schemes uses only single binding between a node and a home region [4,2,3]. In order to reduce the querying delay time, we bind multiple anchor clusters to one node. However, it is not that the querying latency decreases if we just increases the number of bindings carelessly. For example, in

Table 1. Illustration of Size-based Hash Function

$L_{cluster}$	$Pr_{cluster}$	L_{group}	Pr_{group}	Tbl_{group}	$L_{cluster}$	$Pr_{cluster}$	L_{group}	Pr_{group}	Tbl_{group}
1	31	{1,14}	35.5	35.5	25	50	{25,14}	45	266.5
2	17	{2,37}	27.5	63	28	51	{28,5}	41	307.5
5	31	{5,56}	32	95	37	38	{37,19}	31.5	339
9	37	{9,25}	43.5	138.5	51	20	{51,2}	18.5	357.5
14	40	{14,25}	45	183.5	56	33	{56,14}	36.5	394
19	25	{19,28}	38	221.5					

the case that all anchor clusters of one node gather in one small area of one side, far away nodes from the opposite side cannot contact to any near anchor cluster and have to communicate with the anchor cluster by a long distance. So, in this section, we present the multiple anchor cluster distribution algorithm which aims to select a number of evenly scattered clusters of one group. Our scheme does not rely on GPS device and can work with any shape of cluster topology.

The algorithm takes a cluster j as an input then returns a cluster group G. The number of clusters in one group is predefined by parameter S_{AG}. Notice that by the mechanism of 2.1, each cluster-head can have the cluster topology which includes IDs of available clusters $L_{cluster}$, cluster size, and logical links between two adjacent clusters. There is another parameter, $r_{reverse}$ $(0 < r_{reverse} < 1)$, which defines the ratio of reverse length in this algorithm. The distance between two clusters is measured by cluster-hops in the following procedure.

In the pseudocode, the algorithm puts the input cluster j into the result group G initially. It then checks the condition of number of clusters, S_{AG}, on group G. If the condition is satisfied, the algorithm terminates with the result G in line 2. Otherwise, it runs through a loop until the number of clusters in group G is equal to S_{AG}. In each iteration, the distance for each cluster in the networks to the nearest cluster in group G is obtained. Then based on the distance calculation, the procedure selects list of clusters, L_1, which are the furthest away from clusters in group G. Among clusters in L_1, the algorithm chooses an unique cluster x for the case of cluster j. In lines 5 and 6, we use modulo operation to resolve the uniqueness globally. The distance between cluster x and cluster group G is $dist_{max}$. The reverse distance $dist_{reverse}$ is then obtained by multiplying $dist_{max}$ with the predefined parameter $r_{reverse}$. Next, group L_2 of clusters which are $dist_{reverse}$ cluster-hops away from cluster x are chosen in line 8. Once again, we resolve to the global unique cluster y of list L_2 by modulo operation as in lines 9 and 10. Cluster y is then added to group G and the loop is restarted.

We show an illustration of this algorithm by Fig. 2. The shape of the networks is drawn by the contour. First, in Fig. 2(a), the cluster j is put into the group G. Second, Fig. 2(b) shows three possible clusters which are furthest away from cluster j. These three clusters are with the same distance to cluster j, the maximum distance $dist_{max}$. Next, the unique cluster x is selected in Fig. 2(c). There are two possible clusters which are $dist_{reverse}$ hops away from cluster x. Unique cluster y is then chosen in Fig. 2(d). The procedure continues as in illustration. Finally, we have a group of three clusters of group G by two iterations in

Algorithm 1. Multiple Anchor Cluster Distribution

Input: The initial cluster j
Output: The result group G of clusters
Parameters:
The number of clusters in result group, S_{AG}
The list of available clusters, $L_{cluster}$
The ratio of reverse length, $r_{reverse}$

1 $G := \{j\}$
2 **if** $|G| == S_{AG}$ **then** return G
3 $CH := measure\_distance(L_{cluster}, G)$
4 $(L_1, dist_{max}) := max\_distance(L_{cluster}, CH)$
5 $index_1 := j \ mod \ |L_1|$
6 $x := L_1[index_1]$
7 $dist_{reverse} := \lceil r_{reverse}.dist_{max} \rceil$
8 $L_2 := find\_cluster(L_{cluster}, x, dist_{reverse})$
9 $index_2 := j \ mod \ |L_2|$
10 $y := L_2[index_2]$
11 add cluster y to group G
12 goto 2

Fig. 2(f). As we can see, these three clusters are scattered quite evenly in the networks. For our lookup service, they act as anchor clusters for a subset of nodes. A querying source can obtain the information by contacting the nearest anchor cluster. As a result, the average querying delay time is reduced.

3 Performance Evaluation

In this section, we simulate, analyze the performance result of our proposed scheme, CHLS, with reactive lookup service, RLS with Expanding Ring Search [5,1], and KCLS [7]. The simulation time for each test is 15 minutes and the mean result is obtained by 500 running times. There are n nodes scattered in squared networks with node density $d = 8$, the average number of neighbors per node. The communication transmission range is $250m$. The total latency of sending a message from one mobile node to its neighbor follows exponential distribution with expectation $2ms$. The mobility model is random waypoint with node velocity from 0 to $10m/s$. We set the parameters of our scheme as follows: $\gamma_0 = 2;\ \gamma_1 = 1; r_{reverse} = 0.2;\ \Delta = 5\ minutes$. The maximum distance, k-hop value, between cluster-head and members is set to 3. For the clustering topology which is used as the backbone for both KCLS and CHLS, we summarize the number of clusters in Table 2. For the lookup information, we assume that each node has its own information and the source wants to query the information of specific destination. In this general lookup application, r_{update} is the percentage of nodes which update their own information in one second. With $n = 1000$ and $r_{update} = 0.5\%$, there are 5 nodes updating their information per second and each

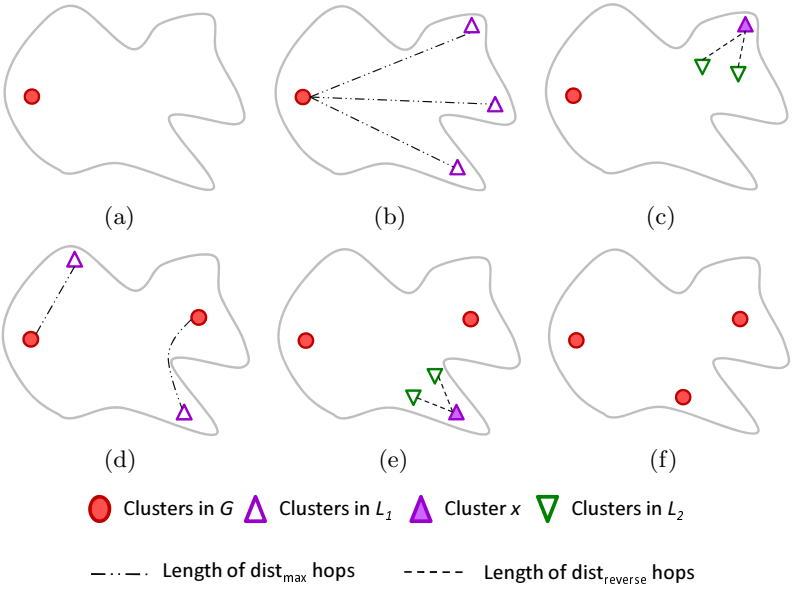


Fig. 2. Illustration of Multiple Anchor Cluster Distribution

node updates its information in 200 *seconds*. The rate r_{query} is the percentage of source nodes which start a new query in one second. In our scheme, CHLS-1 uses the uniform hash function while CHLS-2 is with size-based function. We run CHLS-2 with different values of number of anchor clusters per node, S_{AG}.

We show the query overhead in Fig. 3(a). All communication overhead metrics are measured for one node in one minute in average. As RLS is based on flooding, its query overhead is tremendous. KCLS has the smallest query overhead. However, in Fig. 3(b), the update overhead of KCLS is significant since each node needs to update the information to every available cluster-head. The total overhead is shown in in Fig. 3(c)-(e) and the querying delay time is in Fig. 3(f). We do not show the querying delay time of RLS because it is very high, about 5 times of CHLS-1' and it obstructs the observation of other schemes. When the number of node n, the query rate or the update rate increases, the total overheads of KCLS and RLS increase significantly. Hence, these two schemes are not scalable and not suitable with scenarios of high query rate and update rate. On the other hand, our scheme uses uses anchor cluster system and achieves the scalability requirement with small communication overhead.

Table 2. Number of Clusters

n	400	600	800	1000
Number of Clusters	26.2	38.8	51.3	69.7

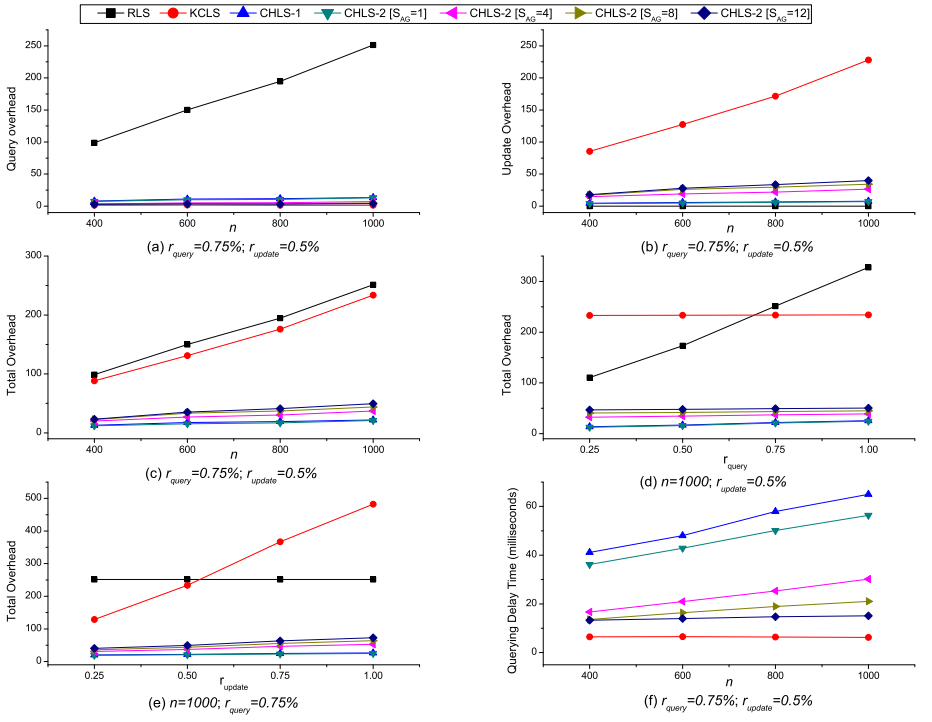


Fig. 3. Communication Overhead and Querying Delay Time

Compared with uniform hash function, size-based hash function is able to reduce about 10% of querying delay time, as in CHLS-2[$S_{AG} = 1$] and CHLS-1 of Fig. 3(f). KCLS has the smallest querying delay time as each node only needs to contacts its cluster-head for the desired information. KCLS' querying delay time is almost the lower bound for lookup service. So, in order to reduce the latency, we use multiple binding between one node and anchor clusters together with anchor cluster distribution algorithm. As the number of anchor clusters are scattered evenly in the networks, the source can get the desired information by contacting the nearest anchor cluster of the destination. Hence, the latency is small. For CHLS-2, when the number of anchor clusters per node S_{AG} increases, the query overhead decreases and the update overhead increases. More importantly, the querying delay time of CHLS-2 is reduced if S_{AG} increases. With $S_{AG} = 12$, CHLS-2's querying delay time is about twice of KCLS'. Moreover, the total overhead of CHLS-2[$S_{AG} = 12$] is about 1/5 of KCLS' in the case $n = 1000$. In general, depending on the scenario and parameters, CHLS-2' overhead may be only about 1/10 of KCLS'.

4 Conclusions

In this paper, we investigate the combination of global hashing lookup service and cluster structure. Compared with other home region based lookup services which depend on GPS device, our scheme is built on a multi-hop cluster topology. The clustering allows us to avoid geographic issues such as empty home region, predefined circle or rectangular area and costly, energy-consuming GPS device. According to our knowledge, there is still no research studying about the hashing-based lookup service in a self-adaptive self-organizing clustering topology. By using the dynamic hashing function, our scheme, CHLS, achieves much smaller communication overhead than RLS and KCLS. We also analyze the effect of cluster size to dynamic hash function and introduce an advanced size-based hash function. Then, the multiple anchor cluster distribution algorithm evenly scatters anchor clusters of one node in the networks. Consequently, the whole scheme successfully shortens the querying delay time significantly with extremely low communication overhead.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0024695).

References

1. Pham, N.D., Choo, H.: Energy Efficient Expanding Ring Search for Route Discovery in MANETs. In: Proceedings of the IEEE International Conference on Communications, pp. 3002–3006 (2008)
2. Erikson, J., Faloutsos, M., Krishnamurthy, S.V.: DART: Dynamic Address Routing for Scalable Ad Hoc and Mesh Networks. IEEE/ACM Transactions on Networking 15(1), 119–132 (2007)
3. Woo, S.C., Singh, S.: Scalable routing protocol for ad hoc networks. Wireless Networks 7, 513–529 (2001)
4. Cheng, H., Cao, J., Chen, H.H.: GRLS: Group-based Location Service in Mobile Ad hoc Networks. In: Proceedings of the IEEE International Conference on Communications, pp. 4734–4740 (2007)
5. Camp, T., Boleng, J., Wilcox, L.: Location Information Services in Mobile Ad hoc Networks. In: Proceedings of the IEEE International Conference on Communications, pp. 3318–3324 (2002)
6. Kasemann, M., et al.: A reactive location service for mobile ad hoc networks. NEC Network Laboratories NL-E Heidelberg. Tech. Rep. TR-02-014 (November 2002)

7. Leng, S., Zhang, L., Fu, H., Yang, J.: A Novel Location-Service Protocol Based on k-Hop Clustering for Mobile Ad Hoc Networks. IEEE Transactions on Vehicular Technology 56(2), 810–817 (2007)
8. Yu, J.Y., Chong, P.H.J.: A survey of Clustering schemes for Mobile Ad hoc Networks. IEEE Communication Surveys and Tutorials 7(1), 32–48 (2005)
9. Yu, J.Y., Chong, P.H.J.: 3hBAC(3-hop between adjacent cluster-heads): A novel Non-overlapping Clustering Algorithm for Mobile Ad hoc networks. In: IEEE Pacific Rim Conference on Communication, Computers and Signal Processing, August 2003, pp. 318–321 (2003)

Performance Analysis of MIMO System Utilizing the Detection Algorithm

Sungsoo Ahn and Dongyoung Lee

Dept. of Information and Communication, Myong-ji College
356-1 Hongeun3-Dong, Seodaemun-Gu, Seoul 120-776, Korea
{ssan,dylee}@mail.mjc.ac.kr

Abstract. This paper presents a performance analysis for MIMO system using detection algorithm based on zero-forcing and maximum likelihood. Proposed algorithm provides optimal gain on a real-time basis arbitrarily according to angle spread produced at every snapshot. From the result of performance analysis through various simulations in WiBro, it is confirmed that proposed method is far superior compare to without detection algorithm.

Keywords: MIMO, WiBro, Zero-forcing, Maximum likelihood, STC, SM.

1 Introduction

The main goal in next generation wireless communication system is to increase the data throughput and the user capacity. It has been confirmed that the performance is far superior when multiple antennas are applied at both transmitter and receiver side. MIMO(Multiple input Multiple output) techniques[1] can increase the performance or capacity according to input-output channel modeling implementation. Also MIMO system can be divided into two categories:STC(Space Time Coding) and SM(Spatial multiplexing)[2]. STC improves the performance of the communication system by coding over the different transmitter branches, whereas SM achieves a higher throughput by transmitting independently data streams on the different transmitter branches simultaneously and at the same carrier frequency. In case of SM method, it is divide the ML(Maximum Likelihood) and ZF(Zero Forcing) algorithm[3] for optimal detection.

This paper applied to the WiBro system based on MIMO for performance analysis. Also, in this paper, we present the results of channel estimation utilizing the LI(Linear Interpolation) in down link system.

This paper organized as follows. Section 2 contains concept of MIMO system. Section 3 describes the structure of WiBro system[6]. In Section 4, we present the performance of proposed method in terms of BER(Bit Error Rate). The conclusions from this research are outlined in Section 5.

2 MIMO System

MIMO system is to increase throughput utilizing multiple antenna[4] in TX side without frequency allocation or increasing transmitter power. MIMO system can be

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 248–257, 2011.

divided into STC and SM which are concern with trade off each other. Two methods can be select to the finally better one according to communication environment and user's requirement. In this section, let's consider concepts and characteristic of STC and SM[5].

2.1 STC(Space Time Coding)

STC technique obtains diversity gain that occurs by transmit encoding signal with space and time axis at transmit antenna. For example, we explain systems that have 2x1 and 2x2 TX antenna, i.e. 2 transmit antenna and 1 or 2 receive antenna.

2.2 2 x 1 STC

As shown in Fig.1, when two symbol signal s_1, s_2 are transmit at the transmit side, STC encoder make encoding during two symbol period T following as,

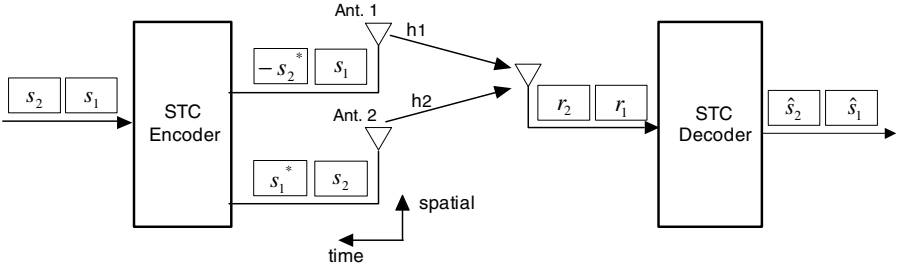


Fig. 1. 2 x 1 STC system block

	Antenna 1	Antenna 2
Time 0	s_1	s_2
Time T	$-s_2^*$	s_1^*

In 2x1 STC systems, throughput is 1 during 2 symbol period because of sending of 2 symbols. Signal r_1 and r_2 received to the first and second symbol duration are written as follows,

$$\begin{aligned} r_1 &= h_1 s_1 + h_2 s_2 + n_1 \\ r_2 &= -h_1 s_2^* + h_2 s_1^* + n_2 \end{aligned} \tag{1}$$

Where h_1, h_2 assume flat-fading channel which have fixed value of the two symbol period. n_1, n_2 denote additional white Gaussian noise.

Equation (1) can be express as a matrix form as below.

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \tag{2}$$

STC decoder detected $\hat{s}_1, \hat{s}_2$ from two of received signals r_1, r_2.

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2^* \end{bmatrix} = \begin{bmatrix} (|h_1|^2 + |h_2|^2) \cdot s_1 + \tilde{n}_1 \\ (|h_1|^2 + |h_2|^2) \cdot s_2 + \tilde{n}_2 \end{bmatrix} \tag{3}$$

As shown in Equation (3), 2x1 STC obtain diversity gain to the 2nd order.

2.2.1 2 x 2 STC

Fig. 2 shows 2 x 2 STC system block diagram. Since receiving antenna consists of two, it is perform STC decoding at the each receiving antenna. We can estimate transmit signal through that technique immediately. 2 x 2 STC decoding notation can be expressed as shown in Equation (4)

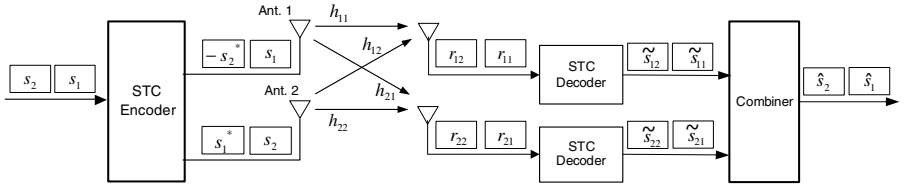


Fig. 2. 2 x 2 STC Transceiver structure

$$\begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix} + \begin{bmatrix} n_{11} & n_{21} \\ n_{12} & n_{22} \end{bmatrix}$$

$$= \begin{bmatrix} s_1 h_{11} + s_2 h_{12} + n_{11} & s_1 h_{21} + s_2 h_{22} + n_{21} \\ -s_2^* h_{11} + s_1^* h_{12} + n_{12} & -s_2^* h_{21} + s_1^* h_{22} + n_{22} \end{bmatrix} \tag{4}$$

Where r_{ij} is received signal, n_{ij} denotes additional white noise, i indicates receiving antenna index, and j is time index. It is assume that subscript i of channel h_{ij} denotes receiving antenna index, j is transmitting antenna index, and h_{ij} indicates flat fading channel.

Thus, signal of each antenna after decoding is expressed as below

$$\begin{bmatrix} \tilde{s}_{11} \\ \tilde{s}_{11} \end{bmatrix} = \begin{bmatrix} h_{11}^{*} & h_{12} \\ h_{12}^{*} & -h_{11} \end{bmatrix} \begin{bmatrix} r_{11} \\ r_{12}^{*} \end{bmatrix} = \begin{bmatrix} (|h_{11}|^2 + |h_{12}|^2) \cdot s_1 + \tilde{n}_{11} \\ (|h_{11}|^2 + |h_{12}|^2) \cdot s_2 + \tilde{n}_{12} \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} \tilde{s}_{21} \\ \tilde{s}_{22} \end{bmatrix} = \begin{bmatrix} h_{21}^{*} & h_{22} \\ h_{22}^{*} & -h_{21} \end{bmatrix} \begin{bmatrix} r_{21} \\ r_{22}^{*} \end{bmatrix} = \begin{bmatrix} (|h_{21}|^2 + |h_{22}|^2) \cdot s_1 + \tilde{n}_{21} \\ (|h_{21}|^2 + |h_{22}|^2) \cdot s_2 + \tilde{n}_{22} \end{bmatrix} \tag{6}$$

Where $\tilde{s}_{ij}$ is signal of each antenna after decoding, $\tilde{n}_{ij}$ denotes additional white noise, and subscript i and j are receiving and transmitting antenna index, respectively.

Accordingly, it can estimate receiving signal theoretically from signal combination within same symbol period at each antenna.

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} \tilde{s}_{11} \\ \tilde{s}_{12} \end{bmatrix} + \begin{bmatrix} \tilde{s}_{21} \\ \tilde{s}_{22} \end{bmatrix} = \begin{bmatrix} (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2) \cdot s_1 + \tilde{n}_{11} + \tilde{n}_{21} \\ (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2) \cdot s_2 + \tilde{n}_{12} + \tilde{n}_{22} \end{bmatrix} \tag{7}$$

From Equation (7), 2 x 2 STC is obtained diversity gain to the 4th order. Also, it has achieve signal to noise ration of 3dB additionally because of two receiving antenna.

2 x 1, 2 x 2 STC system shows excellent performance cause of diversity gain according to number of antenna in RF communication. But, it has been shortage that data throughput is almost same as one of single antenna even through increasing number of antenna.

2.3 SM(Spatial Multiplexing)

In this subsection, we illustrate SM technique. SM has a function that is increase the system throughput by sending data to the each different type at transmit side, in which no exist additional bandwidth.

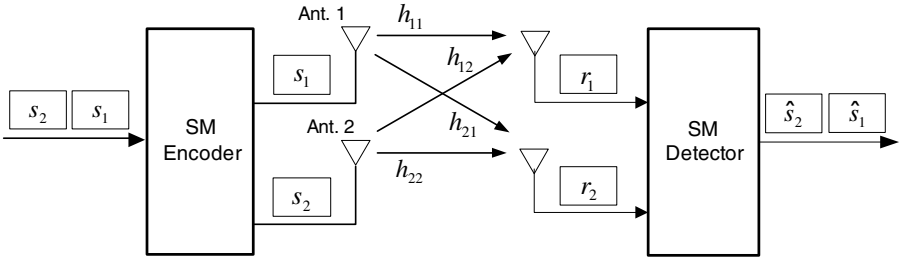


Fig. 3. 2 x 2 SM Transceiver structure

As shown in fig. 3, symbol to input to the encoder as a serial type change the parallel type through SM encoding, is send to transmit antenna. SM detector of receive side is estimate the transmit signal from received one utilizing adopted algorithm. Received signal is following as.

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \tag{8}$$

From Equation (8), it can be express as a matrix form below as.

$$\underline{r} = \underline{\underline{H}}\,\underline{s} + \underline{n} \tag{9}$$

Where $\underline{r}$, $\underline{\underline{H}}$, $\underline{s}$, $\underline{n}$ denote receiving signal vector, channel matrix, transmitting signal vector, and additional white Gaussian noise vector, respectively. In this paper, we illustrate on ML(Maximum Likelihood), ZF(Zero Forcing) that are detecting technique of transmit signal.

2.3.1 ML(Maximum Likelihood) Detection

In this subsection, we illustrate ML detection algorithm. Criterion for ML detection algorithm is written by below.

$$\hat{\underline{s}} = \arg\min_{\underline{s} \in S} \left\| \underline{r} - \underline{\underline{H}}\,\underline{s} \right\|^2 \tag{10}$$

In Equation (10), $\hat{\underline{s}}$ is estimated signal vector, S denotes set of signal to enable all sending, $\underline{r}$ indicates received signal, and $\underline{\underline{H}}$ is channel matrix. ML algorithm is estimate to the largest correlation value between received signal and transmitted signal which has optimal performance. But, there is some too much calculation amount for searching the largest correlation value. When modulation order is M, it means that number of operation is M^{N_t} (N_t is number of transmit antenna). For example, in case of two transmit antenna, basic operations are different from modulation type, i.e., QPSP modulation($M = 4$) required 16 operation and 16QAM has 256 operation. From the this result, it is confirm that operation amount is increased exponentially when modulation order has larger gradually

2.3.2 ZF(Zero Forcing) Detection

ZF is algorithm which is detect transmit signal by multiplying the inverse matrix of channel matrix. Detection method of ZF algorithm is following as

$$\underline{r} = \underline{\underline{H}}\,\underline{s} + \underline{n} \tag{11}$$

$$\hat{\underline{s}} = \underline{\underline{H}}^{+}\,\underline{r} \tag{12}$$

$$\underline{\underline{H}}^{+} = \left(\underline{\underline{H}}^{H}\underline{\underline{H}}\right)^{-1}\underline{\underline{H}}^{H} \tag{13}$$

Where $\underline{\underline{H}}^{H}$ indicates complex conjugate transpose of channel matrix, $\underline{\underline{H}}^{+}$ denotes pseudo-inverse. Therefore estimated signal is below as

$$\underline{\hat{s}} = \left(\underline{\underline{H}}^{H}\underline{\underline{H}}\right)^{-1}\underline{\underline{H}}^{H}\left(\underline{\underline{H}}\,\underline{s} + \underline{n}\right) = \underline{s} + \underline{\underline{H}}^{+}\underline{n} \tag{14}$$

Operation amount of ZF algorithm is small, whereas system performance is severely degraded because of noise amplification caused $\underline{\underline{H}}^{+}$ matrix.

3 WiBro Standards System

3.1 MIMO Encoding in PUSC(Partial Usage of Subchannels)

In this section, we propose the MIMO encoding of IEEE802.16e specification. This paper use to 2 X 2 MIMO system, it is possible to both STC and SM technique when transmit side adapt two antenna. As it is mention, space-time rate of STC is 1, SM has 2. We refer to matrix A, matrix B in WiBro system.

Matrix A and Matrix B can be defined as follow

$$A = \begin{bmatrix} s_1 & -s_2^{*} \\ s_2 & s_1^{*} \end{bmatrix}, \quad B = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

3.2 MINO Encoding in Downlink PUSC

Downlink MIMO encoding is allocated the pilot subcarrier primary in PUSC cluster, and set to data subcarrier in next stage.

3.2.1 Pilot Allocation

Pilot allocation method have to modify as fig. 4 in which pilot use two antenna which have same estimation capability. As shown in figure 3-1, pilot position change the 4 symbol interval periods.

As see figure 3-1, we known that transmitting pilot signal at each antenna send with no overlapping such that channel estimation can easy from that structure. However, pilot power have to transmit in higher 3dB than single antenna one because number of pilot within one cluster is decrease to the half rate.

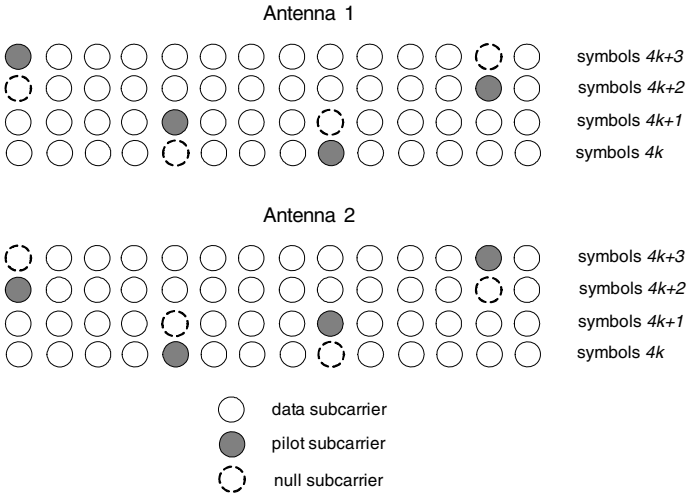


Fig. 4. Cluster structure of MIMO PUSC in two transmit antenna

4 Performance Analysis

This section describes the simulation results obtained using the proposed method of various algorithms. This paper shows the performance of STC, SM, and several combination algorithms. First of all, simulation parameter shows table 1 as

Table 1. STC simulation parameter

parameter	value
channel	Rayleigh fading
Doppler frequency	128 Hz
modulation	QPSK
Number of transmit antenna	1 or 2
Number of receive antenna	1 or 2

Fig. 5 shows the performance of uncoded BER(Bit Error Rate) to the SNR(Signal to Noise Ration) in STC system. In case of 10^{-3} BER, 2 x 1STC is better 10dB diversity gain than SISO method. When we consider 2 x 2 STC, it shows much better 17dB than SISO in 10^{-3} caused by affect to 3dB SNR gain additionally.

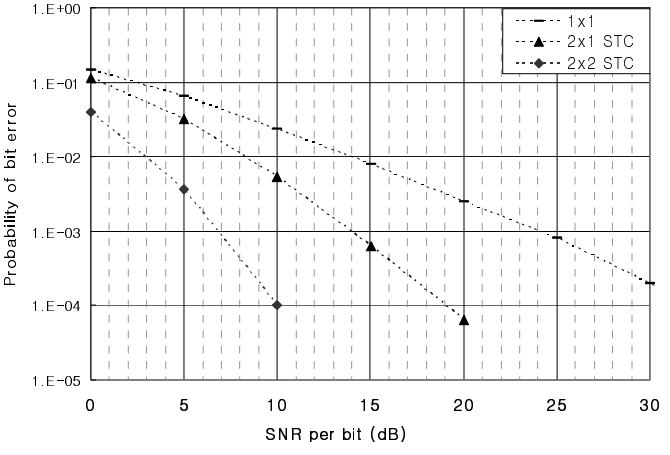


Fig. 5. Performance of STC system (Uncoded BER)

Table 2. 2 x 2 SM simulator parameter

parameter	value
channel	Rayleigh fading
Doppler frequency	128 Hz
modulation	QPSK
Number of transmit antenna	2
Number of receive antenna	2

Fig. 6 shows the performance of uncoded BER in SM system. Performance of ZF method is degraded 2dB compare to SISO in 10^{-3} BER. While throughput increased twice at same condition. However, we can confirm that performance of ML is far superior 8.5dB than SISO in term of BER.

As shown in Fig. 7, performance of 2 x 2 STC is best one among 4 algorithm. When it is focused on SM algorithm, 2 x 2 ML is more outperform 2 x 2 ZF method.

Fig. 8 shows the performance of WiBro MIMO system in coded BER. As shown in fig. 8, the improvement provided by the proposed system is comparable that of provided fig. 7. But, we can observe that fig. 8 maintains much better performance 3.5~4dB than fig. 7 due to coding gain.

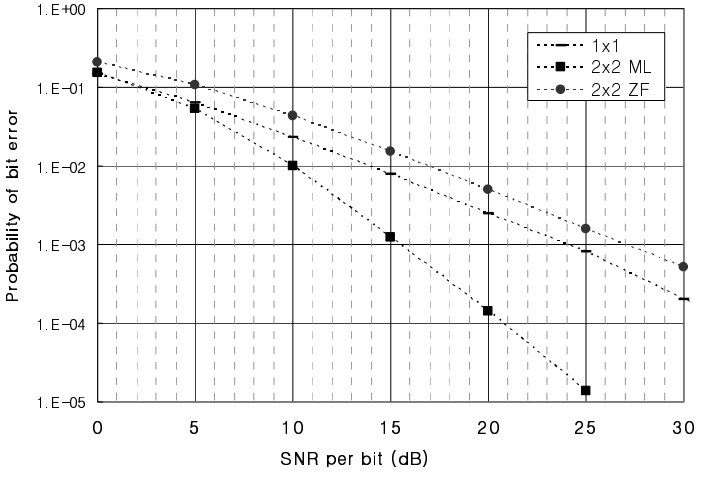


Fig. 6. 2 x 2 Performance of SM system (Uncoded BER)

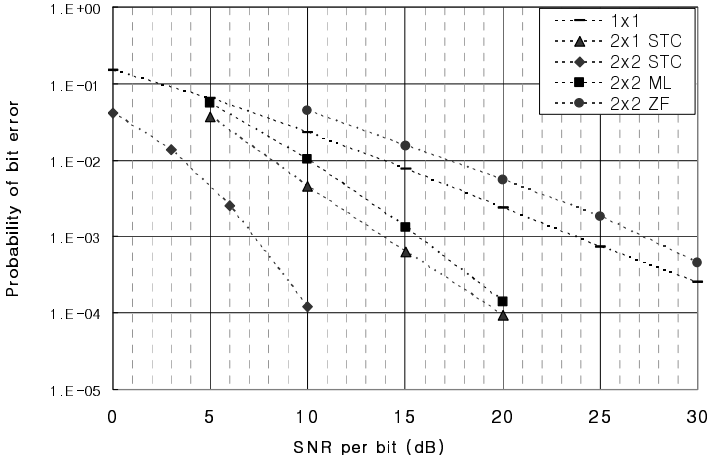


Fig. 7. Performance of WiBro system(Uncoded BER)

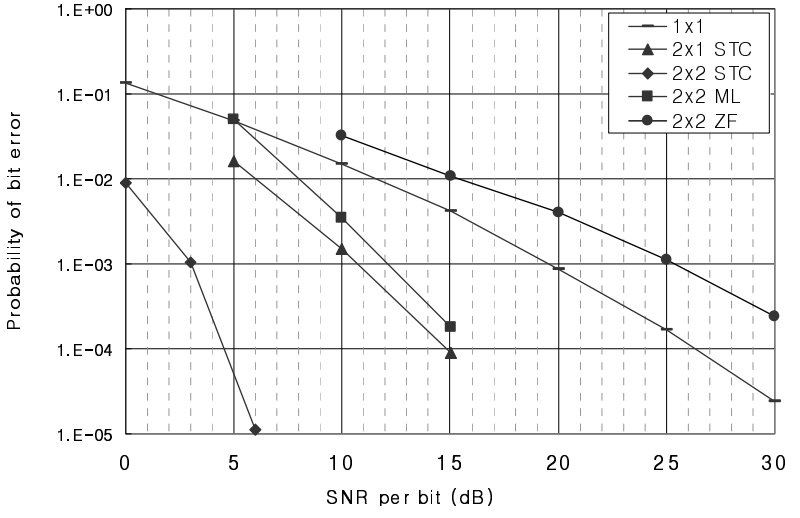


Fig. 8. Performance of WiBro system (Coded BER)

5 Conclusions

In this paper, we present a result of various simulation for improving the BER in WiBro MIMO system. From the simulation results, it is confirm that STC method outperforms the SM method include the ML and ZF algorithm when using the MIMO transceiver antenna in WiBro environments. Based on the analysis from various simulation, it is conclude that the STC technique is suitable for the MIMO operating system in practical WiBro environments.

References

1. Gesbert, D., Kountouris, M., Heath Jr., R.W., Chae, C.-B., Salzer, T.: From single user to multiuser communications; Shifting the MIMO paradigm. IEEE Signal Process. Mag (2007) (in press)
2. Knopp, R., Humblet, P.A.: Information capacity and power control in single-cell multiuser communications. In: Proc. IEEE Int. Conf. Commun., June 1995, vol. 1, pp. 331–335 (1995)
3. Yoo, T., Goldsmith, A.: On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming. IEEE J. Sel. Areas Commun. 24, 528–542 (2006)
4. Viswanathan, H., Venkatesan, S., Huang, H.: Downlink capacity evaluation of cellular networks with known-interference cancellation. IEEE J. Sel. Areas Commun. 21(6), 802–811 (2003)
5. Sharif, M., Hassibi, B.: On the capacity of MIMO broadcast channels with partial side information. IEEE Trans. Inf. Theory 51, 506–522 (2005)
6. Alex, S.P.L., Jalloul, M.A.: Performance evaluation of mimo in IEEE802 16e/WiMAX. IEEE Journal on Selected Topic in Signal Processing 2(2), 181–190 (2008)

Overview and Comparison of Global Concentrating Solar Power Incentives Schemes by Means of Computational Models

M. Villarini[1], M. Limiti[1], and R. Impero Abenavoli[2]

[1] CIRPS Sapienza Università di Roma, piazza San Pietro in Vincoli, 10 - 00184 Rome Italy
[2] DIMA Sapienza Università di Roma, via Eudossiana, 18 - 00184 Rome Italy

Abstract. The present paper gives an updated picture of concentrating solar power incentives schemes in the World. It resumes the main figures of the incentives schemes per Country such as plant size limit, feed-in tariff value, duration, inflation recovery and the eligibility to hybrid plants.

It also gives a comparison of the incentives schemes in terms of revenues and profitability. The investment revenues have been calculated considering the values of the incentives and local radiation. The profitability has been computed, as internal rate of returns of the project (IRRs) considering the incentive value, the solar radiation and the incentive length per each Country and a fixed value of capital expenditure equal for all Countries.

It shows that South Africa and Spain grant the most profitable incentive schemes with IRRs[1]of 28% and 25% respectively and that USA grant the lowest feed-in tariff, as benefit the highest solar radiation, and thus show the lowest profitability.

Keywords: Concentrating solar power, Feed-in-tariff.

1 Introduction

Solar incentives schemes are incentives offered to electricity producers to install and operate solar thermal power plants given by a government in order to encourage the industry to achieve the economies of scale necessary to compete where the cost of renewable-generated electricity is above the cost from the existing grid. Such policies are implemented to promote national or territorial energy independence, high tech job creation and reduction of carbon dioxide emissions which cause global warming.

The most common mechanisms are feed-in tariffs which reward the number of kWh produced over a certain period of time. The price paid per kWh under a feed-in tariff exceeds the price of grid electricity and is calculated as a single value consisting of the following components:

[1] IRR values are calculated using ebitda and are thus overestimated as pretax. They are not to be taken as a reference as absolute values but rather as terms of comparison.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 258–269, 2011.

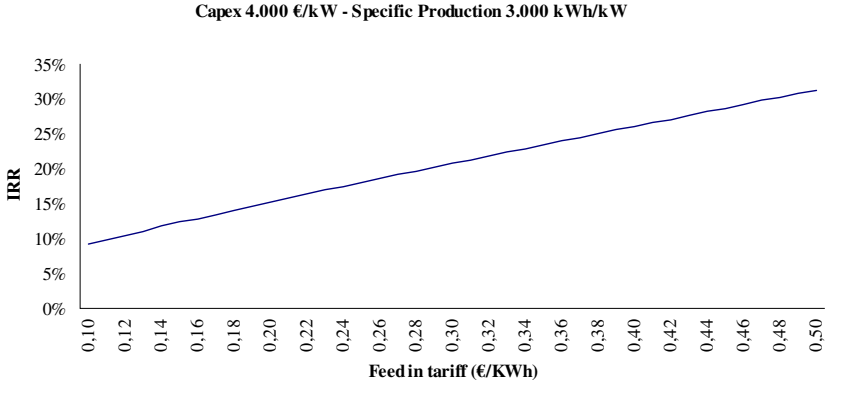
- Return on equity
- Interest on loan capital
- Depreciation
- Interest on working capital
- Operation and maintenance expenses

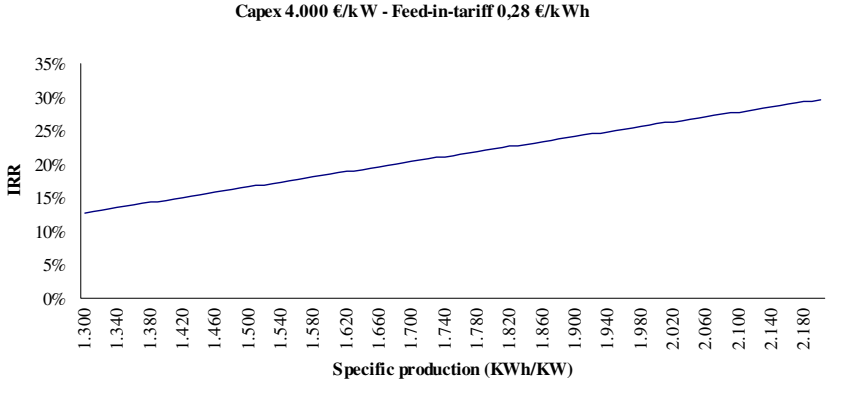
This is done in order to guarantee the electricity producers at least the same profitability of a traditional power plant despite the higher capital expenditure and the higher risks of unproven technologies. However, when in a given country or territory the cost of solar electricity falls to meet the rising cost of grid electricity, then so called grid parity is reached, and in principle incentives are no longer needed.

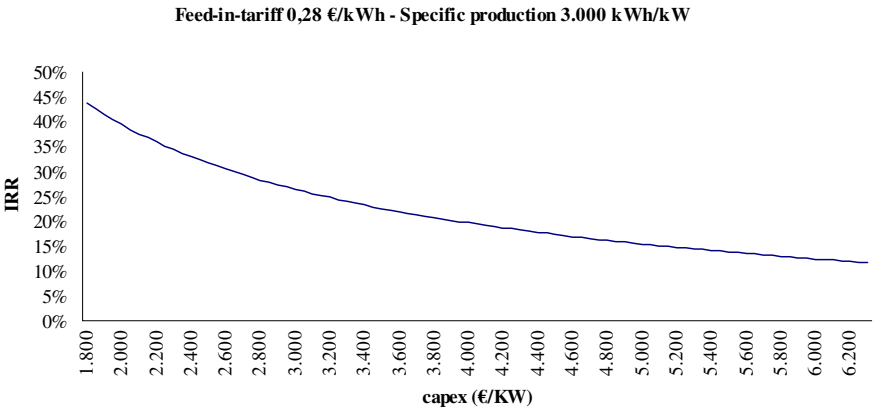
The profitability of the plant depends mainly on the following drivers:

- Feed-in tariffs incentive schemes
- Specific production
- Capital expenditure

In the following figures the IRR of a 10 MW CSP plant, as an index of profitability, is varied with the feed-in tariff, the specific production and the capital expenditure according to the Italian Decree.

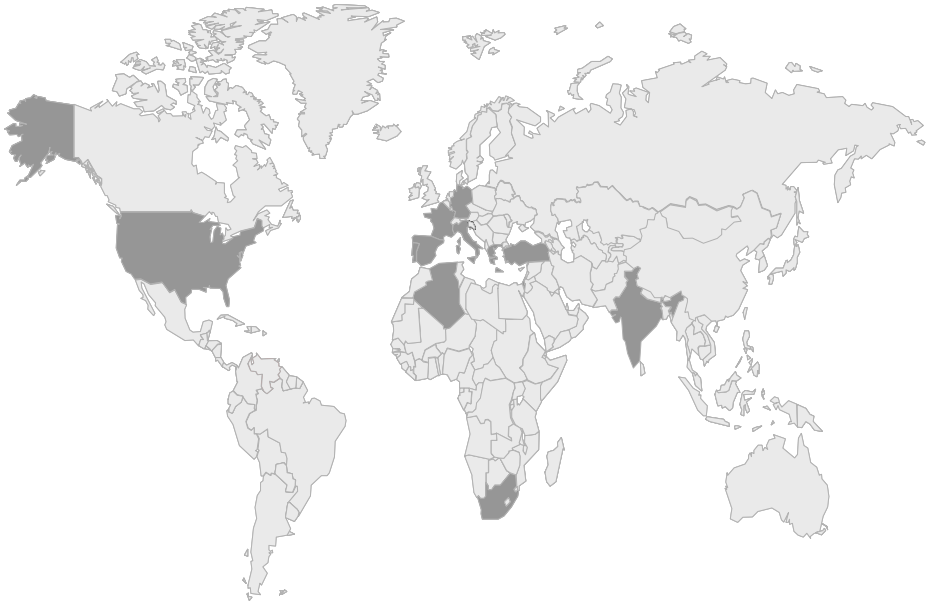






Incentives schemes worldwide

The following figure shows the Countries that have introduced feed-in tariffs to encourage CSP system:



Incentive schemes are typically characterized by a maximum power level, a feed-in tariff value, the length of the incentive, the inflation recovery and the eligibility to hybrid plants.

In the following table are listed the main characteristics of the incentive schemes worldwide.

Country	Power	Tariff	Period	Inflation recovery	Hybrid plant
Algeria	Up to 50MW	0-300% market price	All life	No	Yes
France	Up to 250kW	0,314-0,40 €/kWh	20	Yes	No
	Over 250kW	0,314*R-0,40 €/kWh	20	Yes	No
Germany	-	0,32 €/kWh	All life	Digression 9%	No
Greece	Up a 5MW	0,23-0,25 €/kWh	10+10	No	Max 10%
	Over 5MW	0,25-0,27 €/kWh	10+10	No	Max 10%
India	-	15,04 Rs/kWh	25	No	No
Israel	Up to 20MW	0,20 $/kWh	20	Yes	Max 30%
	Over 20MW	0,16 $/kWh	20	Yes	Max 30%
Italy	-	0,28 €/kWh (solar fraction>85%)	25	No	Yes
		0,25 €/kWh (solar fraction between 50% and 85%)	25	No	Yes
		0,22 €/kWh (solar fraction<50%)	25	No	Yes
Portugal	Up a 10MW	Calculated	15	No	No
	Over 10MW	Calculated	15	No	No
Spain	Up to 50MW	0,27 €/kWh	25+	Yes	Max 15%
South Africa	Over 1MW	3,14 R/kWh (Parabolic Trough)	20	No	Max 15%
		2,31 R/kWh (power tower)	20	No	Max 15%
Turkey	N/A	9.13 YKr/kWh	10	No	N/A

2 Middle East

Israel

In 2002, the Israeli Ministry of National Infrastructures, which is responsible for the energy sector, made CSP a strategic component of the electricity market and introduced since September 2006 feed-in tariffs for solar IPPs effective for 20 years.

In particular in September 2006 the Israeli Ministry of Infrastructure decided to promote the installation of CSP plant introducing a feed-in tariff of 0,8760 NIS/kWh (about 20,4 c$/kWh) for plants with a capacity from 100kW up to 20MW the fare and of 0,7005 NIS/kWh (about 16,3 c$/kWh) for plants with a capacity greater than 20MW. Fossil fuels maximum contribution allowed is 30% of the total energy produced by the plant.

Turkey

Turkey enacted its first specific Renewable Energy Law in May 2005 called “Law on Utilisation of Renewable Energy Sources for the Purpose of Generating Electrical Energy”.

The Renewable Energy Law works in line with “Renewable Energy Source Certificates” (RES Certificate). The law introduced fixed tariffs for electricity generated out of renewable energy sources and a purchase obligation for the distribution companies holding retail licenses from the certified renewable energy producers. The price of electricity bought in accordance with this provision is determined by the Energy Market Regulatory Authority (EMRA). The initial amount was 9,13 YKr/kWh in 2007, (approximately 5,2 c€/kWh) for the first 10 years of operation for a renewable energy generation facility.

Currently there are amendments being made to the RES law. The Draft Law for RES includes a feed-in tariff for CSP of 24 c€/kWh for the first 10 years, dropping to 20 c€/kWh for the next 10 years. Legislations are also discussing an additional tariff for the first five years if at least 40% of the equipment is manufactured in Turkey. There may be further changes to the draft law and the final outcome in Turkey, by the time this report is printed.

Africa

Algeria

Algeria is the first country among the Organization for Economic Cooperation and Development (OECD) to introduce a premium feed-in tariff for thermodynamic systems, publishing in 2004 the "Decree Executif 04-92" listed in the Official Journal of Algeria No.19 in addition to the law n ° 02-01 of 22 Dhou El Kaada 1422 corresponding to February 5, 2002 on electricity and gas distribution.

The Decree establishes a premium on the price of electricity produced, sold or consumed by thermodynamic systems up to 50MW, which varies depending on the solar share as follows:

- Solar share equal to 100% (non-hybrid plants) the premium on the price per kWh is 300%;
- Solar share greater than or equal to 25% premium on the kWh price is 200%;
- Solar share between 20% -25% premium on the kWh price is 180%;

- Solar share between 15% -20% premium on the kWh price is 160%;
- Solar share between 10% -15% premium on the kWh price is 140%;
- Solar share between 5% -10% premium on the kWh price is 100%;
- Solar share less than 5% premium on the price per kWh is equal to zero.

In the case of plants that produce less than 20% of energy used, the price premium is reduced as follows:

- usable output between 15% -19% of the premium is 120%;
- usable output between 10% -15% of the premium is 80%;
- usable output below 10% the premium is zero.

South Africa

On 26 March 2009, the National Energy Regulator of South Africa (NERSA) has approved a 20 year feed-in tariff scheme for renewable energy systems of 2,1 R/kWh for solar-thermodynamic. The rate is annually reviewed for the first five years and every three years by the fifth year onwards.

In October 2009 NERSA reviewed the tariffs with the REFITs Phase II covering several technologies, including parabolic trough without storage and tower systems with storage.

The approved Renewable Energy Feed-in Tariffs (REFITs) Phase II are:

trough without storage: 3,14 R/kWh
tower with storage of 6 hrs per day: 2,31 R/kWh.

NERSA also approved that fossil fuel allowed for the CSP technology is limited to a maximum of fifteen percent (15%) of the total primary energy input and that the power plant must be greater than or equal to 1 MW.

Asia

India

In 2010 India launched the Jawaharlal Nehru Solar Mission in order to achieve 20 GW of solar power capacity and achieve grid parity in pricing by 2022. The 25 years incentive tariffs are notified annually by the Central Electricity Regulatory Commission (CERC). For 2009-10, tariff in the case of solar thermal was 13,45 Rs/kWh. For 2010-11 CERC has increased tariffs to 15,04 Rs/kWh for solar thermal plants.

Europe

Spain

The first step of Spain to the promotion of renewable energy and in particular to the CSP started in 1998 by Royal Decree 2818/1998 which promoted the development of solar technologies (non-hybrid) with an incentive of 30 pesetas/kWh for systems of size up to 5kW up to a total installed power of 50MW.

The subsequent Royal Decree 841/2002 established the distinction in terms of tariff between systems using photovoltaics as a primary source and the systems that use solar thermal, introducing them at a price of 0,120202 €/kWh (20 pesetas/kWh).

On 27 March 2004 the Royal Decree 436/2004 increased the incentive rate for plants that supply power to distribution companies, as 300% of the reference price for the first 25 years, indexed to the fixed price of gas in 2004 equal to 7,2072 €c/kWh, and as 240% of the reference price after the first 25 years. For plants that sell electricity in the electricity market, the Decree provided a premium of 250% of the market price for the first 25 years and of 200% plus an incentive of 10% after the first 25 years.

In addition, in cases the system sells electricity to electricity distribution companies at a fixed price, the Decree limited the use of gas for the purpose of keeping the storage temperature at less than 12% of electricity produced and only during periods of interruption of electricity generation.

Said percentage may reach 15%, without time restrictions of use, if the installation sells its electricity freely in the market.

On 25 May 2007, the Royal Decree 661/2007 introduced a new incentive arrangements under which the investor may decide to sell electricity at a fixed price, the same for all the exercises, or alternatively to sell the electricity directly into the electricity market daily, term or a bilateral contract, receiving in this case a negotiated price plus a premium.

In the latter case, the Royal Decree 661/2007 introduced for some systems, such as thermodynamic systems, an upper and a lower limit for the sum of the hourly rate of the daily market with a premium reference, that the premium receivable per hour may be reduced according to the limits. This new system allows the investor protection in the event that market prices were too low and eliminates the premium when the market price is high enough to ensure recovery of costs. Another innovation introduced by the Decree is the annual adjustment of the tariff on the basis of inflation.

The Decree provides for the thermodynamic a fixed price of 26,9375 c€/kWh for the first 25 years and 21,5498 c€/kWh for the years to follow, a premium benchmark of 25,4000 c€/kWh the first 25 years and 20,3200 c€/kWh for following years, and the lower and upper limits respectively of 25,4038 c€/kWh and 34,3976 c€/kWh.

France

On 26 July 2006 the French government published in the Journal Officiel de la République Française n° 171, a feed-in tariff for solar systems with a capacity up to 12MW and specific production less than 1500 hours/year, equivalent to 30 c€/kWh for installations in France and 40 c€/kWh for installations IN overseas (eg Corsica). For systems with more than 12MW capacity and specific production greater than 1500 hours/year the rate was 5 c€/kWh. The legislation provided an annual adjustment of the scale with the index of the consumption and wage growth.

On 14 January 2010 the government introduced an amendment to the tariff in the Journal Officiel de la République Français n° 0011, according to which the rate is equal to (31,4 c€/kWh)*R for installations in France, where R is a coefficient equal to 1 for power plant less than 250 kW and greater than 1 and variable depending on the location of the plant for power plant over 250 kW. For installations in overseas (eg Corsica) the rate is unchanged at 40 c€/kWh.

Germany

The first steps in Germany towards the development and promotion of renewable energy began with the approval of the Renewable Energy Sources (RES) Act by the German Government 17 March 2000. Compared to the previous Electricity Feed Act, which covered only the photovoltaic industry, the text extended the subject to CSP systems introducing a minimum rate of 99 pfennigs/kWh for 20 years. On 21 July 2004 the Government approved a revision of the Renewable Energy Sources Act, which updated the minimum rate to 45,7 c€/kWh. In 2008 a further update of the Renewable Energy Sources Act, as well as updating the tariff at 31,94 c€/kWh, introduced a reduction of the tariff of 9% per year.

Greece

On 27 June 2006 the Greek Government approved the Law 3468/2006 establishing a 10 year tariff, with the possibility to obtian an extension for further 10 years, depending on the installed capacity. For systems with power capacity less than or equal to 5MW the rate is 25 c€/kWh if connected to the grid, and equal to 27 c€/kWh if installed on islands and not connected to the grid; for systems with power capacity greater than 5MW the tariff is equal to 23 c€/kWh if connected to the grid and of 25 c€/kWh if installed on islands and not connected to the grid.

The law also provides the application of tariffs to solar thermodynamic hybrid plant where energy from conventional sources does not exceed 10% of the total energy produced annually.

Italy

In Italy the Decree of Ministry of Economical Development 11/4/2008, published in May 2008, establishes criteria to promote the production of electric power from the solar source by means of thermodynamic cycles. Incentives are recognized to solar electricity produced by solar thermal power plants or by hybrid solar power plants without any limitation in solar share (solar electricity share or solar capacity share) for 25 years, as a fixed value without any inflation adjustment and can be accumulated with the selling price on the grid. The incentive values vary from 0,22 to 0,28 €/kWh depending on the value of the solar share.

Portugal

On 31 May 2007 Portugal introduced, by Decree-Law 225/2007, a feed-in tariff valid 15 years according to the following formula:

$$VRDm = \{KMHO\ m \times [PF(VRD)m + PV(VRD)m] + PA(VRD)\ m \times Z\} \times [IPCm\text{-}1/IPCref] \times [1/(1\text{-}LEV)]$$

where

VRDm represents the remuneration system in the month m;

KMHOm is a factor that modulates the values of PF(VRD)m, PV(VRD)m e PA(VRD)m as a function of time of day electricity supply;

PF(VRD)m is the fixed part of the price in month m;

PV(VRD)m is the variable part of the tariff in month m;

PA(VRD)m is the environmental part of the tariff in month m;

IPCm-1 is the consumption index in the month m-1;

Z is the dimensionless coefficient on the technology used;

IPCref is the consumption index in the month preceding the start of the grid;
LEV is the losses avoided in the transmission network and distribution.

America

USA

The U.S. legislation has both federal incentives and ad hoc solutions prepared by individual Member

Federal States

The Energy Policy Act of 1992 introduced the federal REPI (Renewable Energy Production Incentive) in order to establish incentives for renewable energy systems including thermodynamics solar power systems. The REPI provided an incentive of 1,5 c$/kWh with a length of 10 years.

Washington

In 2005 the State of Washington introduced, with the Senate Bill 5110, a basic feed-in tariff of 15 c$/ kWh multiplied by a coefficient that varied depending on the technology and origin of the components of the system. The subsequent Senate Bill 6170 and Senate Bill 6658 extended the application of legislation at a plant owned by local Government.

California

In September 2006, the state of California has introduced Assembly Bill 1969 which required each utility to establish, in agreement with the California Public Utilities Commission (CPUC), tariff for electricity produced from renewable sources by water supply companies. Subsequent provisions introduced in 2007 with Decision 07-07-027 in 2008 with Senate Bill 380 and in 2009 with Senate Bill 32 extended the application of tariffs also other companies

Hawai

In October 2008 the State of Hawaii has released the 2008-0273 docket for the development of a feed-in tariff. In September 2009 it introduced a feed-in tariff incentive scheme, through the local utility, with 20 year length and rates still being defined.

Maine

In February 2010 the State of Maine has introduced an incentive system, through the local utility, with 20 year rates average 10 c$/kWh and with a cap equal to the unit cost of the project plus a return on investment.

Comparison

In order to compare the incentive schemes worldwide the profitability of a 10 MW CSP plant, as IRR value, has been calculated according to the following assumptions:

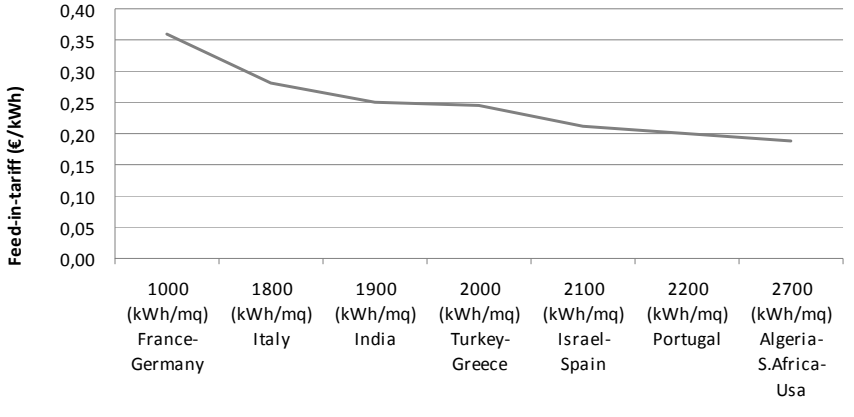
- the feed-in tariff value varies according to the Country
- the specific production value varies according to the Country
- the capital expenditure is given as a fixed value without taking into account possible saving in the construction costs in certain Country

The following table represents the specific production for each Country.

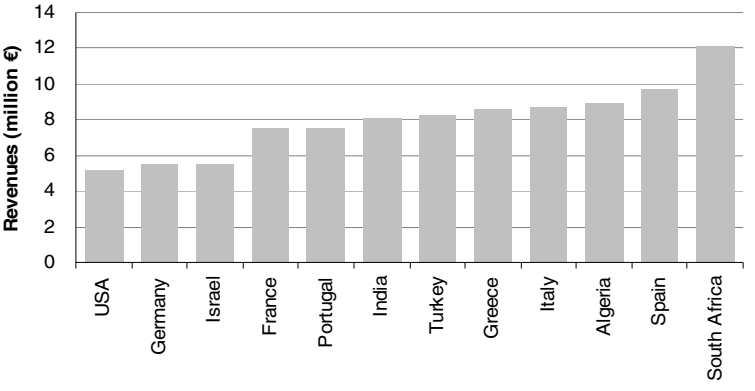
	Solar radiation (kWh/mq)	Specific production (kWh/kW)
Middle East		
Israel	2.100	3.600
Turkey	2.000	3.429
Africa		
Algeria	2.700	4.629
South Africa	2.700	4.629
Asia		
India	1.900	3.257
Europe		
Spain	2.100	3.600
France	1.100	1.886
Germany	1.000	1.714
Greece	2.000	3.429
Italy	1.800	3.086
Portugal	2.200	3.771
Americas		
USA	2.650	4.543

Feed-in tariffs are calculated in order to guarantee a certain return on equity covering project costs. As the profitability depends mostly on revenues and those depend on both solar radiation and feed-in tariff, generally the higher solar radiation, the lower the feed-in tariff is.

As shown in the following figure feed-in tariffs decrease with solar radiation so that those in Germany are much higher than those in USA.

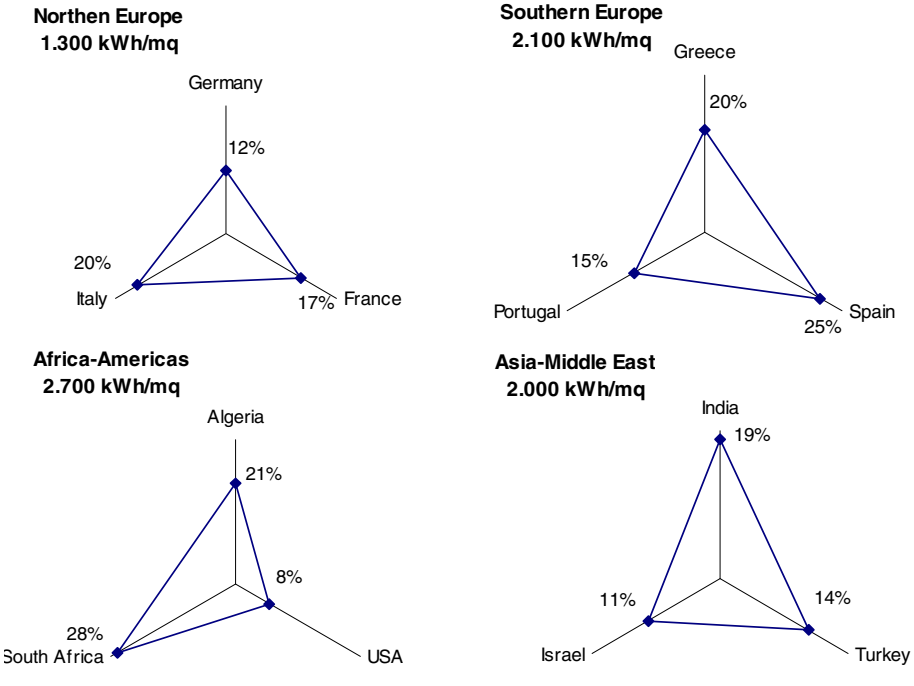


Combining the solar radiation with the feed-in tariff values for each Country the revenues result as follows.



In terms of profitability Spain and South Africa show the highest IRRs respectively of 25% and 28% as Spain grants one of the highest feed-in tariff and South Africa benefits both of the highest solar radiation and of one of the highest feed-in tariffs.

As USA benefit the highest solar radiation, they grant the smallest feed-in tariff and show the lowest profitability.



3 Conclusion

Concentrating solar power technology took a significant step forward at the end of the 1980s with the construction of the SEGS plants in California. Since there and until the massive construction of solar plants in Spain started from 2005 no additional solar power plants have been build following the bankruptcy of the developer of the SEGS projects, LUZ International Limited. SEGS projects represented a highly attractive opportunity for investors with more than one billion dollars in private capital raised in debt and equity financing. This demonstrates that CSP development relies on its attractiveness form a financial point of view. Indeed the following failure of the development is due to USA policy that avoided implementing feed-in tariffs, preferring policies that have proven less effective for encouraging the development of the solar industry. Like in USA, recent CSP development in Spain is primarily due to the considerably attractive financial package to potential investors raised by the Royal Decree.

Looking at the figure comparing the profitability, Spain stands up with one the highest IRR value, also thanks to inflation recovery. Only the recent South African incentive schemes guarantee higher IRR and for this reason CSP is expected to be an attractive technology in this Country for many investors in the near future. Below those two Countries there are many others such as Italy, Algeria, Greece, India and Turkey that have more or less the same value of IRR, around 20%, which is still quite adequate for investors. However CSP development in those Countries does not depend only on its profitability. In particular Greece still suffers from the economic downturn and no projects are expected to be built and Italy still focuses its financial resources on photovoltaic solar plants. India, Algeria and Turkey on the other hand represent an interesting potential target for CSP development.

Comparing the revenues figures, apart from South Africa, most of the Country have similar revenues and this demonstrates an adequate incentive schemes for investors. Indeed there is no reason why the success gained in Spain can not be replicated anywhere else. Countries with lower revenues, such as Israel, Germany, France, Portugal and USA, instead are not expected to be an attractive target for investors in CSP technology.

Economical Analysis of SOFC System for Power Production

Andrea Colantoni[1], Menghini Giuseppina[1], Marco Buccarella[2], Sirio Cividino[3], and Michela Vello[3]

[1] Department of Geology and Mechanical, Naturalistic and Hydraulic Engineering for the Territory – University of Tuscia 01100 Viterbo (Italy)
ergolab@unitus.it
[2] Department of Industrial Engineering - University of Perugia
Via Duranti 67 06125 Perugia (Italy)
[3] Department of Agrarian and Environment Science - University of Udine
Via delle Scienze 208, 33100 Udine (Italy)

Abstract. A comprehensive economic analysis of solid oxide fuel cell (SOFC) systems is presented in this paper. The analyzed system consists of a desulphurization system, a pre-reforming reactor, a fuel cell stack, an after-burner, heat exchangers, a power converter, a control system, blowers, pumps, start-up heaters and piping. Since most of these components are still in the prototype or early commercialization phase, their unitary cost (€/kW) as well as the global SOFC system cost is assessed as a function of time. Current investment costs and assumptions regarding future technological developments have been taken into account to evaluate the characteristic cost curves. These curves shows three segments with different slopes in costs trend which correspond to research and development stage, introduction of the product into the market and "market stabilization". A sensitivity analysis has been carried out to identify the most critical components within the system. Results demonstrated the criticality of fuel cell stack, balance of plant and control system costs; their optimization could contribute to a drastic reduction in the global cost. Finally, the assessment of the cost using net present value (NPV) as control parameter was performed.

Keywords: SOFC system, biomasses, renewable energy, biofuels.

1 Introduction

The problem of electrical energy generation is a controversial topic which from a long time has attracted and still attracts the interest of many different research sectors. Problematic topics such as scarcity of fossil fuels or high CO_2 emissions directly related to worldwide energy consumption (estimated around 25000 million metric tons [1]) lead to investigate new energy sources. One of them, linked to the global hydrogen project, is the development of fuel cell technology. Among different types of fuel cells, high temperature fuel cells (molten carbonate, MCFC, and solid oxide fuel cells, SOFC) present a special interest. All around the world important efforts are made on research of high temperature fuel cells manufacture and performance.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 270–276, 2011.

On SOFC technology, main efforts are focused on a power range between 1 and 25 kW. Most relevant pilot experiences have been developed by private companies. Sulzer Hexis tested a 1 kW stack for residential applications. Ceramic Fuel Cells has designed and manufactured a 25 kW system [2]. Studies dealing with SOFC system operation and economical feasibility have been performed for different applications [3]. Present costs are still too high, but the further research and optimization of operation will lead to a diminution of costs allowing also market penetration of this technology. Studies financed by US Department of Energy's Solid State Energy Conversion Alliance (SECA program) [4] have as final objective the reduction of kW cost (produced by SOFC). There are many different applications for the SOFC: auxiliary power units used in mobile applications, distributed power generation… Also Delphi Automotive Systems and Siemens Westinghouse have developed SOFC systems of different power capacity [5], while General Electrics Hybrid Power Generation System (GE HPGS) is focused on the study of power generator systems of 3 to 10 kW. The relevance of these studies is not only limited to energy production or energy savings topics but also have influence on environmental or social matters.

This paper presents a detailed analysis of the total costs of a 1 kW SOFC system which will be helpful to assess which part/component of the system has a higher impact on the global investment and to perform a comparison of component costs along time. The curve describing cost evolution has been built based on the assumption of the existence of three differenced phases on cost history depending on market position:

- Cost trend prior to introduction in market (research and prototype stage),
- Cost trend during market penetration,
- Trend to the minimum cost (cost stabilization).

Taking from literature the total cost of an internal combustion engine (ICE) or gas turbine (GT) system and considering that they are already consolidated technologies, thus with a narrow margin of technical, environmental impact or economical improvement; fuel cells become a promising field to develop/investigate. Low pollutant emissions, high efficiencies and flexibility on fuel feeding (fuel cells are capable of aprovechar low calorific power gases) are some of the important advantages presented by fuel cells when compared to traditional technologies such as ICE or GT system.

2 Material and Methods

2.1 SOFC System Description

This paper deals with the economical analysis of SOFC systems fuelled by natural gas for residential applications. As seen in figure 1, the SOFC system is composed of:

- Fuel Processing:
    - Desulphurizer: Reactor with an activated carbon bed which absorbs H2S contained in natural gas composition.
    - Pre-reformer: Natural gas is partially converted in hydrogen in this block prior to entering the stack to avoid carbon deposition and displace equilibrium reaction to hydrogen generation.

- Power Generation:
  - Fuel Cell Stack: Planar fuel cell stack consisting on 40 cells of 100 cm^2 active area (temperature 850º C).
  - After-burner: Combustor where fuel in anode off-gas flow mixed with remaining oxygen in cathode off-gas is burnt. It produces a supplementary heat source to be used in and outside the system.
- Power Conditioning: Converter and inverter which transform DC to AC
- Control System
- Start-up System
- Heat Management

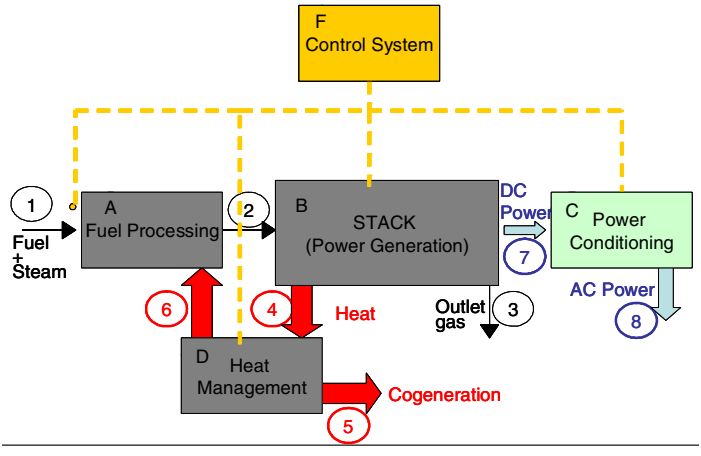


Fig. 1. SOFC system block diagram

2.2 Cost Analysis Methodology

Costs of the system have been studied from three different perspectives: evolution in time, sensitivity analysis and net present value. The methodology used for developing any of these studies is described in this section. From a parallel simulation study of a 1,12 kW SOFC system working at 800º C with fuel utilization 75%, current density 0,4 A/cm2, electrical efficiency 43,5% and thermal efficiency 22,9% are obtained. At present, the total cost of the SOFC system is 6700 €/kW when fuel cell stack is estimated to cost 2600 €/kW. Traditional power generators such as internal combustion engines (ICE) cost around 900 €/kW [7].

2.3 Cost Time-Evolution

This section, cost time-evolution, is an essential part to understand the economical feasibility of a SOFC system. The cost of every single component of the system has been independently analysed as a function of cost assumptions and initial year of market penetration. Cost trend is assumed to follow a curve consisting of three different phases:

1. First phase: represents the introduction into market
2. Second phase: represents the commercialization period,
3. Third phase: represents the trend to the minimum cost asymptote.

First and second phases are assumed to be linear and the third phase is assumed to follow an hyperbolic curve. As all commercialised objects, the SOFC system components will tend to a minimum cost in the time towards an asymptote. As the information dealing with the year of penetration into the market is difficult to find for every SOFC system components, estimations are made. From literature, 1955 is chosen for the balance of plant components and 2025 is assumed for SOFC stack introduction in the market. The starting year for the development of every technology required in the system is considered 1915 for all components of the BoP and 1980 for the SOFC stack. The simple assumption of choosing the same year of market penetration or initial development for every equipment of the system except the stack, even if it is not extremely accurate does not introduce any relevant error in the present costs of technology. As prior of 1980 SOFC stack was not a reality, 1980 has been the starting point for analysing costs time-evolution of the whole system. Apart from SOFC stack, technology of all other components within the system is situated/located in the third stage of cost evolution, following an asymptotic trend to minimum cost. As SOFC stack is a relatively recent technology and due to the fact that it has not still been introduced as a product into market, its cost is defined through the first phase of the curve.

2.4 Sensitivity Analysis of Partial Costs

Once the costs time-evolution has been established/calculated, these results are analysed through comparison of histograms built for years: 1980, 2005 and 2085. Every one of these graphs presents the costs distribution by component highlighting which parts of the system have a heavier cost weight. Present cost of single parts (expressed as €/kW) has been gathered from literature [6] and shown in table 1.

Table 1. Reference costs of the SOFC system components

Components	Assumed Initial Year	Initial Cost (€/kW)	Present Cost (€/kW)
Fuel Cell	1980	3300	2600
Converter-Inverter	1915	715	150,59
Reformer	1915	775	90,91
Heat Exchanger	1915	1125	290,32
Burner	1915	1010	250,37
Control System	1915	2025	1190,08
Peripheric	1915	3065	2050,05
Blower	1915	625	90,91

2.5 Net Present Value

Net Present Value (NPV) is the difference between cash flow and initial investment; with cash flow that depend on benefits, personal and maintenance costs. This economical parameter allows to calculate the payback time of a project (product or system) and permits to make the comparison between the convenience along time of any technologies.

Net Present Value is calculated as:

$$NPV_n = -I_0 + \sum_{k=0}^{n} ACF_k \tag{1}$$

where:
I_0 = investment cost [€],
n = years.

$$ACF_k = N_k \cdot \frac{1}{(1+i)\,\hat{}\,k} \tag{2}$$

where:
i = discount rate,
N = Cash Flow [€/year],
ACF = Discounted Cash Flow [€/year],

The cash flow is the difference between the benefits (through the sale of the thermal and electrical power) and total cost (P + M).

$$N_k = R_k - (P_k + M_k) \tag{3}$$

$$P_k = P_{k-1} \cdot (1+r) \tag{4}$$

$$M_k = M_{k-1} \cdot (1+i) \tag{5}$$

where:
R = Benefits [€/year],
P = Personal cost [€/year],
M = Maintenance cost[€/year],
r = Inflation rate.

Maintenance costs are assumed to depend on discount rate, while personal costs is assumed to depend on inflation rate. Electrical power costs have been considered with and without governmental incentives.

2.6 Results of the Economical Study of the SOFC System

Through the methodology described in section 3, the economical results of the SOFC system are obtained.

The boundary conditions which define the curves, shown in table 1, are: initial year for technology birth, initial cost, year of market introduction, etc.). The two first phases indicate a dramatic decrease in SOFC stack cost in the research stage and market stabilization.

Table 2. Sensitivity Analysis of Partial Costs of the System along time

	2005		2085	
Components	**Cost (€/kW)**	**%**	**Cost (€/kW)**	**%**
Fuel Cell	*2600,00*	*38,73*	*501,14*	*11,87*
Inverter	*150,59*	*2,24*	*102,40*	*2,42*
Reformer	*90,91*	*1,35*	*52,63*	*1,25*
Heaters	*290,32*	*4,32*	*230,77*	*5,46*
Burner	*250,37*	*3,73*	*193,14*	*4,57*
Control System	*1190,08*	*17,73*	*1116,28*	*26,43*
Peripheric	*2050,05*	*30,54*	*1973,77*	*46,74*
Blower	*90,91*	*1,35*	*52,63*	*1,25*
Total	6713,23	100	4222,76	100

Also in the curves corresponding to the other components of the system, three different phases are pointed out. The most relevant difference between SOFC stack and the BoP cost evolution curves is the present situation as developed technologies due to the assumption of initial year of market introduction. SOFC stack is clearly the youngest technology in the system, in fact SOFC stack is the only component which has not yet been introduced into the market. The trends of others system components aren't interesting because they are more old technology.

Cost data for every component expressed as cost percentage and €/kW are gathered in table 2.

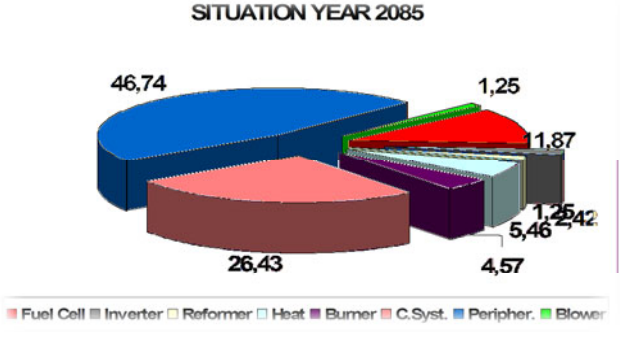


Fig. 2. SOFC system block diagram

In Figures 2 present the division of partial costs in SOFC system represented as pie chart graphics under three temporal situations: 2085. This graphic show that the most

influential technologies on total costs are the SOFC stack, the peripheral components (piping, desulphurizer, starter heater, installation, etc.) and the control system (safety equipment, pressure-relief valves, sensors, computer hardware, etc.).

3 Conclusions

The cost analysis of the total system and single system components shows that the SOFC will be as competitive as traditional technology (ICE and GT), but it is required the diminution of costs in general and specially the SOFC stack and the fuel processing (desulphurizer and pre-reformer). When the technology of fuel cells would enter into the market, its cost will be lowered. For others components the diminution in costs will depend on the different materials used.

The maintenance (ordinary and extraordinary) costs of the fuel cell system are of great importance. The total cost depence of the maintenance; if the maintenance is 10% of initial investment the fuel cell system not convenient, because the initial investment is never recovered. If the maintenance of 5% of initial investment the payback time is seven years, but more bottom of the ICE. The lowering of maintenance costs is very important and is the main goal that will be analyzed in future. The critical costs of the total system from the economical point of view are the initial investment, the maintenance costs and the income in the market of the fuel cell. Thus, these parameters/values would be the main goals to be optimized.

Acknowledgment. The authors gave equal contribution to the realization of the present work.

References

1. International Energy Annual 2005, International Energy Agency (2005)
2. Foger, K., Godfrey, B., Pham, T.: Development of 25 kW SOFC system. Fuel Cells Bulletin 5(2), 9–11 (1999)
3. Singhal, S.C.: Solid oxide fuel cells for stationary, mobile and military applications. Solid State Ionics, 152-153, 405–410 (2002)
4. Singhal, S.C., Mcvay, G.L.: In: Yokokawa, H., Singhal, S.C. (eds.) Proceeding of the Seventh International Symposium on Solid Oxide Fuel Cells, The Electrochemical Society, vol. 16, p. 53 (2001)
5. Williams, M.C., Strakey, J.P., Singhal, S.C., Subhash, C.: US distributed generation fuel cell program. Journal of Power Sources 131, 79–85 (2004)
6. Blesi, M., Ohl, M.: Fuel Cells: Bottom-up interpretation of the experience curve. In: Workshop of EU-EXTOOL and IEA EXCEPT (2003)
7. Alanne, K., Saari, A.: Sustainable small-scale CHP technologies for buildings: the basic for multi-perpective decision-making. Renewable and Sustanaible Energy Reviews 8, 401–431 (2004)

Modelling the Vertical Heat Exchanger in Thermal Basin

Maurizio Carlini[1] and Sonia Castellucci[2]

[1,2] CIRDER, University of Tuscia, Via San Camillo de Lellis s.n.c, 01100 Viterbo (Italy)
maurizio.carlini@unitus.it
[2] CIRDER University of Tuscia, Via San Camillo de Lellis snc, 01100 Viterbo (Italy)
sonia.castellucci@unitus.it

Abstract. In geographical area characterize by specific geological conformations such as the Viterbo area which comprehend active volcanic basins, it is difficult to use conventional geothermal plants.

In fact the area presents at shallow depths thermal falde ground water with temperatures that varies from 40 to 90°C geothermical heat pumps cannot be utilized.

In these area the falde thermal can be exploited directly as hot source using vertical heat exchanger steel tubes without altering the natural bilance of the basin.

Through the heat exchange that occurs between the water in the wells and the fluid that circulates inside the heat exchanger ,you can take the heat necessary to meet the thermal requirements.

The thermal energy used is transferred by natural convection from the thermal water to the heat exchanger in the well. The heat exchanger consists of a U-shaped tube where inside there is, as circulating fluid, water.

To conduct the experiment it is necessary to examine in detail the internal temperature of the well, and the characteristics both input and output of fluid in the plant. For this reason, were inserted in the well three temperature probers for measure the temperature inside. There will be, at constant time intervals, some sampling to evaluate pH, electrical conductivity, density and salinity. These parameters are needed to evaluate the possible corrosion of the materials included in the well.

The target of the project is to analyze in detail the plant for the exchange of heat with the thermal basin creating a model of heat exchange.

Keywords: heat, thermal aquifer, thermal energy.

1 Introduction

The geology of the volcanic basin area in the Italian region of Viterbo, defines a situation which is very rich and unique, from the point of view of energy source, but it is difficult to be approached with the conventional low-enthalpy geothermal systems for domestic, industrial and agricultural applications. [1]

The geological area results in the coexistence of overlapped interacting aquifers. The shallow volcanic aquifer, characterized by fresh waters, is fed from the area around the Cimini Mountains and is limited at its base by the semiconfining

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 277–286, 2011.

marly-calcareous-arenaceous complex and low-permeability clays. To the west of Viterbo, vertical upflows of thermal waters of the sulphate-chloridealkaline- earth type with higher gas contents, are due to the locally uplifted carbonate reservoir, the reduced thickness of the semiconfining layer and the high local geothermal gradient. The hot waters (30–60°C) are the result of deep circulation within the carbonate rocks (0.5–1.8 km) and have the same recharge area as the volcanic aquifer. The upward flow in the Viterbo thermal area is at least 0.1 m3/s. This flow feeds springs and deep wells, also recharging the volcanic aquifer from below. [2]

Interpretation of the stratigraphy of wells drilled in the area, correlated with surface geology, allowed the reconstruction of the hydrostratigraphy of the thermal area.

Pyroclastic and lava formations belonging to the Cimini and Vico complexes occur from the first 10 m down to about 100 m depth. These volcanic rocks constitute the shallow unconfined aquifer of the area, due to the primary and secondary porosity that characterize volcanic rocks. The thickness of the volcanic aquifer decreases in the thermal area where it includes more layers of travertine deposits. Pliocene-Pleistocene clayey deposits were found beneath the volcanic aquifer.

Tale area è stata monitorata in vari studi geologici sono disponibili, infatti una serie di rilevazioni che permettono di conoscere le temperature e le portate dei pozzi (Tab1).[3]

Table 1. Elevation discharge rate and temperature of the main thermal springs of the area of Viterbo

Spring or group of springs	Name	Elevation (m asl)	Discharge (L/s)	Temperature (°C)
3	Bagnaccio group	310–319	10	38–64
4	Monterozzo group	305–319	5	27–51
7–8	Zitelle group	289–308	6.5	39–65
9	Carletti	285	1.5	57
10	Bullicame	298	10.4	57
17	S. Cristoforo	245	0.1	53
21	S. Sisto	230	3.0	57
26	Bacucco group	315–320	2	36–49
27	Urcionio group	305–319	20	28–60

The thermal waters are analised in the main springs and deep wells of the area.

The thermal springs and deep wells of the area considered includes sulphate-chloride alkaline-earth waters, with high salinity. (EC=2,700–3,300 μS/cm) and temperature (T=30–62°C).

Given the geological characteristics of the municipality of Viterbo and given the impossibility of legislation drilling new wells, the first phase of the project has served to raise on the territory of a well with the characteristics required. In that respect, it was a well identified and holding Plant Bulbs, Mr Daniel Cortese has provided the area for the construction of the trial.

The well tested relate to the volcanic aquifer in the central sector of the thermal area, in the zone which has a low hydraulic gradient Fig 2. The volcanic aquifer is interbedded with low permeability swampy deposits and travertine. The local substratum, at a depth of more than 50 m, consists of the marly-calcareous-arenaceous complex [4].

According to the drilling log, the well depth is 66 m below grade with an 150 mm nominal diameter casing installed to the well bottom. The static water level in the well is, about 1.5 m above campaign plan.

The rebar on the DHE assembly was placed on the well bottom at approximately 44 m below grade, some 13 ft (3.96 m) shallower than indicated on the drilling log. This indicates that either debris has fallen down the well, or that about 13 ft (3.96 m) of rock fragments have entered the well through the casing perforations. The Geo-Heat Center commonly observes this phenomena of basalt rock fragments entering wells through torch-slotted well casings.

2 Modelling Heat Exchanger in the Well Thermal

The thermal energy used is transferred by natural convection from the thermal water to the heat exchanger in the well. The heat exchanger consists of a two U-shaped pipe stainless steel where inside there is, as circulating fluid, water [5].

The heat exchanger was placed on the well bottom at approximately 44 m below grade, some 22 m shallower than indicated on the drilling log. This indicates that either debris has fallen down the well, or that about 22 m of rock fragments have entered the well through the casing perforations. commonly observed this phenomena of basalt rock fragments entering wells through torch-slotted well casings.

The starting point is the general equation for heat transfer across a surface which is:

$$Q = A \cdot U_D \cdot \Delta T_{LM} \tag{1}$$

Q : heat transferred per unit time, W
A : heat-transfer area, m^2
U_D : overall heat transfer coefficient, $W/m^2\ ^oC$
ΔT_{LM} : mean temperature difference, the temperature driving force, ^oC :

$$\Delta T_{LM} = \frac{(T - t_2) - (T - t_1)}{\ln\left[\frac{(T - t_2)}{(T - t_1)}\right]} \tag{2}$$

T : temperature hot fluid
t_1 : temperature cold fluid, in the feeding pipe
t_2 : temperature cold fluid, in the outgoing pipe

For the heat exchange across a typical heat-exchanger pipe the relationship between the overall coefficient and the individual coefficients, which are the reciprocals of the individual resistance, is given by [7]:

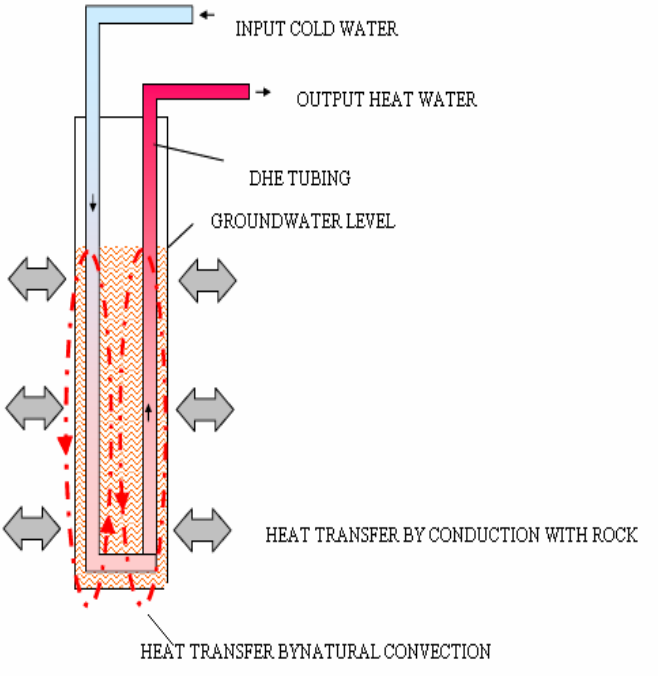


Fig. 1. Schematic heat exchanger in the well[6]

$$\frac{1}{U_D} = \frac{1}{h_{i0}} + \frac{d_0}{2k}\ln\frac{d_i}{d_0} + \frac{1}{h_0} + R \tag{3}$$

U_D : the overall coefficient based on the outside area of the tube, W/m^2 ^{o}C
k : thermal conductivity of the tube wall material, steel 50 W/m ^{o}C
d_i : inner pipe diameter, m
d_0 : outer pipe diameter, m

h_{i0} : outside dirt coefficient (fouling factor), W/m^2 ^{o}C
h_0 : outside fluid film coefficient, W/m^2 ^{o}C

R : dirt factor pipe for water 0,0002 (m^{2}h ^{o}C/kcal) for T < 50^{o}C and 0,0004 (m^{2}h ^{o}C/kcal) for T>50^{o}C.

The first step is to calculate the overall heat transfer coefficient according to the following scheme [8]:

The coefficient of convention is calculated by the number of Nusselt (Nu) with the relations :

$$h = \frac{Nu\lambda}{D} \tag{4}$$

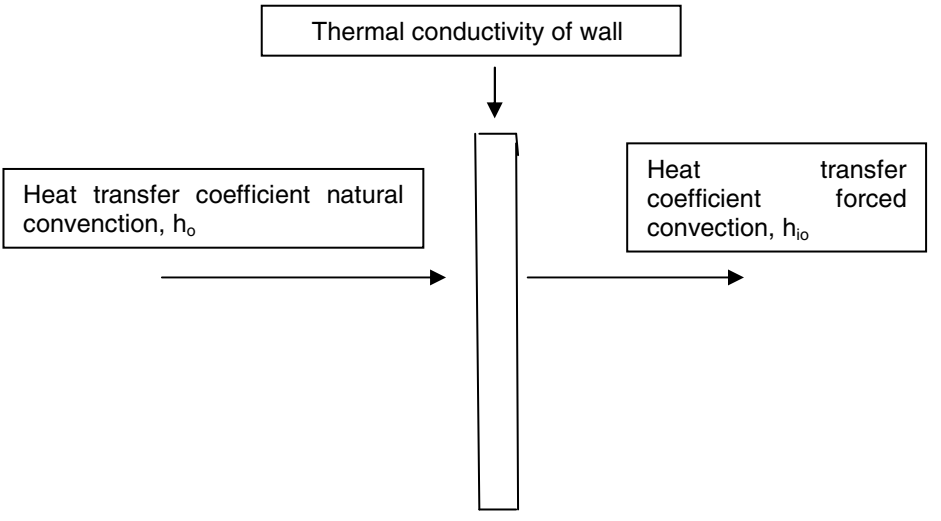


Fig. 2. Heat transfer coefficients

The number of *Nu* generally depends by Grashof (Gr), number of Prandl (Pr), number of Reynolds (Re) [9]:

$$Nu = f\left(\mathrm{Re}, Gr, \mathrm{Pr}\right) \tag{5}$$

In case of natural convention the number of Nu does not depend by the number of Reynolds, because the fluid is resting , therefore:

$$Nu = f\left(Gr, \mathrm{Pr}\right) \tag{6}$$

Therefore in the case of a vertical stratum of a resting fluid [10]:

$$Nu = \int_0^l \left(-\frac{\partial T}{\partial Y}\right)_{Y=0} dX \tag{7}$$

where:
-T nondimensional temperature
-X,Y: non dimensional space coordinate

The number of *Nu* depends by two parameter ,the number of *Rayliegh:*

$$Ra = Gr \cdot \mathrm{Pr} \tag{8}$$

and the thermal stratification parameter, S, defined as:

$$S = \frac{1 dt_{\infty,x}}{\Delta t dX} \tag{9}$$

where:

- $t_{\infty,x}$, temperature reference
- Δt, temperature potential (t_w- $t_{\infty,x}$)

If the ambiental conditions are isothermal, the numerical value of *S* is zero. In this case you can use a sperimental formula for the number of Nu and the heat transfer of coefficient (Cheng Minkowycz)[11]:

$$Nu = cGr^{b}\,\mathrm{Pr}^{a} \tag{10}$$

The parameters a , b, c depence on the geometrical system and by the movement of the fluid during the natural convention wich can be laminar or turbulent. Becouse the fluid is resting , it is not possible to use Reynolds numbers to calculate the regime of moviment, that is why we utilize the number of *Rayleigh* (Ra) [13]:

$$Ra = Gr \cdot \mathrm{Pr} \tag{11}$$

Once we have calculated the numbers of Pr e Gr we calculate Ra, we take a look at the table wich will give us the parameters a,b,c,.The Scientific laminare regime is valid for:

$$Ra \leq 10^{9} \tag{12}$$

In this case you can use the formula of the number 2 table [14].

Table 2. Coefficient natural convection

Situazione geometrica	*Campo di validità (Ra)*	$Nu = C\,Gr^{b}Pr^{a}$ C	b	a	*Autori*	*Note*
Strato verticale di altezza *H* e spessore *L*: una parete verticale più calda dell'altra	$< 2.000\ Pr$	1	0	0	Jakob	*Nu* e *Gr* calcolati in funzione di *L*. Relazioni valide per l'aria
	$(2 \times 10^{4} \div 2 \times 10^{5})\ Pr$	$0{,}18\left(\frac{H}{L}\right)^{-1/9}$	0,25	0		
	$(2 \times 10^{2} \div 11 \times 10^{6})\ Pr$	$0{,}065\left(\frac{H}{L}\right)^{-1/9}$	0,33	0		
	$< 10^{3}$	1	0	0	Emery e Chu	Idem, relazioni per liquidi, con $3 < Pr < 30.000$
	$10^{3} \div 10^{7}$	$0{,}28\left(\frac{H}{L}\right)^{-1/4}$	0,25	0,25		

3 Result and Discussion

During 10 months we have collected the materials from the thermic probes and at the inut/output of the hydraulic system referring to the temperature of the water in the sink and in the input/output of the hydraulic system (see also Fig. 6) [15].

The graphics 5,6 desings the stable temperature of the thermal water in function of a deep well with or without heat exchanger. During the testing we have two transitory phase followed by stable:

- Transitory phase depending on the exit of gasess of the well without heat exchange.
- Transitory phase at the begining of the heat exchange.

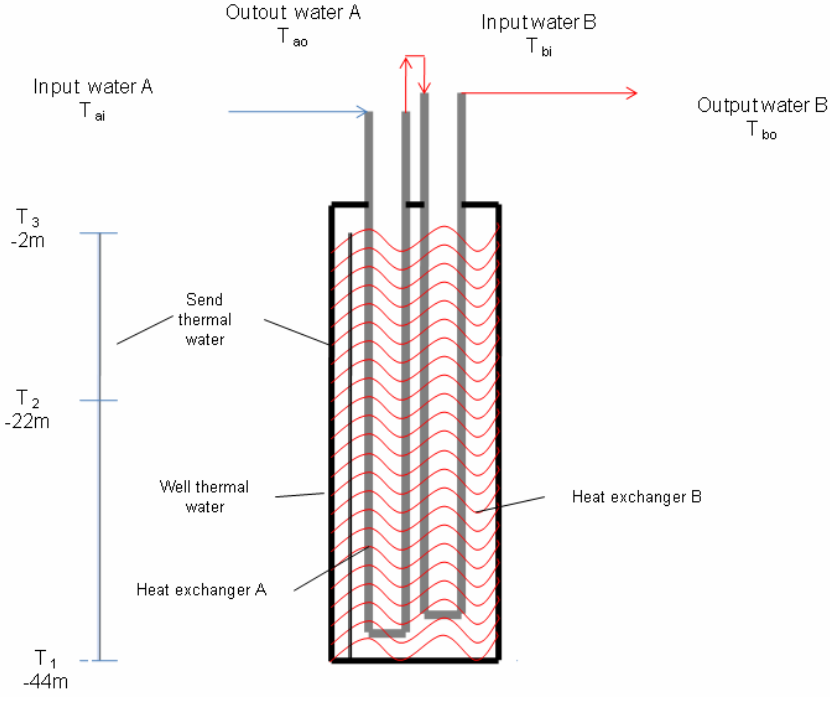


Fig. 3. Plan

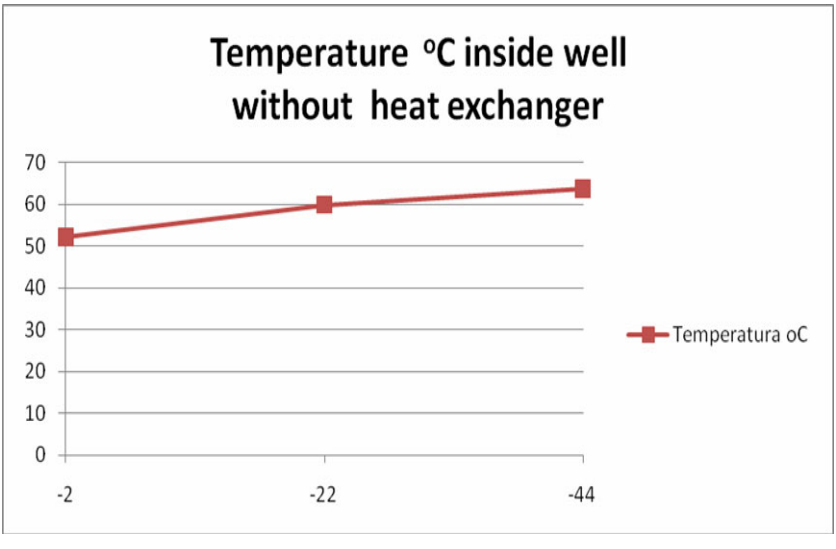


Fig. 4. Temperature °C inside well without heat exchanger

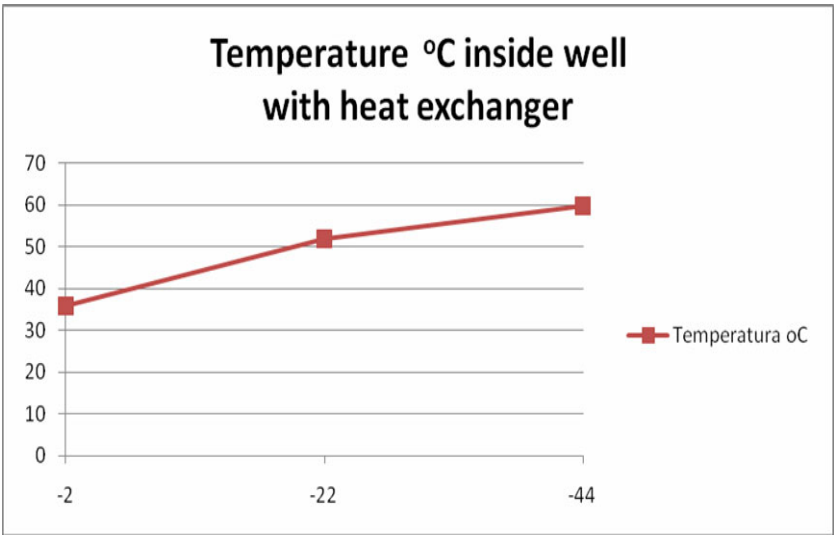


Fig. 5. Temperature °C inside well with heat exchanger

In figure number 7 you can see the temperature of the heat-carrying fluid during the heat exchange, wich you can see in figure number six .

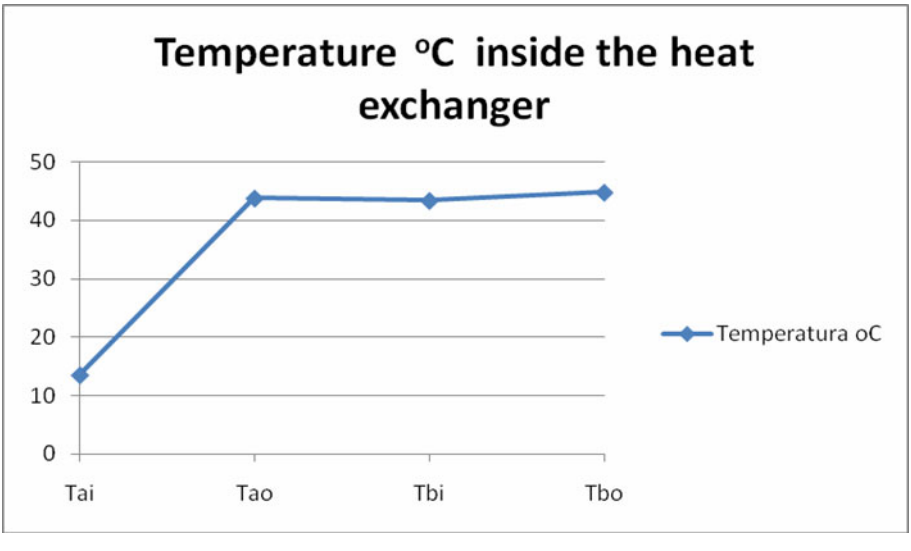


Fig. 6. Temperature °C inside the heat exchanger

4 Conclusion

Based on the collected informations we have verified the reliability of the heat exchanger inserted in the well. The heat exchanger appears to be sufficient on heating the water without any disturbance to the thermal system. For a correct heat exchange we must open the valve of the outlet gas The system is efficient for the realization of low temperature system for heating homes, industrial, commercial, craft and low-cost, without altering the delicate balance of the water thermal.

Bibliography

1. Lazzarin, R.: Ground as a possible heat pump source. Geothermische Energie 32/33, marzo/giugno (2001), http://www.geothermie.de/gte/gte3233/ground_as_a_possible_heat_pump_s.htm
2. Piscopo, V., Barbieri, M., Monetti, V., Pagano, G., Pistoni, S., Ruggi, E., Stanzione, D.: Hydrogeology of thermal waters in Viterbo area, central Italy. Hydrogeology Journal 14, 1508–1521 (2006)
3. Carotenuto, A., Casarosa, C., Dell'isola, M., Martorano, L.: An aquifer-well thermal and fluid dynamic model for downhole heat exchanger with a natural convection promoter. Heat mass Transfert 40, 4461–4472 (1997)
4. Angirasa, D., Peterson, G.P.: Natural covenction heat trasfer from an isothermal vertical surface to a fluid saturated termally stratified porous medium
5. Yavuzturk, C.: Modelling of Vertical Ground Loop Heat Exchangers for Ground Source Heat Pump Systems. Thesis to the Faculty of the Graduate College of the Oklahoma State University in partial fulfilment of the requirements for the Degree of Doctor of Philosophy (December 1999)
6. Lund, J.W., Chiasson, A.,"Toni" Boyd, T., Carr, D.: Assessment of Downhole Heat Exchangers in Existing Wells in Puna District, Hawaii, Geo-Heat Center (June 2007)
7. Talleri, M.: Applicazioni geotermiche negli impianti di attivazione termica della massa. In: Seminari Velta (February 2001)
8. Kavanaugh, S.P., Rafferty, K.: Ground source heat pumps - Design of geothermal systems for commercial and institutional buildings. In: ASHRAE Applications Handbook (1997)
9. Bonacina, C., Cavallini, A., Mattarolo, L.: Trasmissione del calore. Cleup, Febbraio (1994)
10. Olesen, B.W., Meierhans, R.: Attivazione termica della massa. In: Seminari Velta, 2001 (February 2001)
11. Cheng, Minkowycz: Free convenction about a vertical flat plate embedded in a porous medium with application to heat transfer from a dike. Journal of Geophysical Research 82, 2040–2044 (1977)
12. Yavuzturk, C.: Modelling of Vertical Ground Loop Heat Exchangers for Ground Source Heat Pump Systems. Thesis to the Faculty of the Graduate College of the Oklahoma State University in partial fulfilment of the requirements for the Degree of Doctor of Philosophy (December 1999)

13. Carlini, M., Castellucci, S.: Efficient energy supply from ground coupled heat transfer source. In: Taniar, D., Gervasi, O., Murgante, B., Pardede, E., Apduhan, B.O. (eds.) ICCSA 2010. LNCS, vol. 6017, pp. 177–190. Springer, Heidelberg (2010)
14. Carlini, M., Monarca, D., Biondi, P., Castellucci, S., Honorati, T.: A Simulation Model for the Exploitation of Geothermal Energy for a Greenhouse in the Viterbo province. In: Work safety and risk prevention in agro-food and forest systems. International Conference Ragusa SHWA 2010, Ragusa IblaCampus- Italy, September 16-18 (2010)
15. Sebastiani, E.: Lezioni di Impianti chimici, edizioni scientifiche SIDERA
16. Bakhoum, E., Toma, C.: Mathematical Transform of Travelling-Wave Equations and Phase Aspects of Quantum Interaction. Mathematical Problems in Engineering 2010, Article ID 695208 (2010), doi:10.1155/2010/695208

Optical Modelling of Square Solar Concentrator

Maurizio Carlini[1], Carlo Cattani[2], and Andrea O.M. Tucci[3,*]

[1] DiSAFRi, University of Tuscia
[2] Dip Mat, University of Salerno
[3] [CIRDER] University of Tuscia
uba@unitus.it

Abstract. This paper deals with the optical design of a photovoltaic solar concentrator composed by two squared reflection mirrors.

The optical configuration of the device, is based on the Cassegrain telescope and designed in order to maximize the fill factor over the irradiated area and by providing an easily manufactured shape.

The collection unit has been simulated and optimized by a ray tracing method aiming to obtain the concentration ratio about 1000 over the solar cell which has been placed behind the primary mirror to allow simple cooling.

This nominally independent collection unit can be assembled into modules and arrays that produce almost any requested power level.

Keyword: Photovoltaic, Ray tracing, Solar cell.

1 Introduction

Solar energy may be collected and exploited by solar concentrators.

The principal component of this device is the optical collection unit which gathers the radiated power from the sun and concentrates it over small area where it may be made available for various applications: indoor illumination through optical fibers [1], coupled system that can merge photovoltaic and thermal applications [2] and many other uses.

In the photovoltaic system the solar collectors help to decrease the amount of the expensive cells [3,4].

These components have been hence investigated, developed and optimized [5-8], in order to improve the efficiency and compactness and ease of fabrication.

Among the used optical shapes, the two stage mirrors configuration has been harnessed to achieve these results [9-12].

A solar collector based on the Cassegrain telescope configuration was investigated for the storage of solar energy in a chemical system [9]. In this paper the authors present a two stage configuration to design a first high-performances solar collector.

A two-stage solar concentration designed with imaging techniques was investigated and optimized for high flux concentration at high collection efficiency [10].

* Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 287–295, 2011.

The research discussed in this paper, was conducted using a modified telescope Cassegrain configuration. The modification consists of replacing the secondary mirror with a complementary hyperboloidal branch. This layout allows a large gap between the secondary mirror and the absorber, and has a restricted exit angle for the high efficiency configurations.

A solar concentrator with two circular reflective surfaces has been optimized to reach a high concentration ratio [11]. In this paper the authors investigate a small two stage solar concentrator configuration to obtain the highest concentration ratio possible when the concentrator tracing error exists. The best combination between the shapes of the mirrors is the parabolic and ellipsoidal one.

In general, it can be observed that the two stage configuration can be a basis for the development of a module of high performance.

The efficiency of a system composed by various elements of this type can be enhanced considering the combination of squared rim elements which presents a negligible packing losses [12].

The use of this shape has been evaluated for a high-concentration photovoltaic design based on parabolic dishes system. The collection unit is a miniaturized paraboloidal dish that concentrates sunlight into an optical fiber which is coupled with the solar cell [12].

This paper, addresses the optimization of the optical design of a bigger collection unit composed by two squared mirrors. The designed collection unit has been parameterized setting four variables and this module has been optimized by the code Zemax.

The optical quality of the final design has been evaluated through the spot diagram which are classified by field of view . This configuration aims to obtain an optimized fill factor and allows the manufacturing of a simple mounting structure exploiting the combine between the edges. Furthermore, the resulting minor weight reduces the power supply which is necessary for the engines of the tracking system.

2 Design of the Module

The collection unit is based on the Cassegrain telescope configuration and aims to a concentration ratio about 1000 over the solar cell, which has a squared area of 10mm side.

The first structure has been designed by considering a system which combines two circular mirrors [13]. The first mirror (also called primary mirror), M1, (see Fig.1) has a 320mm diameter, the conic constant K_1= -1 and the focal length $f1$= 200mm.

If we assume the angular size of the sun equal to 0.5° and the scale defined as:

$$S(\text{arcsec/mm}) = 206265/\text{f(mm)} \quad , \tag{1}$$

then the focal length of the second mirror (secondary) M2 is f=1145.92mm.

The constant conic for M2 and the transverse magnification may be obtained from these data and they are K_2= -2.02 m=5.72961 respectively.

The diameter of the secondary mirror has been fixed to 50mm, and the ratio of the ray heights at mirror margins is k=0.147. The position of M2, considering K and the

$f1$, is $S2$=29.41mm. The radius of curvature of M2 may be deduced from the relationship between the constant conic and the transverse magnification m: $R2 = 71.26\text{mm}$.

At last the position of the solar cell has been defined as $S'2 = 170.59\text{mm}$.

The significant data of the device are collected in the following table and represented in the following figures.

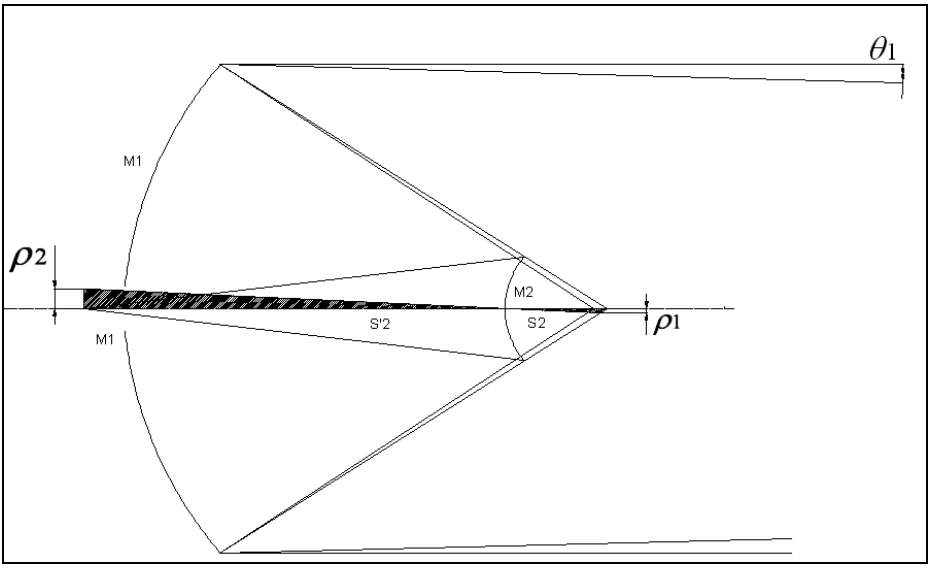


Fig. 1. Scale of a Cassegrain Telescope

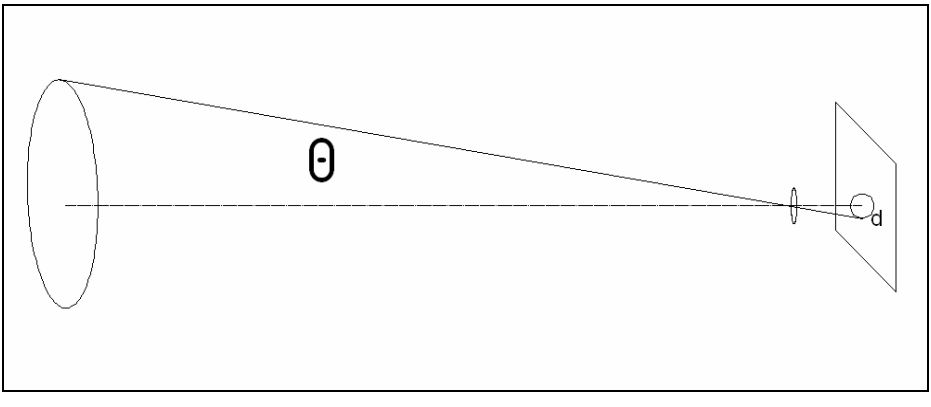


Fig. 2. Scale of a telescope

Table 1. Device data

SYMBOL	Description	Dimension
LM1	Diameter side M1	320mm
LM2	Diameter side M2	50mm
K1	Constant conic K1	-1
K2	Constant conic K2	-2.02
*f*1	Focal length f1	200mm
F	Focal length of solar concentrator	1145.92mm
R2	Radius of curvature of the M2	71.26mm
DM1M2	Distance between M1 and M2	170.590mm
S	Scale	179 arcsec /mm

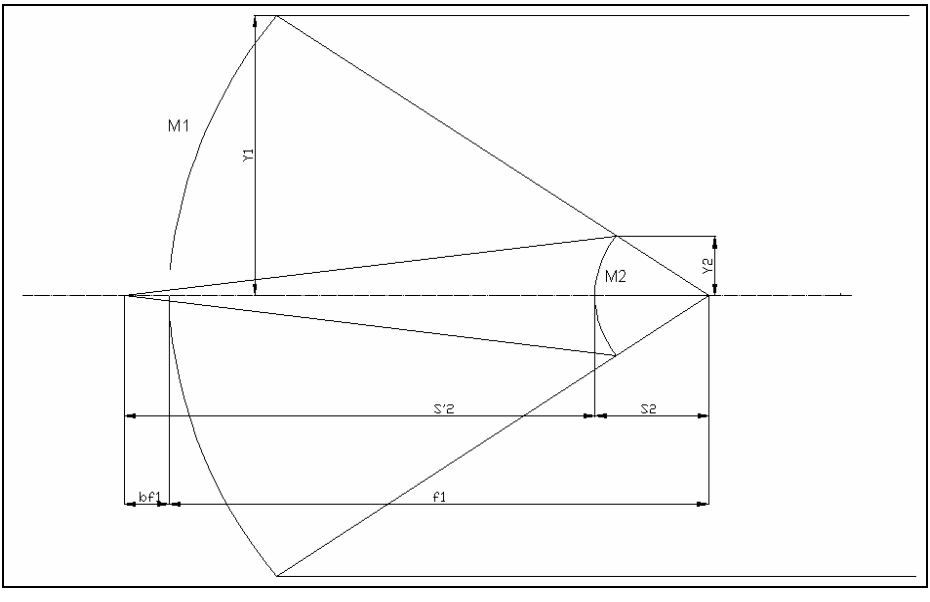


Fig. 3. Characteristics and parameters of the Cassegrain telescope

The calculated optical configuration has been then computed and simulated by Zemax software.

The field of view of 0.5° has been split into five sub fields (0°, 0°), (0°; 0.25°), (0.25°;0°), (-0.25°;0°), (0°;-0.25°) while the aperture stop of mirror M1.

Three wavelengths: 450nm, 1000nm, 1400nm are used to represent the adsorption spectrum of the solar cell and the coating of the surfaces has been set as a full spectral reflective one.

It has not been considered neither the misalignments nor the atmosphere.

A second run simulates the same collection unit with square reflective surfaces whose dimensions of the sides are equal to the previous mirror's diameters. A compared view between the two solar collector has been shown in figure 4. The collected surfaces of the square model is bigger than the circular one and the module can be packed without almost any losses in the array. The gain can be evaluated almost in a minimum of 10%. An example of array has been shown in figure 5.

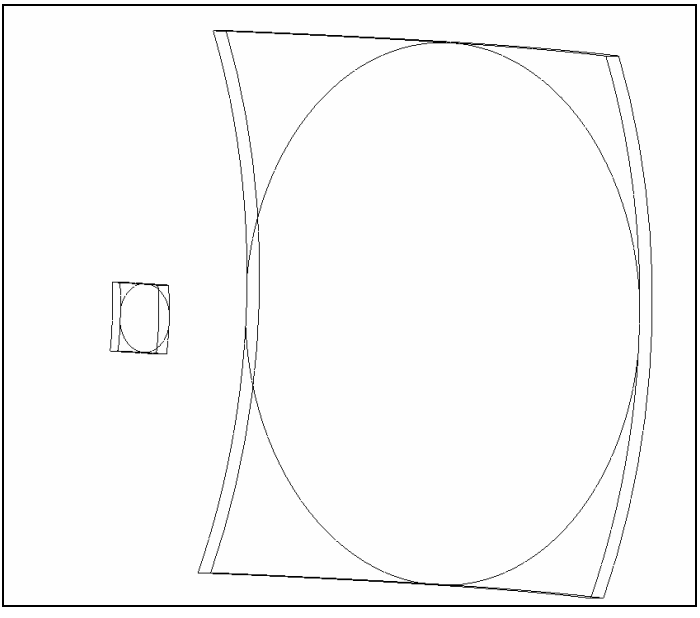


Fig. 4. Comparison between the two models

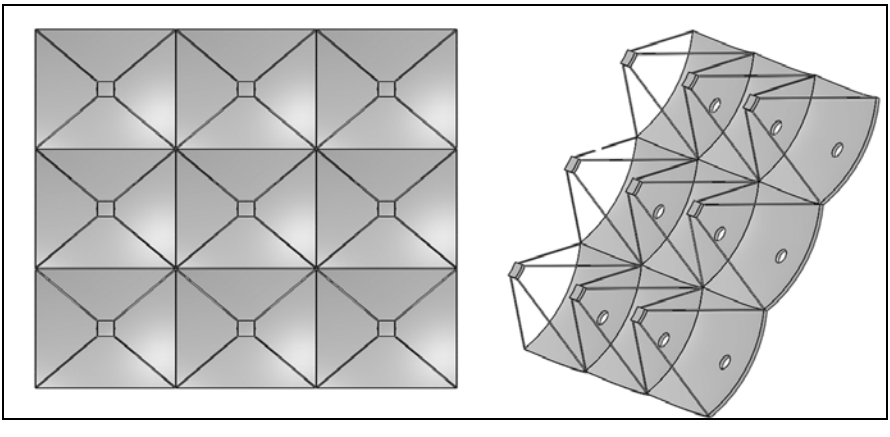


Fig. 5. Fill factor over radiated area

This new configuration has been computed and optimized to shift the focus of the device 20mm behind of the primary mirror (figure 6 and 7) maintaining the same concentration ratio.

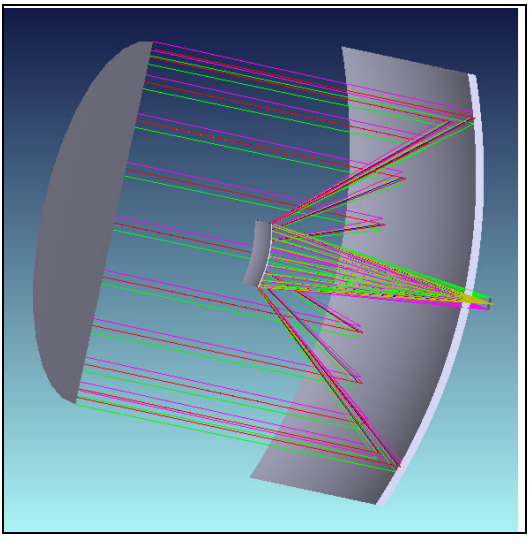


Fig. 6. Focus shifted behind the primary mirror

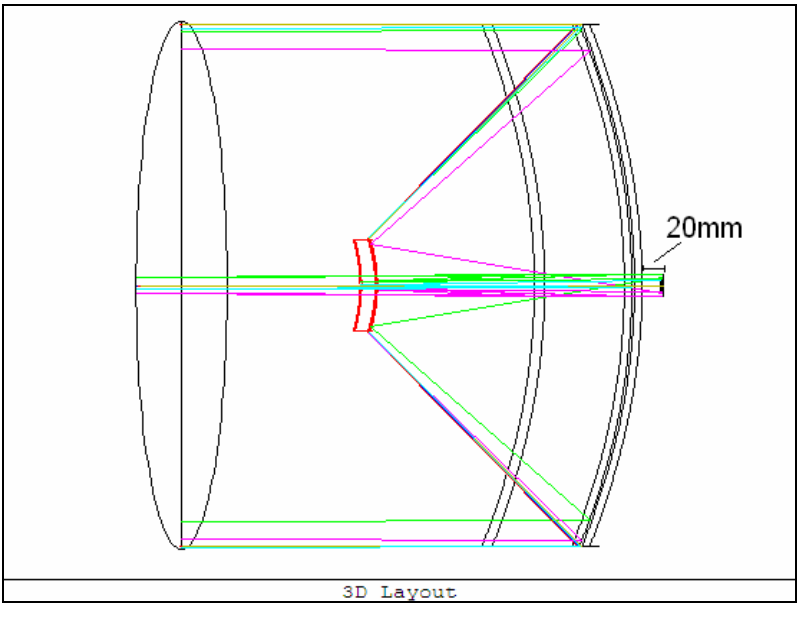


Fig. 7. Focus shifted behind the primary mirror

The backward position of the focus enable us to design an easily housing for the pv cell and for its cooling system. An example of integrated housing has shown in the following figure.

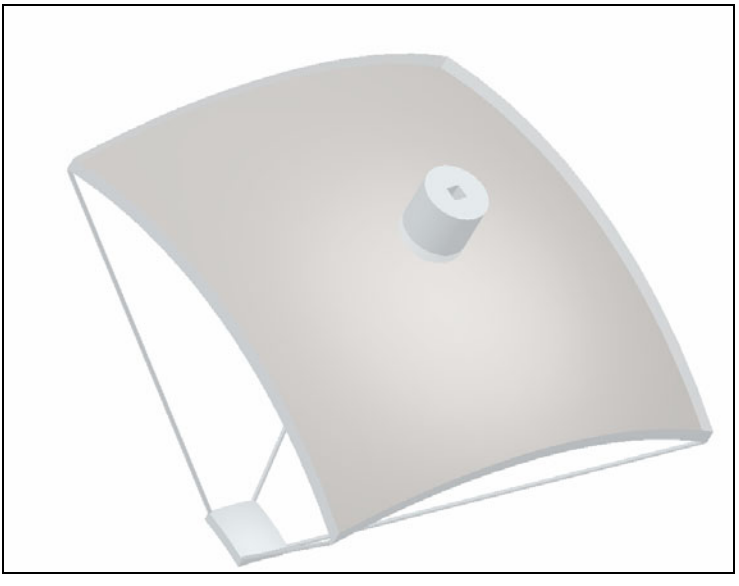


Fig. 8. Housing for the pv cell

Therefore the conic K2, the distance between M1 and M2 and both two mirrors' radii of curvature (related to the merit function [14]) have been assumed as variable parameters of the device.

The optimization algorithm exploited by Zemax, based on damped least squares method, has been executed about fifty cycles until the local minimum for the MF has been found.

The geometrical characteristics of the corresponding configuration for the optimized device are summarized in table 2. Figure 9 shows the spot diagram of the final configuration, collected as functions of field of view.

Table 2. Optimized Device data

Symbol	Description	dimension
LM1	Length side M1	340mm
LM2	Length side M2	50mm
K1	Constant conic K1	-1
K2	Constant conic K2	-2.157
R*1*	Radius of curvature of M1	405.101mm
F	Focal length of solar concentrator	1068.74mm
R2	Radius of curvature of M2	71.26mm
DM1M2	Distance between M1 and M2	166.520 mm
S	Scale	192.998 arcsec /mm

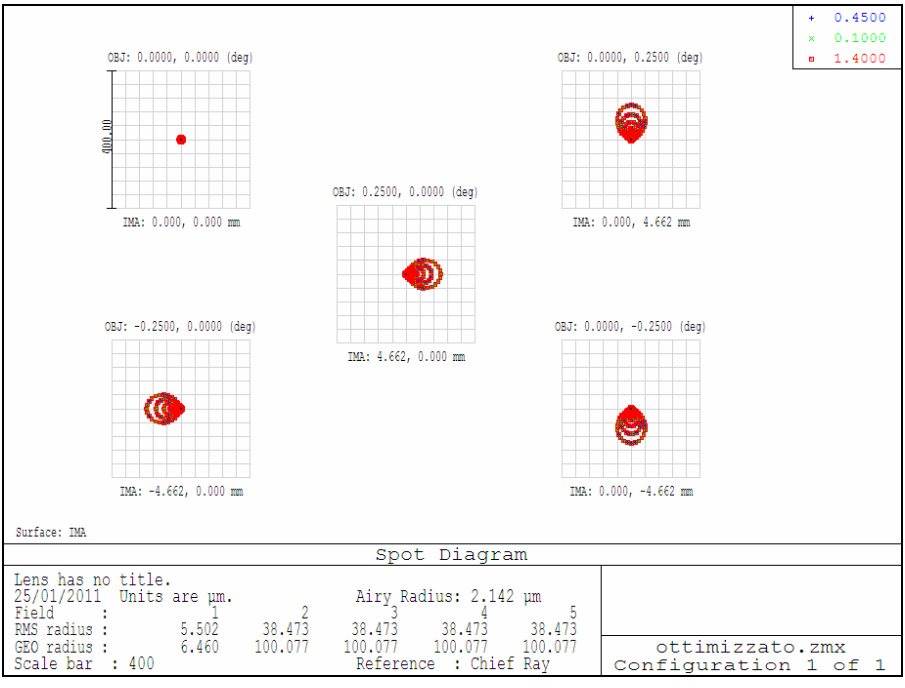


Fig. 9. Spot diagram of the optimized configuration

3 Conclusion

In this paper has been investigated a solar collector with squared mirrors. This kind of mirrors have been recently developed and used on the Cassegrain telescope configuration.

It has been computed a starting optical design with traditional circular mirrors and its layout has been simulated to calibrate a first optical setup.

The shape of the mirrors has been then modified into square ones to enhance the collected surface and the fill factor over the radiated area.

This last configuration has been finally optimized placing the housing of the photovoltaic cell behind the primary mirror to allow a simpler cooling together with an easily housing of the transducer.

The final module presents a RMS radius spot of about 5.5mm for the in axis field and 38.47mm for the other ones. The present aberrations in the off axis fields can be considered negligible for a non imaging device.

In the table 2 are collected the geometrical characteristics of the optimized solar collector: although the surface of the secondary mirror is hyperbolic, the radii with respect to the dimensions and the distances between the reflective surfaces allow an easily manufacturing of both surfaces which can be also assembled exploiting the edges with a minimal mounting.

Hence the computed optical module can be considered a starting point for design a high performance solar collector.

Further studies will focus on the optical tolerances of the system and termo-mechanical analyses of the prototype.

References

1. Hyunjoo, H., Jeong, T.K.: Application of high-density daylight for indoor illumination. Energy 35(6), 2654–2666 (2010)
2. Shouli, J., Peng, H., Songping, M., Zeshao, C.: Optical modelling for a two-stage parabolic trough concentrating photovoltaic/thermal system using spectral beam splitting technology. Solar Energy Materials & Solar Cells 94, 1686–1696 (2010)
3. Swanson, R.M.: The promise of concentrators. Progress in Photovoltaics 8(1), 93–111 (2000)
4. Luque, A., Sala, G., Luque-Heredia, I.: Photovoltaic concentration at the onset of its commercial deployment. Progress in Photovoltaics 14(5), 413–428 (2006)
5. Allen Jong-Woei, W., Yi-Yung, C., Bo-Yi, W.: Innovative design of Cassegrain solar concentrator system for indoor illumination utilizing chromatic aberration to filter out ultraviolet and infrared in sunlight. Sol. Energy 83, 1115–1122 (2009)
6. Jia-hong, H., Wun-Ciang, F., Wei-Chi, H., Jui-che, T.: Solar concentrator constructed with a circular prism array. Applied optics 49, 4472–4478 (2010)
7. Duerr, F., Meuret, Y., Thienpont, H.: Miniaturization of Fresnel lenses for solar concentration: a quantitative investigation. Applied Optics 49, 2339–2346 (2010)
8. Panagiotis, K., Eleni, C., Vijay, M.: Nonimaging lenses for solar concentration. Applied Optics 49, 5183–5191 (2010)
9. Mauk, C.E., Prengle, H.W., Eddy, C.H.S.: Optical and thermal analysis of a Cassegrainian solar concentrator. Sol. Energy 23, 157–167 (1979)
10. Feuermann, D., Gordin, J.M., Ries, H.: High-flux solar concentration with imaging designs. Sol. Energy 65, 83–89 (1999)
11. Chen, C.F., Chih-Hao, L., Huang-Tzung, J.: A solar concentrator with two reflection mirrors designed by using a ray tracing method. Optik - International Journal for Light and Electron Optics 121, 1042–1051 (2010)
12. Feuermann, D., Gordin, J.M.: High-concentration photovoltaic designs based on miniature parabolic dishes. Sol. Energy 70, 423–430 (2001)
13. Schroeder, D.J.: Astronomical Optics. Academic Press, London (2006)
14. Zemax User Manual, ZEMAX Development Corporation (2010)
15. Geary, J.M.: Introduction to lens design. Willmann-Bell (2002)

Plant for the Production of Chips and Pellet: Technical and Economic Aspects of an Case Study in the Central Italy

Danilo Monarca, Massimo Cecchini, and Andrea Colantoni

Department of Geology and Mechanical, Naturalistic and Hydraulic Engineering for the Territory, University of Tuscia, Via S. Camillo De Lellis, s.n.c., 01100 Viterbo, Italy
ergolab@unitus.it

Abstract. This work is the result of a technical and economic analysis about the process of transformation of forest wood in biofuels, pellets and wood chips. The experimental pilot plant is managed by a forest consortium located in the province of Terni, Umbria Region (central Italy), near the Mountain Community "Valle del Nera" and Mountain San Pancrazio. The structure and all the machinery of the consortium have been realized utilizing government incentives and involving public and private subjects that operate in the district. The production process of woodpellet is more articulated: it is necessary to refine the mixture a bit more, creating a wood dust which is immediately compressed to form the classical cylindrical pellet's shape and to provide a uniform dough-like mass. The study of the productivity of machinery employed in the transformation of the wood material was made analyzing the cards compiled daily from the technicians of the consortium. The economic evaluation has been completed following an analytical procedure considering the hourly manpower cost and the single machines utilized in all the process phases.

Keywords: chips and pellet, biomass, costs, wood materials.

1 Introduction

The consortium examined is located near the village of Arrone, in Terni administrative department: it is a favorable position for harvesting and transforming vegetable biomass, due to the presence of forests, pine groves, coppice stands, olives.

Characterized as strategic point in order to pull down the handling costs for the material, the consortium is situated not far from the places of withdrawal of forest biomass and near one of the most important traffic routes of the territory. It was promoted from the Mountain Community "Valle del Nera and S. Pancrazio Mount" and is constituted by the same agency with the active participation of the Municipalities of Montefranco, Arrone, Polino, Stroncone, Ferentillo, and of the agricultural and forest enterprises that operate in the area. It pursues the following objectives:

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 296–306, 2011.

- management of the forest properties conferred from the associates;
- management of the wood - energy row, regarding supplying of the material in entrance, its transformation in biofuel and its commercialization;
- advisings about the administrative management of the properties conferred from the associates;
- advisings for associates and enterprises interested to join to the project.

The consortium was constituted in order to value and to manage the agro-forest resources through the transformation of the same ones in biofuels, standardized chips and pellet, with the dimension and the calorific value for domestic and industrial heating. It is important to know the processes through which the crude biomass is converted in better usable and commercializable products in order to improve qualitative and quantitative level of production. In fact the employment of the renewable sources only turns out sustainable and convenient when the costs supported for supplying and transformation turn out inferior to the revenues obtained from their sale. It is therefore of fundamental importance to understand and to analyze the transformation process, with the aim to find and to propose valid solutions in order to take advantage of renewable energy from biomasses, with the maximum yield.

2 Material and Methods

2.1 Description of Process

Supplying: Thanks to the management of supplying of the raw materials and to contracts stipulated with the associated suppliers of lumber, the consortium guarantees, during the year, a constant contribution of vegetable biomass which is composed mainly from the following species: Pinus halepensis, Pinus pinea, Populus nigra, Fagus sylvatica, Quercus ilex, Quercus pubescens and many pruning residuals of Olea europea and street trees.

The suppliers directly unload the lumber in the great large storage square (ca. 1 ha) near the transformation shed. The material is stacked using a hydraulic loader (model Dalla Bona AS 410) set in action from a tractor Mc Cormik C85 MAX. Dimensions of stacks are15 x 2,5 x 3 meters. These measures are nearly a forced choice, because all the suppliers of the consortium work this type of lumber cutting and logging it at a length of 1 m. As soon as completed, these stacks are quickly covered with burlaps preventing in this way the rain water bathing the wood. All the material is held in stack for approximately three months with the aim of a humidity reduction.

Production of chips: The working process begins transferring the wooden material from the large square to the transformation shed. This phase is carried out loading a towing (with hydraulic loader AS 410) which comes immediately transported and unloaded, thanks to an old tractor Carraro 48,4, in the appropriate space for the feeding of the chipper. The chipper is a Pezzolato model PTH 700/660. The feeding system of the machine is constituted of a small hydraulic loader (model Dalla Bona AS 31) which is set in action from an electrical engine of 4 kW. This element is fixed

to earth in a precise point, in lateral position regarding the feeding system of the machine so that the operator has a lateral vision of the process and can take part timely in order to carry out eventual manual regulations.

Chips are pushed from a screw towards the back fan and expelled through the 360° adjustable drainage conveyor. The next machine along the chipping line is the grinding machine MAC 1300. Such machine has a loading hopper in the upper side, where the material expelled from the chipper is directly introduced (thanks to a hydraulic jack) and is crushed against a trituration organ made by a tooth cylinder (that works at a low speed) which grinds the inserted product. In the lower part of the cylinder there is a pierced sheet grill which has the scope to make homogenous the size of the product.

The material exits from the holes of the sheet and falls directly in a hopper that bring it into a screw which transfer the chips to the winnowing silos. This is a great container, equipped to its inside of a three plans vibrating sieve, that realizes product sorting in function of the product destination. In fact the rougher part of chipped material and powder, that constitute the working refuse, are aspired and sent into a container.

The container is transferred afterwards through an elevator OM model XD 25 and it is emptied in a covered greenhouse, used for the storage of the material for the production of the pellet. The scales of intermediate dimensions fall, thanks to gravity, on the bottom of the silos where a mechanical agitator is installed and allows the screws to fish all the material in order to subdivide it in chips to assign to the wrapping in big-bag, in single bags and loose. This last one, like the working refuse, is transported with the elevator in the greenhouse for storage and successively introduced in the pellet line.

The bags of chips are wrapped using a composed machine from a first hopper that is filled up from the material unloaded directly from the screw of silos; a vibrating slab has the task to fill up a second hopper which is connected to an electric balance. The machine creates equal doses based on the inserted parameter. In order to finish such operation, the full bag is closed in the upper extremity through a thermical welding machine, which is situated beside the drainage hopper. Than the bag is put on a pallet. The final product is wrapped with an extensible film and at last fixed with strip, to assure stability during the transport. Using the elevator, the confections in bag or in big-bag are stored in a warehouse, ready for the sale.

Pellet production: The phase of pelletizing is a directly connected process to that one of the chipping, in fact the pellet is the product obtained from an ulterior working of the chips. The material is kept from the greenhouse and unloaded by means of bobcat model 242 B in silos nr. 1, equipped of an inferior mixer that allows to the screw to transfer the material in the continuous dryer, located inside the shed.

The production of pellet is facilitated if the product has a humidity in the dry state comprised in range 12-13% [1] [2]. It is necessary therefore to subject the product to a process of artificial drying that in this system is carried out in the continuous towel, constituted by a great rotary cylinder (handmade prototype), inside of which there are some shovels that stir the cips continuously. The cylinder rotates horizontally with an axis not aligned, since this position determines the advance of the product taking advantage of a game of levels. All the process happens in an atmosphere heated by

electric resistances sited longitudinally to the structure. The watery vapor that manifest as a result of the loss of humidity is aspired and transferred in the culvert of drainage from a cyclone with forced air.

The material, after the process, is transported by means of a system of screws to silos nr 2 that is found in the upside part of the hammerings refiner. In fact, when the phase of drying is finished, the material needs an ulterior preparation before being definitively transformed. Through a hammerings mill the dimensions of chips are reduced to obtain a final size of 3 - 5 millimeter This machine is composed from a cut organ that turns at high speed inside of a circular structure composed by a pierced sheet. This operating phase happens under the action of a forced air cyclone, which has the task to unload the refined material.

Now the product is ready for being transformed definitively in pellet through a specific machine (General Dies), which has the ability to transform the material from the form disintegrated to the compact form. The loading of such machine happens through a complex (feeder + air conditioner + forced feeding) located in the upper part of the same one.

The material comes down directly from the silos to the screw feeder, with a modular spin, which allows to vary the material amount that is wanted to be transformed. From the feeder the product comes down in the air conditioner, in which the modification of the humidity of the material can happen through a tap for the water. This passage is carried out above all when the product to press turns out much dry, episode never happened during the period of observation.

The last stage of the phase of loading happens through the forced feeding, that is an impeller that push violently the material inside the draw-plate, where happens the real phase of pressing. The draw-plate is a circular steel element with holes of 6 millimeter disposed in the center, inside of which are some tooth rollers.

The rollers remain fixed and the draw-plate rotates thanks to the action of a powerful electrical propeller with a maximum power of 110 kW. The rollers are lubricated through a specific pump. Moreover is present a pump and a closed circuit for oil necessary for the lubrication of the bearings of the main axis of machine. A radiator cooled from an impeller lowers the temperature of the oil.

The rollers are located in contact with the draw-plate, so the material that interposes between they is pressed. This continuous action of these elements generates also the successive expulsion of the little cylinders of pellet from the opposite part of the draw-plate, where there are some adjustable knives that cut the product in standard dimensions [3]. Such process provokes an excessive increase of temperature in relation to the great pressure that is developed; consequently the product turns out warm excessive for being able to be sacked. The problem is resolved for means of a cooling machine disposed after the press. The finished product comes in a hopper and is carried, through a screw, in a cups elevator; it transfers than it vertically introducing it in it the cooler, that it is a container subordinate to aspiration, for means of a cyclone to forced air. Inside of the cooler the material temperature comes down fastly and once caught up an established level comes unloaded thanks to a moving grid leaves that it to fall directly on a vibrating sieve. This sieve is used to clean up the pellet from the little bits of powder that still remain. After this activity the material is transported again vertically in the last silos

of collection, which is set over to sacking unit. The sacking unit is semi-automatic, so the presence of an operator is necessary. Each bag is situated over a balance, and is filled up from a flow of materials till the attainment of fixed weight. The bags, once filled up, are closed trough a thermichal welding machine, which is located beside the balance; then the bags are disposed over a pallet. When the decided weight is achieved, the pallet is wrapped by an extensible film fixed by strips. Now the product is definitively ended, and ready for being stored in warehouse before the commercialization.

2.2 Methods of Analysis

In the examined system the estimation of the medium productivity of machines was made during the year 2007, on the base of the analysis of the daily production cards that are filled in by the person in charge in the yard at the end of every working turn. In these documents, the quantitative of real production and the effective working hours for every production line are reported.

They also reports the loading of the crude material stored in the large square, the storing of the finished product in warehouse, including pauses, time and motivation. The same analysis of the production sheets has concurred to the quantitative definition of the material worked in the unit time along the two lines of transformation (table 1) and the determination of the annual use that is a fundamental parameter to the aims of the calculation of the operating costs. In this way it has been able to state that the system annually works 45.000 tons of raw materials, which are transformed in three main products in order to answer to the various requirements of the customers (table 2).

Table 1. Medium productivity of chips and pellet lines

Line	Average-production (t/h)
Chips line (loose or in big-bag)	3,55
Chips line (single bag)	3,10
Pellet line (single bag)	2,20

Table 2. Tons of raw materials worked annually and division in assortments

Raw Materials		45.000 (t/year)
Obtained assortment	Chips in big-bag	300 (t/ year)
	Chips in bag	700 (t/ year)
	Pellet in bag	3.500 (t/ year)

The system operating costs have been calculated for every machine, following the several productive passages in every single machinery. Without an official protocol, the cost analysis was done considering the main methods proposals in bibliography from several authors [4] [5], bringing opportune subjective modifications to parameters and coefficients. Moreover the parameters of residual value, economic duration and coefficients of repair and maintenance inherent machines used for chips production and pressing, has been obtained through the direct and professional experience of the suppliers. The lubricating and fuel burnup of machines used for the transport of lumber has been found directly during the refueling operations. The administrative office of the consortium has given us the costs of purchase of the machines, thus like the price corresponded to the associates for supplying of raw materials (50 €/t) and the cost for workers (12,5 €/h). Considering that along the two lines electric power supply is widely used, for the hour cost analysis of each machine it has been of fundamental importance to know the maximum power of motors of the single machinery that composes the plant. The first transformation line needs of electricity from the loading of the lumber in the chipper till the packaging of the material. Machines used in this operation demand the power showed in table 3.

Table 3. tons of raw materials worked annually and division in assortments

Chips line	Max. power (kWe)
Oil pump and hydraulic arm	4
Chipper	90
Ginding machine	55
Oil system of grinding machine	4
Silos screw feeder	1,5
Vibrating sieve	0,4
Silos mixer	1,5
1st screw	1,5
2nd screw	1,5
3rd screw	1,5
Total	**156,9**

Regarding the transformation of the pellet, it begins with the external loading of the disintegrated wood and finishes with the packing of the product (table 4).

For the operation of the plant, the compressed air and aspiration systems are also set in action electrically (table 5).

Tables 3-4-5 show the value of the maximum power demanded from the electric motors in every transformation line. In order to estimate the percentage of use of the maximal electrical power of motors, an amperometric probe has been used. To know the data of amperometric absorption of every motor, allows to know the electrical power really used by them. As an example, for the motor of the press it has been

found that in conditions of maximum production it absorbs approximately 195 A. As the three-phase tension is 380 V the real electrical power absorption is 74,1 kW. As the maximum power of this motor is 110 kW, we can say that the motor itself absorbe approximately 2/3 of the maximum power in working conditions. This procedure is repeated for all the members of the production lines, having stated that for all the users were worth the aforesaid relation. To be more certain and to have further confirmation of such correspondence another test has been done: knowing the total maximum power of the full line, and using the amperometric probe for measuring the current absorption of the full line. The result is that the power absorption of the full line is equal to the 2/3 of the total maximum power. The chips line uses approximately 105 kW, while the pellet line demands approximately 152 kW.

Table 4. Electrical power required (pellet line)

Pellet line	**Max. power (kWe)**
Ext. Silos mixer	1
Ext.-Int. Screw	1,5
Feeding silos mixer	0,75
Star valve	1,5
Hammerings refiner	75
Air forced cyclone	5,5
Radial valve (under cyclone)	5
Silos screw feeder	1,5
Silos mixer	2,2
Screw feeder	2,2
Air conditioner	5,5
Forced feeding	0,75
Press	110
Screw for elevator	1,5
Elevator I	1,5
Screw for cooler	1,1
Air forced cyclone (cooler)	7,5
Radial valve (cyclone)	1,1
1st sieve motor	0,2
2nd sieve motor	0,2
Elevator II	1,5
Oil lubrication group	0,25
Grease lubrication group	0,18
Total	**227,43**

Table 5. Electrical power required (pellet line)

System of aspiration and compressed air	Max. power (kWe)
Main aspiration	18,5
Screw feeder	1,5
Outlet radial valve	1,5
Compressor equipment	11
Total	**32,5**

3 Results

Table 6 shows the synthesis of operating costs of every single working phase turning out from the technical-economic analysis of the machine's parameters.

Adding the costs of the working phases of each production line, the total hour costs for every manufactured product are obtained (table 7).

To determine the real production cost of a ton of material, taking into account the specific productivity that characterizes every working line, the unit costs in weight have been also defined (table 7).

Table 6. Hour cost of the single phases (comprehensive of 10 workers) (net of Vat)

Nr	Element	Hourly cost [€/h]
1	Stacking and loading of the trailer with hydraulic arm and tractor Mc Cormik	27,04
2	Transport of the wood from the large storage square to the shed with tractor Carraro	28,62
3	Physical transformation of the material by means of trituration line (chipper, crushing, silos-sieve, 1 round screw + 2 flat)	49,05
4	Packing of the crushed material	13,71
5	Transport of the chips in bags to the warehouse with undercarriage elevator	20,37
6	Transport of the chips loose in greenhouse with BOBCAT	20,37

Table 6. (*continued*)

7	Transport of the chips from the greenhouse to the external silos with BOBCAT	16,66
8	Transformation process by means of the refining line (external silos, hammerings mill + 5 round screws) + drying	17,80
9	Pressing and cooling of the material	27,57
10	Packing of the finished product (pellet)	12,78
11	Transport of the packed pellet to the warehouse with BOBCAT	20,37
12	Cost of the aspiration system and compressed air module	2,63

Table 7. total hour cost and unitary cost in weight for each working line

Working costs (total)	**Hourly cost [€/h]**	**Unit costs in mass[(1)] [€/t]**
Chips in big-bag [Elements 1-2-3-5-12]	127,71	35,97
Chips in bag [Elements 1-2-3-4-5-12]	141,42	45,62
Pellet in bag [Elements 1-2-3-6-7-8-9-10-11-12]	222,89	101,31

Costs for the packing materials are shown in the following prospect (table 8). Considering all the showed voices, the production value of every type of assortment realized in the consortium can be calculated. It takes in consideration the costs supported for the production of every ton, comprehensive of expenses for the raw materials, the labor and the materials for the wrapping and the packing, according to the effective potentialities of the system in the considered period of analysis (table 9).

[1] Determined according to the productivity showed in table 1.

Table 8. Cost of the materials for wrapping and packing (net of Vat)

	Pallet (100 x 120 cm)	Bags with logo for pellet and chips	Big-bag (90 x 90 x 120 cm)	Strip + Extensible film in polyethylene
Nr of pieces	4.500	289.000	300	4.200
Unit costs (€)	3,5	0,12	4	0,60

Table 9. Total cost of the assortments (net of Vat)

	Working costs (€/t)	Pallet (€/t)	Bags (€/t)	Strip + film (€/t)	Raw materials (€/t)	Total (€/t)
Chips in big-bag	35,97	3,5	(2)	-	50	89,47
Chips in bags	45,62	7 (3)	9,6	0,60	50	112,8
Pellet in bags	101,31	3,5	8	0,60	50	163,4

4 Conclusions

To introduce on the market an economic, therefore competitive, product is the objective of every company. All this is possible examineing with attention the technical data regarding times and production costs of the product. Such aspects necessarily go to affect on the decisions and the organizational choices that a company must take. Placing the attention on the final result, therefore on the real production costs of the products, it is thought, according to the analysis realized, that they could be mainly contained. This supposition derives from the ascertainment of the inadequacy of some members of the plant, which are characterized from operating costs extremely elevated that contribute, consequently, to an increase of production costs. As a result of this study we can formulate some proposals for the reduction of the production costs of biofuels. It is possible to carry out improvements regarding the production of the pellet, eliminating some working phases that are extremely useless and uneconomical. It is emphasized the uselessness of the transport of loose chips and of the refuse of winnowing in the greenhouse through the elevator undercarriage and the successive transport of the same ones, by means of the bob-cat, from the greenhouse to the external silos.

Such phases would not have absolutely to be present in a well conceived system, because every single shift of the material involves remarkable energetic cost, above all if machines not conceived in order to carry out such kind of operations are used. These processes could be replaced from specific equipments for the transfer of the disintegrated material, like pneumatic conveyors or simple screws that would increase the speed and the continuity of operation of the transformation line, inducing an increase of the system efficiency.

[2] Loan of big-bags for use.
[3] For a ton two pallets are necessary.

The new systems of transfer could moreover be employed also in order to avoid that along the production line of the pellet are used chips journeyed in the crushing machine and the sieve, operations that remarkably affect the final production cost of the pellet. In fact, the grinding can happen also only in the hammerings mill, that has the ability to work without problems much crude scales.

In order to achieve this objective a silos for chips could be installed with a system of drainage with stellar valve, connected to the chipper and that it would feed the chips line or the pellet line. Such silos must have adapted dimensions, so as to guarantee, to every working turnover, the amount of necessary raw materials. Bringing such modifications, the greenhouse could be used, due to its vicinity to the packing area, as warehouse for pallets of finished product; infact today the product is stored in atmospheres not completely sluices and therefore not suitable to the conservation of a material extremely sensitive to the atmospheric humidity. Other fundamental parameter to which it must paied attention is the continuous search of the maximization of the hour productivity of machines.

Taking in consideration the carried out analysis, it can be asserted that the biomasses can play an important role in the "energetic competition", but such objective can be hit only perfecting and optimizing the performances of the technologies and the processes of transformation of the raw materials, reducing the production costs and obtaining therefore biofuels that can be commercialized at a competitive price.

The authors contribution in this paper can be considered equal.

References

1. Hellrigl, B.: Il pellet: notizie e informazioni sulla produzione e l'impiego. Aula magna N.3 (2003)
2. http://www.generaldies.com
3. CTI SC09, R03/01 Raccomandazione del Comitato Termotecnico Italiano sui biocombustibili solidi: specifiche e raccomandazioni, Milano, p. 54 (2003)
4. Franco, S., Monarca, D.: La valutazione preventiva dei costi di riparazione e manutenzione delle trattrici. Rivista di Ingegneria Agraria, March 1998 pp.138-147 (1998)
5. Miyata, E.S.: Determining fixed and operating costs of logging equipment. North Central Forest Experiment Station, Forest Service, 14 p. USDA. St.Paul (1980)
6. Biondi, P.: Meccanica agraria. Le macchine agricole. UTET, Torino (1999)

Feasibility of the Electric Energy Production through Gasification Processes of Biomass: Technical and Economic Aspects

Danilo Monarca, Massimo Cecchini, Andrea Colantoni, and Alvaro Marucci

Department of Geology and Mechanical, Naturalistic and Hydraulic Engineering for the Territory, University of Tuscia, Via S. Camillo De Lellis, s.n.c., 01100 Viterbo, Italy
ergolab@unitus.it

Abstract. Biomass is one of the main sources for energy production, indeed, due to its chemical and physical peculiarities, it can be used very well in thermo chemical processes such as combustion, pyrolysis and gasification. Furthermore, the considerable variability of bio fuels that can be produced allows the experimentation of new technologies that with a higher performance are able to produce not only heat but also electric power. On the other hand, due to problems in supplying and technology, the main disadvantage of such energetic system consists in its difficult application in a large scale production. The present study aims to highlight the possible use of biomasses in small energetic districts such as Viterbo area: the exploitation of wood and cellulose biomasses are described and the possible application with flowing bed technology is considered. This study wants to technically test the feasibility of a biomass supplied system, taking care of stocks availability and amount, use of gasification process based on flowing bed technology, chemical characteristic of syngas and energy power production.

Keywords: biomass production, gasification, economic aspect.

1 Introduction

The alternative energy sources, are becoming increasingly a global reality, their use becomes necessary to try to meet the increasing demand of energy, trying to reduce the possible environmental impact. To do this it is necessary, as well as develop alternative energy systems of great power, that distribute primary energy supply big cities, you must also try to act in small portions of territory (hence the concept of ecological district). This concept, there are easily applicable in the case of biomass, especially for agro-forestry companies with a high output of residual biomass derived from their work. In particular lignocellulosic biomass are easier to administer, both for their use for the purpose of energy, both for ease of transport in points defined by the plants or use processing, both for the various conversion technologies that are present in the current state (combustion, gasification, pyrolysis, steam explosions).

This study, consider pruning of hazelnut groves, as biofuels, located in the district of Cimini and Sabatini Mountains, in the territory of Viterbo, in central Italy. Each year they provide for a discrete volume of residual biomass. Normally these vegetable

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 307–315, 2011.

waste is disposed of by burning on board the field, this operation completely disadvantage, both from an economic point of view and environment. The use of biomass for energy, presents some difficulties, derived from chemical and physical characteristics of the biomass to be used in the process, from identifying sources of supply in the territory in order to optimize its use, and technology best suited to transformation in the form of primary energy. In particular the latter aspect is closely related to the chemical and physical characteristics (ratio C/N, percentage of moisture present, etc.). The analysis carried out explored aspects relating to the quantities produced in each hectare in the territory of the province of Viterbo, the energy content of residues and their possible use by the gasification process, including whether energy aspects, meaning producible electricity.

2 Material and Methods

2.1 Areas of Study and Surveys in the Field

The testing took place during the period 2004-2007 in some areas located in the province of Viterbo, in particular were considered three different companies located in the municipalities of Ronciglione, Capranica and Caprarola - Bassano Romano. The total area sampled was 48.7 hectares. The parameters business are illustrated in Table 1. Companies examined characteristics on the one hand very similar and the other specific to each. The first, derived mainly as a result of economic and agronomic history of this agricultural district, are: 1.the multi-stem farming system; 2.the form of soil management, consisting to cover grass in a controlled and natural way; 3. the type of pruning carried out (of rehabilitation, of production); 4. adherence to agri-environment (Reg. EEC 1257/99).

Table 1. Winter pruning: characteristics of companies sampled

Companies	Location	Age (years)	Area (ha)	Plants/ha
Mordacchini	Ronciglione	30	17	238
Ginnasi	Caprenica	35	20	333
Guerrieri	Bassano R.	30	1,3	476
Guerrieri	Bassano R.	50	0.6	331
Guerrieri	Caprarola	40	1,0	278
Guerrieri	Caprarola	30	1,5	331
Guerrieri	Caprarola	30	1,2	278
Guerrieri	Bassano R.	50	2,0	278
Guerrieri	Bassano R.	40	1,1	278
Guerrieri	Bassano R.	30	3,0	476

On each plot, has carried out a systematic and randomized sampling, realized by taking along rows and weighing biomass produced by a plant every ten, through a mechanical balance. In doing so were tested 10% of plants of each plot.

After the measurement, loppings have been placed at the centre of the space between the rows, as practice, while the small trunks were stacked on the row.

Regarding loppings, this provision is needed to facilitate their transport on the sidelines of the plot, where then loppings are burned, contrary to the provisions of law.

This displacement happens mechanically using a fork linked to the three points attack of the tractor.

The small trunks are manually loaded on a agricultural trailer and they are used for stoves and fireplaces.

2.2 Laboratory Analysis

A sample of wood pruning was sent to the University of Sassari for the definition of physical-chemical properties of the *Corylus avellana* wood.

It was then determined the gross calorific value, through a bomb Mahler calorimeter, the moisture content and ash content, following the procedure prescribed by ISO 9017. The elementary chemical composition was defined as reported by ASTM D 5373 [1].

Following the provisions of the technical specification of CTI [3], it was also carried out the measure of the apparent density of the chopped wood and through screens it was assessed the flakes size.

2.3 Gasification System

After determining the potential results and the chemical and physical characteristics, the second part of work was to analyse the possible use of the residual biomass of the hazel (o nut-tree), through a thermochemical process: the gasification, this is a process that provides a partial oxidation in the absence of oxygen, with high temperatures (900 - 1000° C, which obtains fuel gas consisting of: H_2, CO, CO_2, H_2O, CH4, C_mH_g (various hydrocarbons) and N_2.

If the oxidizing agent is oxygen the gas produced has average calorific commonly called "synthesis gas" or syngas usable both to burn in a gas turbine or internal combustion engine, and for the methanol, hydrogen, natural gas or methane and ammonia synthesis.

The study of a possible application from a technical and economic point of view was conducted over a multi-reactors gasifier, it has a fluidized bed technology. The multi-reactors gasifier a device which was filed patent application to the Ministry of Industry.

The gasification process provides, in general, reactions with very different characteristics between them.

In traditional gasifiers these stages were physically in a single environment and with identical conditions, so you can only find a compromise but we can not optimize the conditions of process for each reaction.

In multi-reactors gasifier with several reactors, place the various stages in physically distinct environments with the ability to optimize for each phase, the process conditions.

This translates into a better quality of the output gas, in the absence of tar products resulting from an incomplete reaction, and the possibility of using different biomass for composition and size without adversely affect on the quality output gas.

In addition, the reaction can be conducted with temperatures of 800 ° C, which permits to use not particularly expensive materials allowing some significant savings in terms of the finished planting cost. The gasifier is an EDM002_025 model and it has a fluidized bed.

This gasifier separates the three main reactions (pyrolysis, oxidation, reduction) in three different reactors and it is also provided with many filters for cleaning up the syngas, developed by the company Mariani Technology Planting Inc. at Civitacastellana near Viterbo.

Technical characteristics are: thermal power 125 kWt, electrical power of 25 kWe, electrical efficiency of 20%, thermal efficiency 36%, with a ratio kg of biomass/ kWe depending on the characteristics of biomass used. For this type of planting are made different powers, but this study found the size to 25 kWe, the only planting to now achieved.

In this context have analysed the technical and economic aspects in order to assess the benefits of that planting; derivative:

1. disposal of residual biomass;
2. possibility of producing electricity with environmental and economic benefits;
3. more suited planting to farms.

3 Results and Discussions

The results for the amount of biomass detected in the four years of observation (Table 2), show a growing production during the first three years of sampling.

This trend is not respected during the last season where the production of pruning back to fall than the previous year.

This variability is attributable mainly to different intensity with which it has carried out the pruning, modest intensity during the years of the maintenance pruning, more vigorous ones which also played a rehabilitation cut. In the latter case, however, there has been the only increase in production of small trunks while the loppings production remain essentially unchanged.

The values reported in the table show both the mass of biomass with moisture in the field, equivalent to 48.3% (expressed as a percentage of fresh weight), both with humidity of 30%.

This last value of the water content allows storage and the immediate use of biomass inside boilers and it contains the losses due by the fermentative process.

Table 2. Average values of achieved biomass (values at harvest-time and at moisture use)

Year of sampling	t/ha moisture = 48,3%	t/ha moisture = 30%
2003/2004	1,5	1,11
2004/2005	2,8	2,07
2005/2006	3,1	2,29
2006/2007	2,2	1,62

Biomass achieved in the territory of Viterbo, with regard to residues of pruning, in terms of potential supply was represented by geographic information systems. In the province of Viterbo, most of the municipalities in the area, presenting a potential availability of residual biomass on pruning between 100 and 1.000 t / s.s. ha^{-1} and between 1.000 to 10.000 t/s.s. ha^{-1}.

The results of elementary analysis of hazelnut hardwood are summarized in Table 3.

To get a quantitative assessment of the thermal energy contained in the hardwood of *Corylus avellana*, we compared the value of lower calorific value of the latter, with the one belonging to some of the main species present inside the Viterbo district, especially Turkey oak (*Quercus cerris*),Downy oak (*Quercus pubescens*), Beech (*Fagus sylvatica*), Chestnut (*Castanea sativa*), Manna-ash (*Fraxinus ornus*).

Table 3. Chemical and physical characteristics of hazelnut wood

Ash content		
C_O	% w. (dry)	2,36
Elementary membership		
C	% w. (MAF)[1]	47,78
H		5,61
N		0,35
O		46,26
Hight heating value	MJ/kg (dry)	
P_{SO}		17,67
Lower heating value	MJ/kg (dry)	
P_{SI}[2]		16,45

It can be seen (Table 4) such as hazelnut wood has a value close to that of species commonly considered good fuels, as Turkey oak, Downy oak and Beech, whose values stand at around 18 MJ / kg.

Table 4. Average values of achieved biomass (values at harvest-time and at moisture use)

	Woody species	**Potere calorifico inferiore (MJ/kg)**
Quercus cerris	(Turkey oak)	18,12
Quercus pubescens	(Downy oak)	18,01
Fagus sylvatica	(Beech)	18,00
Castanea sativa	(Cestnut)	17,92
Corylus avellana	(Hazelnut)	16,45
Fraxinus ornus	(Manna-ash)	15,69

Another energy comparison was made between the hazelnut wood and diesel, taking into account the reduced conversion efficiency of boilers to biomass (Table 5).

[1] MAF = Moisture Ash Free: devoid of moisture and ash.
[2] Calculated according to the equation provided by ISO 9017:1987.

Considering the lower calorific value of Corylus avellana (hazelnut) with a moisture of 30%, equivalent shows that 1 kg of diesel can be replaced by 4,65 kg of hazelnut wood chips.

Table 5. Calculation of equivalence energy diesel / hazelnut chips

Fuels	P.i. (MJ/kg)	P.i. (kcal/kg)	K[3]	Energy Equivalence
Diesel	42,7	10.200	1	1 kg of diesel
Chips, hazelnut wood (moisture = 30%)	10,79	2.580	0,85	4,65 kg of chips

3.1 Technical-Economic Aspects for the Production of Electricity through Residual Biomass Gasification of Hazelnut.

According to the chemical and physical characteristics of hazelnut biomass, described in Table 3, we can determined the input parameters for the production of electricity, described by the following tables (Table 6).

Table 6. Characteristics of biomass and input data for gasification system

Biomass characteristics	
Higher calorific value [MJ/kg]	17,7
Moisture [%]	30%
Lower calorific value [MJ/kg]	12,4
kWh electrical products/ kg of moist biomass	**0,64**
Moles of C/kg of dry biomass	47,8
Moles of O_2/kg of dry biomass	46,26
Moles of H_2/kg of dry biomass	5,61
Moles of C included into gasifier [moli/h]	1.352
Moles of O_2 included into gasifier with the biomass [moli/h]	1.309
Moles of H_2 included into gasifier [moli/h]	158,7
Needs experimental air [Nm^3/kg dry biomass]	2,22
Needs experimental air [Nm^3/h]	62,9
O_2 included into gasifier with air [moli/h]	514

Capacity of biomass needed [t/h]	**0,040**
Ratio biomass/air [(kg/h)/(m/s)]	**15,94**

[3] Coefficient corrective action based on the performance of the boiler.

As the gasification process in an air deficit and difficult to administer, the reactions that characterize it, are fundamental to the balance in the production of syngas.

In fact, a relationship between sub stoichiometric air / fuel unfit, is to shift the balance of reactions, producing harmful substances to the environment (excess of CO2, N2 excess of TAR, CHAR), to this end is crucial regular these aspects settling in advance, through calculations, reports of the various components depending on the chemical and physical characteristics of biomass, in terms of moisture and calorific value.

The moisture content into the biomass is important because the phase which provides for the gasification (pre-heating) and the energy cost of that, it will be all the greater because the moisture content in the biomass. Regarding the investment costs for a multistage gasification system, they are at around € 3.000,00 kWe h-1 product, in Table 7 and Table 8, are defined major items of cost for construction of the system and subheadings related to realization of the site.

Table 7. Investment costs, fixed costs and variable costs and pay-back for the use of the gasification system fuelled by residual biomass (hazelnut pruning)

PLANT 25 kW_e	
System Characteristics	
Installed Power [kW]	25
Hours of operation per year	4000
Percentage of system use	46%
Efficiency	20%
Biomass kg/ kW_eh	0,64
Duration of the system [year]	15
Costs	
Cost of the system [€/kW]	3.000
Total Cost of the system [€]	75.000
Network connection [€]	5.000
Transport and assembly [€]	2.000
Assistance bureaucracy [€]	2.000
Civil works [€]	4.167
Stock biomass [€]	1.250
Site design [€]	2.000
Site putted under rule [€]	2.000
Total investments [€]	93.417
A grant [%]	40%
A net grant [%]	35%
Total investments with funding [€]	60.721
PAY-BACK with incentive	5
PAY-BACK without incentive	7

Table 8. Analysis costs

Variable costs per year	
Total cost of biomass [€/year]	455,00
Cost water disposal [€/year]	195,00
Cost disposal tar [€/year]	780,00
Cost ash disposal [€/year]	64,35
Total cost disposals [€/year]	1.494,35
Maintenance [€ / year]	3.100,00
Cost management CV [€/year]	550,00
Total variable costs [€/year]	**5.144,35**

Fixed costs per year	
Cost depreciation [€/year]	13.239,58
Interest [€/year]	1.721,15
Overheads [€/year]	10.591,67
Users	1.588,75
Total fixed costs [€/year]	**25.552,40**

Value of Production	
Production of energy [€/year]	29.650,00
Energy for consumption[€/year]	22.750,18
Total [€/year]	**52.400,18**

Production Costs	
Purchases of materials	455,00
Costs for staff	15.281,02
Depreciation	13.239,58
Total [€/year]	**28.975,60**
Difference between Value and Production Costs	**23.424,58**

Table 9. Calculation of R.O.I

SYSTEM 25 kWe	
Total Investment **[€]**	93.417
Gross operatine income **[€]**	23.425
R.O.I. Return on Investment [%]	25%

Taking into account of the biomass moisture and of the average energy efficiency of a small (20%) gasification system (defined by the manufacturer), it was calculated potential electricity you can produce, with this type of systems fuelled by waste pruning.

Table 10. Calculation of land required for the production of 100 MW/year

Power installed [kW]	25
Hours of operation year	4000
Percentage of use	46%
Energy potential estimated [MW/years]	**100**
Efficiency [%]	20
Biomass kg/ kWh electrical	0,64
Biomass (pruning) needed for the estimated production of 100 MW [t]	96
Production biomass residual average for the four [t/ha]	1,77
Surface necessary [ha]	**36**

4 Conclusion

Detailed studies have shown that the optimal management strategy of biomass, for the greenhouse effect reduction, depends largely on the efficiency of energy conversion of woody biomass.

One of the biggest problems to use the agro-forestry biomass is clearly the need to set up a notoriously fragmented system .

Projects for forest planning aimed to the activities of forest use organization in space and time for the electricity production in small and medium-sized system, even if based on the use of geographic information systems and expert systems, they are difficult to apply in the Italian reality precisely because of the fragmentation of private property and the frequent absence of forest owners that characterizes the Cimini - Sabatini mountain community.

According to the above developed, the area required for the electricity production, only from pruning waste, from processing hazelnut, amounting to 100 MW (from small system with nominal power of 25 kWe), it is of 36 ha, much lower then that analysed, amounting to 48.7 ha.

Moreover, this power requires not excessive investment, with brief times return on investment. The main drawbacks, of the gasification described by the system, are operating costs resulting greater with the increasing of power system.

The authors contribution in this paper can be considered equal.

References

1. ASTM D 5373, Standard test methods for instrumental determination of carbon, hydrogen, and nitrogen in laboratory samples of coal and coke. ASTM Book of Standars, vol. 05.05 (1993)
2. AA.VV. L'utilizzo del legno come fonte di calore. Regione Piemonte – Ipla Torino, p.84 (2003)
3. Kramer, J.M.: Agricultural biogas Casebook. Great Lakes Regional Biomass Energy Program Council of Great Lakes Governors (2002)
4. Tchobanoglous, G., Theisen, H., Vigil, S.: Integrated Solid Management, Engineering Principles and Management. Mc Graw-Hill Inc, New York (1993)
5. Energy plan of Viterbo city

Soot Emission Modelization of a Diesel Engine from Experimental Data

Enrico Bocci[1,*] and Lorenzo Rambaldi[2]

[1] Marconi University of Rome, Via Virgilio. 8,
00193 Roma, Italy
[2] Sapienza University of Rome, P.za San Pietro in Vincoli. 10,
00184 Roma, Italy
e.bocci@unimarconi.it

Abstract. In recent years particulate diesel engines emission analysis has become crucial. Indeed the particulate modelling and experimental analysis has been modest, mostly because the lack of adequate measurement instruments and the greater focus on the efficiency. Today many instruments can measure accurately the soot emissions and numerous models are in developing. This paper shows the experimental results of a soot emission measurement campaign to predict the soot emission based on simple motor data. Different series of tests has been performed at constant speed and variable load to estimate the amount of particulate matter emitted from diesel vehicles based only on the parameters provided by the motor vehicle electronics. Making a simple model based only on the engine load parameter it has been possible to have model cumulative values very coincident to those experimentally determined. However the relevant difference between the instant model values and those experimental determined, highlights that other more complex models have to be made in order to predict correctly the instant particulate emissions.

Keywords: Diesel Particulate Matter, soot.

1 Introduction

Today's diesel engines are predominantly direct-injection, four-stroke engines with electronic controls that offer high efficiency, quiet operation, and low emissions. Anyway the pollutants emissions reduction of a diesel engine is still challenging, especially regarding particulate and NOx emissions. The formation of pollutants during the combustion process in a diesel engine is strongly influenced by the non homogeneous condition, due to the variability of the distribution of fuel within the air. The presence of these harmful substances in exhaust gases of diesel engines is of particular concern on transport, contributing substantially to the pollution of urban areas. [1]

The expression "particulate" (synthetically PM = Particulate Matter) indicates the whole solid and liquid particles generated in the combustion process and carried in suspension by the exhaust gases. In the case of diesel engines, the particulate is

* Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 316–327, 2011.

mainly composed by carbonaceous particles on which they are condensed or absorbed high molecular weight organic compounds (ketones, aldehydes, esters, etc.).

Microphotographs analysis have revealed that the basic structure consists of small carbon nuclei which are aggregated in clusters of thousands nuclei of different shapes and sizes, briefly called "particles".[2]

In general, larger ones (average diameter 100 to 200 nm) give the largest contribution to the total mass of diesel particulate, while those of smaller diameters (around 10 nm, also called nanoparticles) are more numerous. These substances are inhaled and largely retained in the lungs, causing direct physiological damage and constituting the vehicle of many harmful substances whose effects are not yet fully known. Especially the nanoparticles seem to be the most dangerous, because it is more difficult for the human immune system expel it from the lungs.[3]

In a more traditional and practical term, these diesel engine exhaust emissions are called "black smoke", as distinct from "white or blue smoke" emitted when the engine is started from cold or at low loads. The latter consists mainly of water vapour condensed combined with droplets of fuel and oil unburned or partially oxidized. It disappears as the load grows and the cylinder walls are heated, allowing the injected fuel to burn properly.

The "black smoke" instead, or particulate matter is emitted during the normal operation of a motor to operating temperature, to a modest extent when it is powered with an air / fuel ratio lean (partial load), but in large quantities when the ratio approaches the stoichiometric (conditions of acceleration or running at full load).[4]

1.1 The Particulate

The particulate is a complex substance formed through a combination of agglomeration, condensation, adsorption, changes to viscosity, and chemical transformations as the exhaust (soot, hydrocarbons, wear metals, oxides of carbon, sulphur, nitrogen, and inorganic oxides) leaves the cylinder and travels along the exhaust pipe. In general the particulate is characterized by the presence of three different types of particle, called nucleation, accumulation, and coarse.

The coarse particles are of various kinds, most of the particles are not emitted directly, but they originate from the other two types. In fact the particles are inclined to lay in the exhaust pipe, join to each other, becoming bigger and fall in the flow of exhaust gases. This process of accumulation and release makes the formation of this type of particle random and unpredictable. Precisely for this reason these particles are poorly studied. Probably consist of a solid core, perhaps slightly more thickened than in other particles, and an outer layer of volatile material.

The particle nucleation are poorly understood, although this situation is changing rapidly. Historically until the mid-nineties, almost nothing was known about it, and this is understandable, since they were below the limits of measurement of most of the instruments. Much of the research, but not all, suggest that the nucleation particles are composed of volatile material. Others suggest that some of these particles are really solid, or at least have a tiny solid nucleus. These issues are carefully studied in many research laboratories at the moment.

The accumulation of particles received by far the most attention. Their most immediate and obvious characteristic is to be formed by a series of smaller primary particles. These primary particles, even if they are not exactly spherical, can be approximated to a spherical shape. The measuring range of these spheres is usually about 20-50 nm [1]. The accumulation of particles have different sizes depending on the number of balls that contain, and not because of the variation in the size of the individual spheres. This assembly of spheres, which form the skeleton of a particle accumulation is an "aggregate" or "agglomeration". The number of spheres that can unite not seem to be subject to any physical law: it varies from tens, hundreds, up to a thousand [4]. There is even a well-defined morphology: there are very elaborate interconnected chains, and clusters of compact highly variable. In some cases, the two ends of a chain of beads can be joined together, forming a necklace structure. All you can really say is that the balls, since they are isolated, have an amazing tendency to aggregate.

Represented schematically in Figure 1 are the three types of particles constituting the particulate matter.

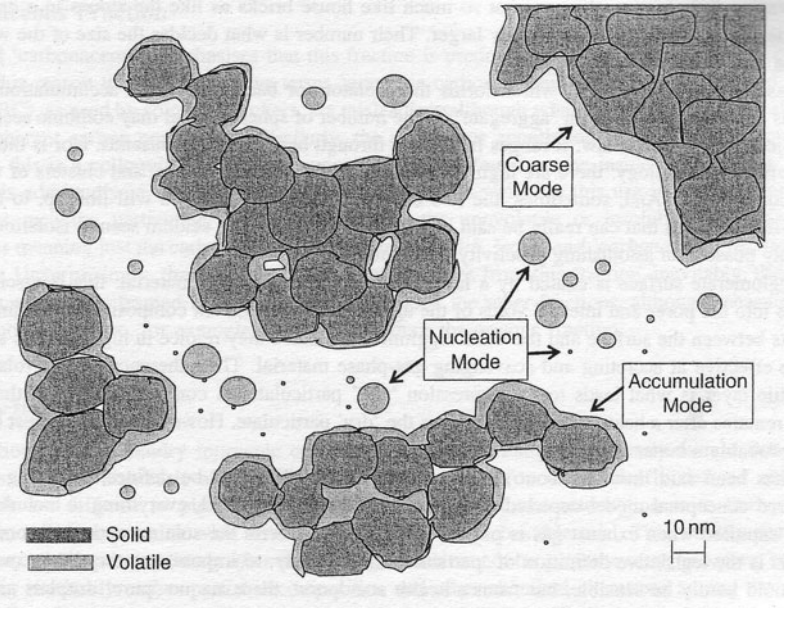


Fig. 1. Schematic illustration of the three groups of particles forming the particulate

The surface of the particles is coated with a layer of liquid or semi-liquid, it is sticky and penetrates into the pores and voids in the interior of the agglomeration. Therefore, there are significant differences in composition between the surface and the nucleus. The adhesion of this layer volatile or semivolatile explains the term "high humidity", as opposed to the solid core that remains after a heating process called

"dry particulates". Little has been said so far about the chemical composition of particulate matter. This can be defined according to the scheme at four levels in Figure 2.

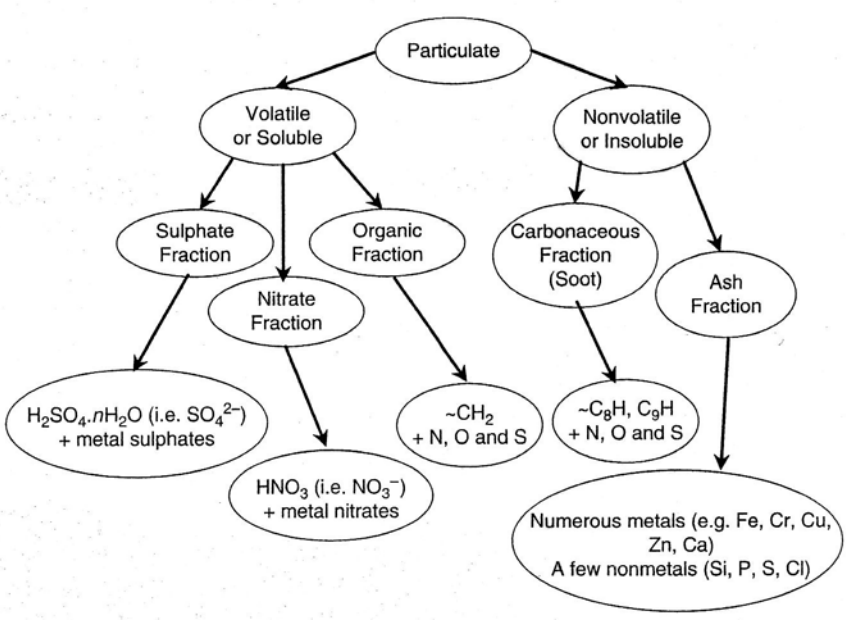


Fig. 2. Schematic illustration of the particulate formation

According to the Figure 2, the first level of particulate includes everything that can be captured when the exhaust gas passes through a filter, with the exception of condensed water and this is the legal definition of "particulate". As mentioned above, during heating, some materials evaporate, some do not, or alternatively, some materials are in solution in certain solvents, some not. This divide between the particles that is volatile and the non-volatile or soluble or insoluble. So, there are five distinct subgroups or fractions: sulphates, nitrates, organic, soot and ash. As we shall see, the chemical composition largely corresponds to the physical representation.

1.2 Model Description

In the framework of the work, a soot model, starting from the parameters acquired through the OBD2 interface, has been developed. The study of the model was carried out on a Fiat Punto 1.4 MJET not equipped with diesel particulate filter (DFP). To calculate the mass of soot emitted using the following formula:

$$m_{soot}\,(kg) = \sum_{i=1}^{n} \frac{C_{soot,i} \cdot \dot{Q}_{asp.,i} \cdot \Delta t_i}{10^6 \cdot \rho_{air}}. \quad (1)$$

Where:

- C_{soot} is the concentration of soot measured by the instrumentation AVL 483 Microsoot Sensor [mg/m^3]
- $\dot{Q}_{asp.}$ is the air flow rate [g/s]
- Δt è range acquisition [s]

In the formula the air flow inlet and not the exhaust is used, because the latter is not present among the parameters that can be acquired through the OBD2 interface. The replacement does not determine, however, a significant error.

The objective of the model is to determine the amount of soot emitted according to the parameters provided by the ECU (Electronic Control Unit). Among the parameters, those that most influence the soot emission and in which we focused attention are:

- Engine load
- Engine speed [rad/s]

As defined by SAE J1979 [5], in the case of a Diesel vehicle, the engine load (CLV) is the ratio, between the flow rate of fuel injected and the maximum flow rate at a certain speed under standard conditions (25 °C and 760 mm Hg).

$$CLV\% = \frac{Actual\_fuel\_injected}{Max\_fuel\_injected@STP(Rpm)\cdot\left(\frac{P.baro}{29.92}\right)\sqrt{\frac{298}{Ambient_T+273}}} \cdot 100 \,. \tag{2}$$

In addition the SAE J1979 states that there is a linear relationship between the engine load and the torque. So it is useful to assess the relationship between the concentration of soot emitted by the load and engine speed.

Figure 3 and Figure 4 shows the trends of the volumetric concentration of soot acquired in a real test in relation with the two engine parameters taken in account.

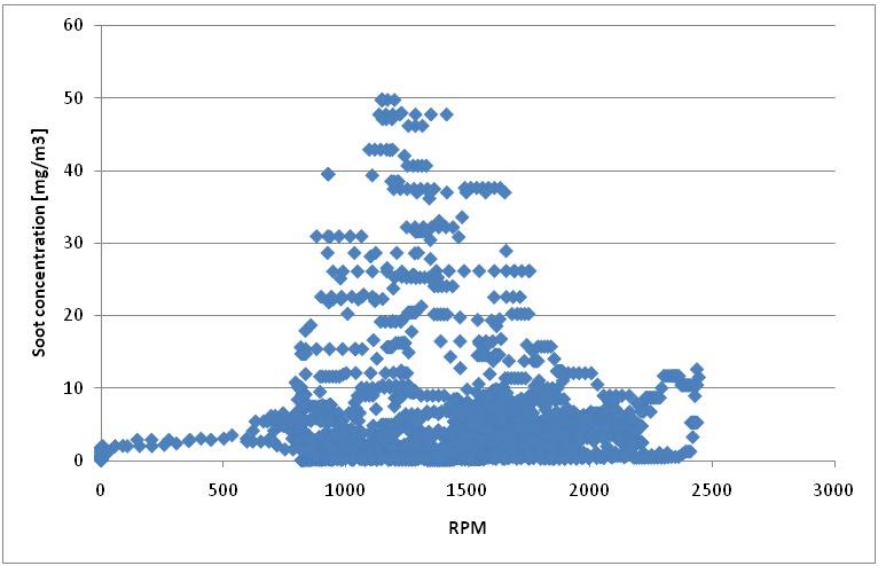


Fig. 3. Soot concentration as a function of the engine speed acquired in a real test

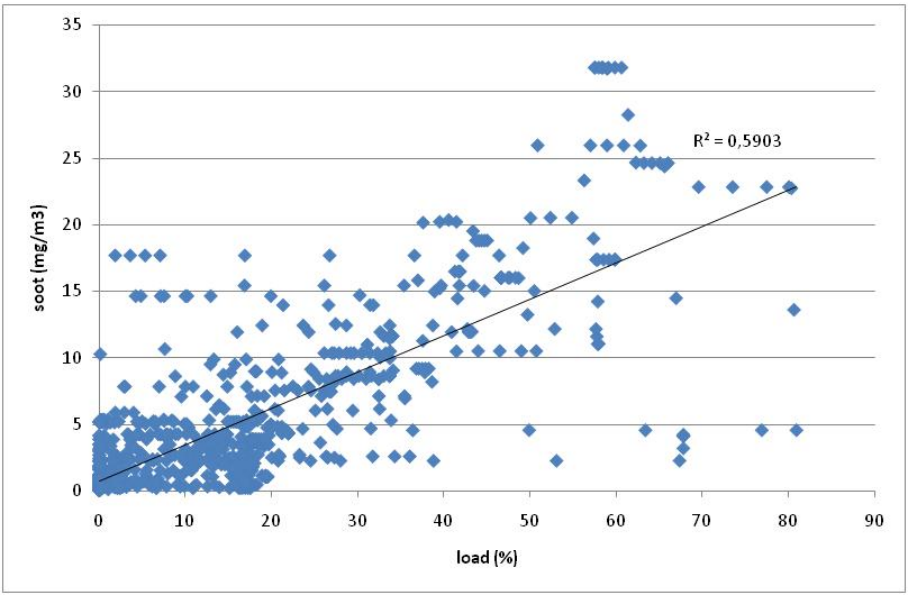


Fig. 4. Soot concentration as a function of the engine load acquired in a real test

There is no a clear correlation with the concentration of soot emitted, but soot concentration and the engine load appears to be in a closer dependence. It is important to underline that the two figures are based on instantaneous data (with some error related to the performance of measuring equipment) and that there is an imperfect alignment between the instantaneous values of load and soot concentration. The latter fact is due to the fact that in this case, the instrument used is the AVL 483 Microsoot Sensor that detects the concentration of soot in the tailpipe of the vehicle where the pollutants are measured with a delay time respect to their production. Since the time taken to travel from the exhaust pipe is variable with the engine speed, it is not possible to realign the data strictly simply shifting back in time the soot concentration measurement.

Finally the engine load has been chosen as the only parameter to base the calculation model of the soot concentration.

$$soot\ (\frac{mg}{m^3}) = A\ + B\ \cdot load \tag{3}$$

However, taking into account the dispersion that characterizes these data and considering that the actual values recorded are afflicted with some uncertainty of measurement, for a more precise identification of the relationship between the two variables is preferable to build the same curve by performing tests in stationary operating conditions.

A series of tests at constant engine speed and variable load have been performed. In particular, in this way the engine load is constant for a certain period of time, these

conditions helps to reduce the uncertainties associated with the quantities recorded by averaging the values of the soot and the engine load during the test period.

Figure 5 shows the values obtained in tests at a constant speed for different values of the load. In this way it is possible to assign the values to the constants A and B in the formula of the used model.

$$soot\ (\frac{mg}{m^3}) = 0.03\ + 0.286\ \cdot load \tag{4}$$

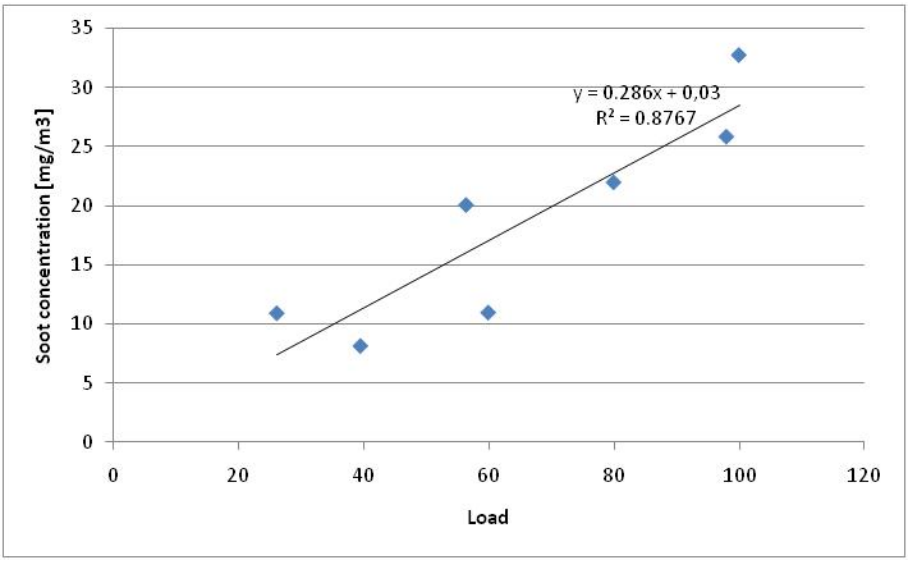


Fig. 5. Soot concentration as a function of the engine load in a constant speed test

1.3 Model Verification

To ensure the accuracy of the model, it has been validated through road tests in real conditions. AVL 483 Micro Soot Sensor [6] has been mounted on the vehicle (Figure 6) in addition to the on-board instrumentation of the vehicle (OBD2).

The tests were carried out on different urban locations, in normal traffic using two different drivers. The duration of the tests varied from 5 to 18 minutes.

In Figures 4.5 and 4.6 are reported the trends of the speed of the two tests under consideration.

In Figures 7 and Figure 8 are reported the drive cycle of the tests used to validate the model. Figure 8 and Table 1 shows as the test 1 was carried out in strictly urban area in the center of Rome, in conditions of high traffic. Average speed is very low and Figure 8 also shows constant stops and starts typical of this condition.

Fig. 6. AVL 483 Micro Soot Sensor mounted on the vehicle for the campaign acquisition

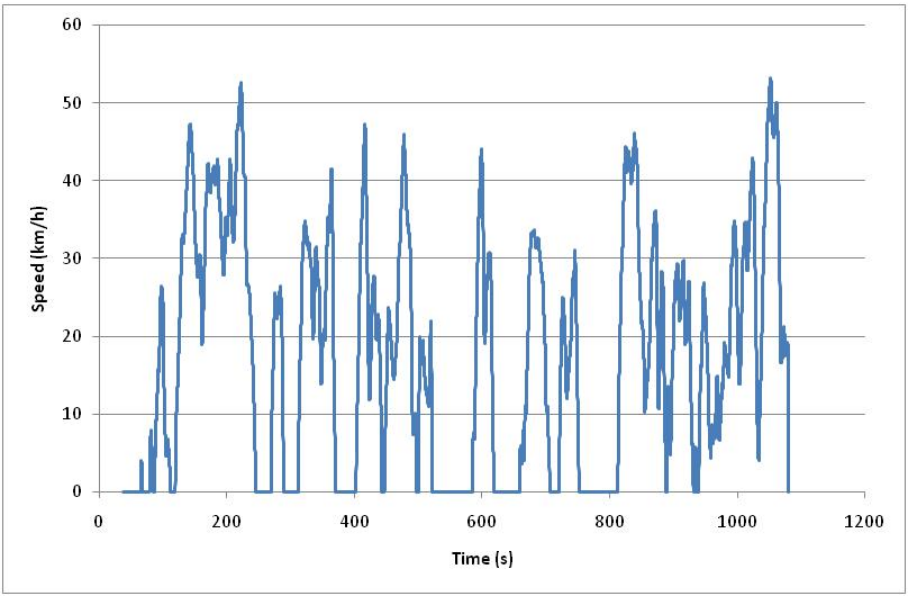


Fig. 7. Drive cycle of the test 1

Test 2 (Figure 9) is instead characterized by higher speeds, it was always carried out in urban areas but in a more peripheral area of the city. Stops and starts due to traffic are still present in this test.

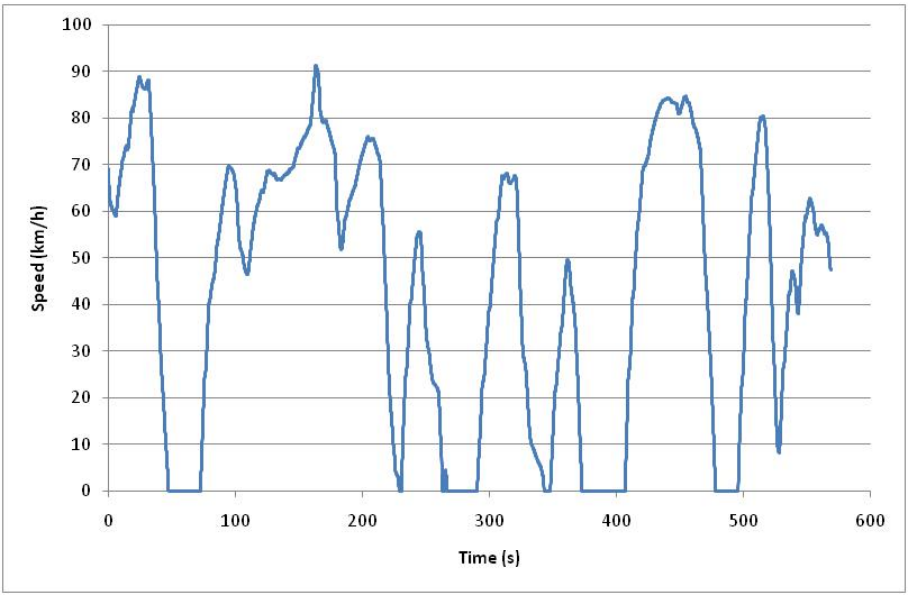


Fig. 8. Drive cycle of the test 2

Table 1. Parameters of the drive cycle

Parameters	Test 1	Test 2
Average speed [km/h]	43.2	16.51
Max speed [km/h]	91.3	53.2

The graphs shown in Figure 9 and Figure 11 are useful to analyze the performance of the model developed. Figure 9 and Figure 11 compare the values calculated by the model with those measured.

From the analytical point of view, it should be noted however, that it is difficult to compare the values measured and those modelled because of their fluctuations, due to the dynamic conditions of detection.

So, to assess the accuracy of the model it is more meaningful to compare the integrated value of the soot emission, checking if during the whole test significant deviations have occurred between the two series (Figure 10 and Figure 12).

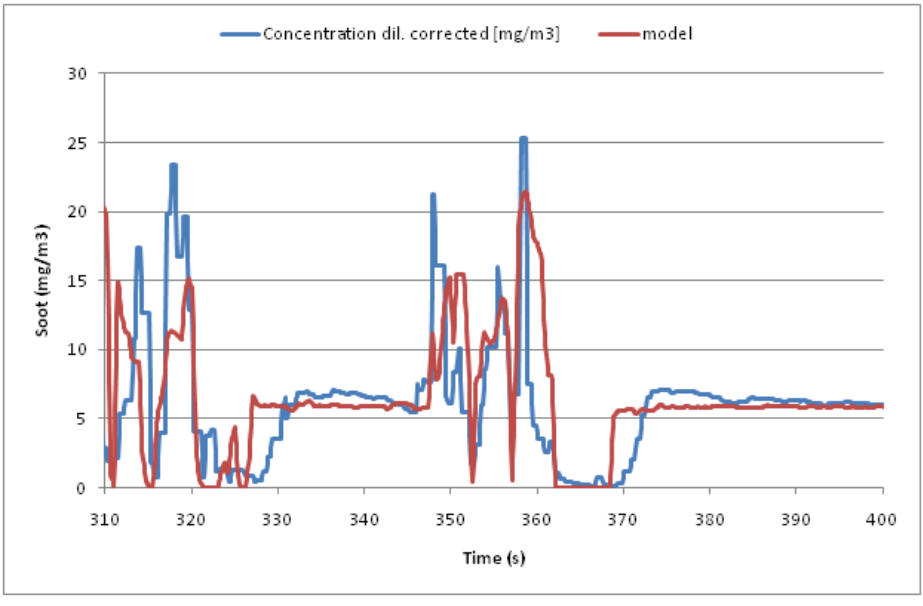


Fig. 9. Soot measured (blue) and soot modelled (red) as a function of time – Test 1

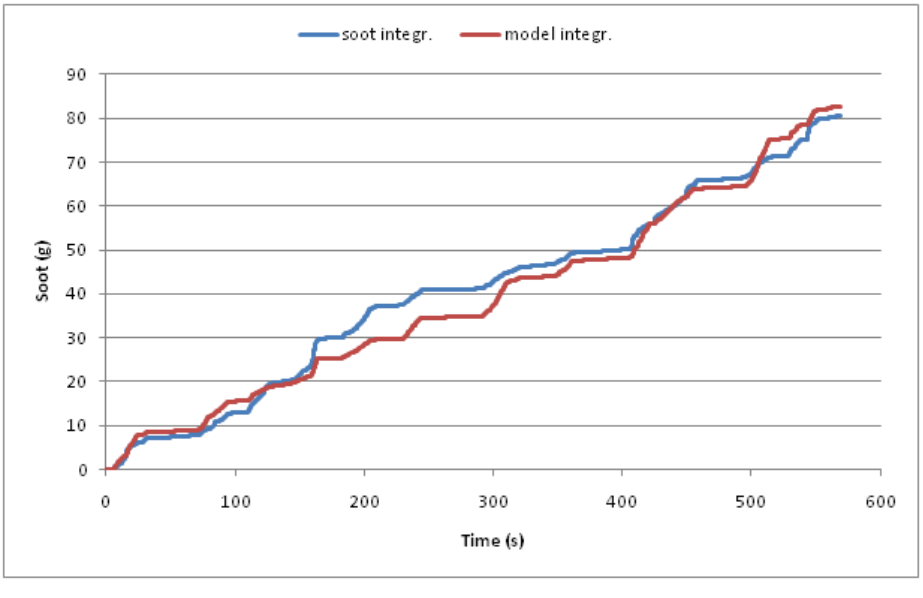


Fig. 10. Integrated soot measured (blue) and integrated soot modelled (red) as a function of time – Test 1

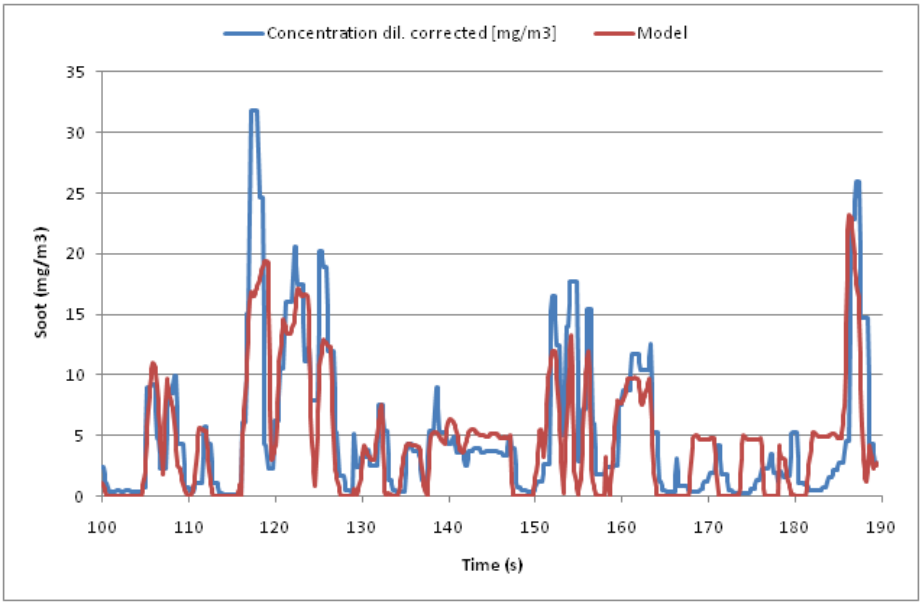


Fig. 11. Soot measured (blue) and soot modelled (red) as a function of time – Test 2

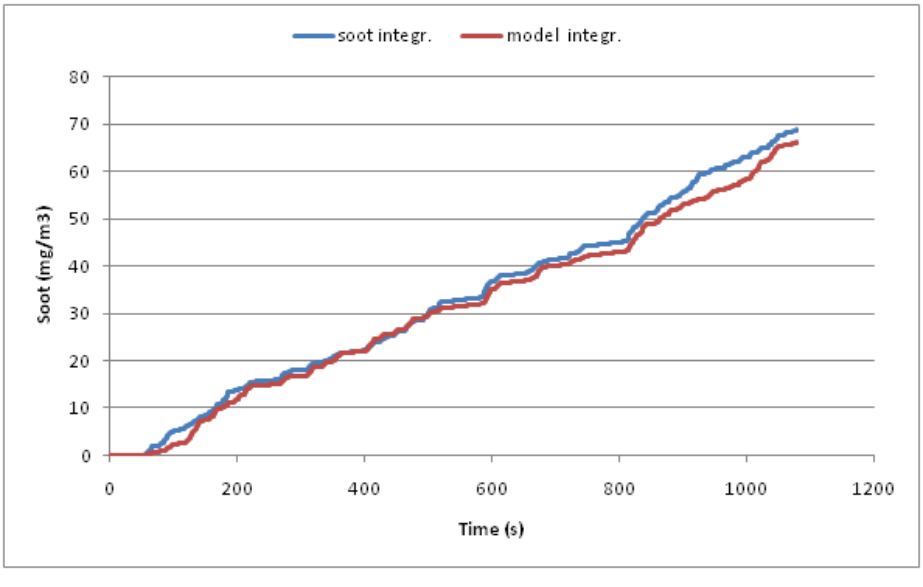


Fig. 12. Integrated soot measured (blue) and integrated soot modelled (red) as a function of time – Test 2

Table 2. Comparison between the model and the measured values

Parameters	Test 1	Test 2
Correlation between values modeled and measured	0.44	0.39
Correlation between integrated values modeled and measured	0.98	0.99
Soot, model (mg)	80.4	68.9
Soot, measured (mg)	82.6	66.2
Error (%)	2.6	4.0

2 Conclusions

The model adopted does not follow accurately the instantaneous values measured of soot concentrations. The correlation coefficients are in fact relatively low (Table 2). However, the model is highly accurate in evaluating the integral emission values with correlation coefficients close to unity and error percentages on the total amount of soot emitted at the end of the tests between 2 and 4%.

Acknowledgments. The author wish to thank their scientific mentors professor Fabio Orecchini, Francesco Filippi and Vincenzo Naso; the researchers Fernando Ortenzi and Adriano Alessandrini; and the thesis student Luca Vercesi for their essential contribution.

References

1. Eastwood, P.: Particulate Emissions from Vehicles. John Wiley & Sons, Ltd, Chichester (2008)
2. Burtscher, H.: Physical characterization of particulate emissions from diesel engines: a review. Aeosol Science 36, 896–932 (2005)
3. Matti Maricq, M.: Chemical characterization of particulate emissions from diesel engines: A review. Aerosol Science 38, 1079–1118 (2007)
4. Walker, A.P.: Controlling particulate emissions from diesel vehicles. Topics in Catalysis 28(1-4) (April 2004)
5. SAE J1979 Revised (April 2002)
6. AVL, AVL Micro Soot Sensor, Operating Manual, Product Guide (September 2009)

Equal Bisectors at a Vertex of a Triangle

R. Losada, T. Recio, and J.L. Valcarce

IES de Pravia, (Asturias, Spain), Universidad de Cantabria, (Santander, Spain), and IES Pontepedriña, (Santiago de Compostela, Spain)

Abstract. Given a triangle ABC, we study the conditions that its vertices must satisfy in order for the internal and external bisectors corresponding to one of the vertices to be equal. We investigate whether there are triangles for which the bisectors at each vertex are equal and other related properties. Automatic Deduction techniques (such as those described in [1]), implemented with CoCoA [2] and the dynamic geometry system GDI ([3], [4]), are used. Moreover, an *ad-hoc* GeoGebra [5] package has been developed (c.f. [6]) to facilitate the exploration of the problem and to improve the analysis and representation of the results in graphical form.

Keywords: Dynamic Geometry; Elementary Geometry; Automatic Deduction; Automatic Discovery; Bisectors.

1 Introduction

It is well known that if a triangle has two internal bisectors of equal length, then the triangle is isosceles (Steiner-Lehmus theorem), and this condition is also sufficient. This theorem was the subject of attention for years, as can be seen, for example, in [7]. The generalization of this result, concerning the equality of internal and external bisectors for two different vertices, was addressed recently in [8], [9] or [10], using computer algebra tools. In [11] another related contribution of the authors can be found, which establishes an open problem concerning the equality of internal or external bisectors for two or three different vertices of a triangle. In that work, automatic deduction tools in geometry, such as those described in [12], in a context of multiple theses where usual techniques are not available, were successfully applied for the first time.

In this new paper, we study the conditions that a triangle must satisfy in order for the internal and external bisectors corresponding to the *same* vertex to be of equal length. Its extension to the case where this property holds simultaneously on several vertices is also considered. We address, as well, some problems on the areas of the so called *bisector triangles*. The combination of three tools (GDI, GeoGebra and CoCoA) in the context of automatic discovery and its application to some challenging problems –if approached in the traditional way– is perhaps the main contribution of this work.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 328–341, 2011.

2 A Short Introduction to Automatic Discovery

Automatic discovery of elementary geometry theorems, although less known than automatic proving is not new. Finding the geometric locus of a point defined through some geometric constraints (say, finding the locus of a point when its projection on the three sides of a given triangle form a triangle of given constant area) can be considered as a task for the "automatic derivation" of properties approach, circa 25 years old.

Although "automatic derivation" (or locus finding) aims to discover some new geometric statements, such as "given this construction and these constraints, point P lies on the curve C", it is not exactly the same as "automatic discovery", that searches for complementary hypotheses for a (perhaps generally false) geometric statement to become true (such as stating that the three feet of the altitudes for a given triangle form an equilateral triangle and finding out what kind of triangles accomplish it). If suitably interpreted (for instance, considering a trivial thesis $0 = 0$ and searching for the conditions to verify it in terms of some specific data, such as the coordinates of point P), automatic discovery tools might as well achieve automatic derivation of properties.

The essential idea behind the different approaches to discovery is, essentially, to consider that the necessary and sufficient conditions that one needs to add to a given collection of hypotheses for a given thesis to hold is... the thesis itself. More precisely, that of adding the conjectural theses to the collection of hypotheses, and then deriving, from the new ideal of theses plus hypotheses, some new constraints in terms of the free parameters ruling the geometric situation. This derivation is achieved by elimination of the remaining (dependent) variables.

For a toy example, consider that $x - a = 0$ is the only hypothesis, that the set of points (x, a) in this hypothesis variety is determined by the value of the parameter a, and that $x = 0$ is the (generally false) thesis. Then we add the thesis to the hypothesis, getting the new ideal $(x - a, x)$, and we observe that the elimination of x in this ideal yields the constraint $a = 0$, which is indeed the extra hypothesis we have to add to the given one $x - a = 0$, in order to have a correct statement $[x - a = 0 \wedge a = 0] \Rightarrow [x = 0]$.

Indeed, things are not so trivial. Consider, for instance, $H \Rightarrow T$, where $H = (a + 1)(a + 2)(b + 1) \subset K[a, b, c]$ and $T = (a + b + 1, c) \subset K[a, b, c]$. Take as parameters $U = \{b, c\}$, a set of $\dim(H)$-variables, independent over H. Then the elimination of the remaining variables over $H + T$ yields $H' = (c, b^3 - b)$. But $H + H' = (a + 1, b, c) \cap (a + 2, b, c) \cap (a + 1, b - 1, c) \cap (a + 2, b - 1, c) \cap (b + 1, c)$ does not imply T, even if we add some non-degeneracy conditions expressed in terms of the free parameters U, since T vanishes over some components, such as $(a+2, b-1, c)$ (and does not vanish over some other ones, such as $(a+1, b-1, c)$).

Bearing these difficulties in mind, an elaborated discovery procedure, with several non trivial examples, is presented in [1]. One interesting example of the power of such automatic discovery protocols, related to the Steiner-Lehmus theorem and accomplishing an original and long time conjectured result, appears in [11].

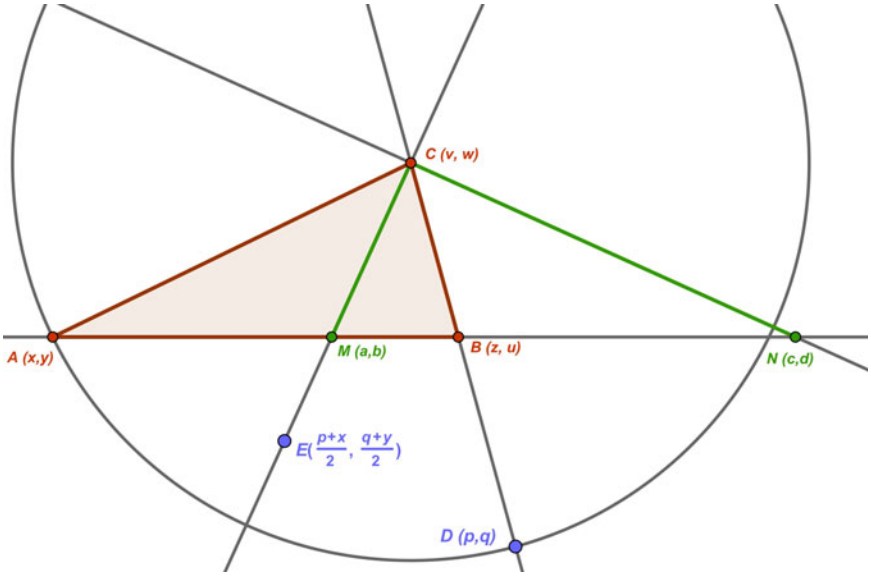


Fig. 1. Construction of the bisectors at vertex C

The discovery method of [1] has been recently revised in [12], showing that, in some precise sense, the idea of considering hypotheses and theses $H + T$ and then eliminating some variables here, is intrinsically the unique way towards discovery. In what follows we will apply, without further details, the methods described in these papers.

3 Equality of Internal and External Bisectors Corresponding to one Vertex

We start by considering the general case of a triangle with vertices *A(x,y)*, *B(z,u)* and *C(v,w)*. First we study the conditions that A, B and C must satisfy for the two bisectors corresponding to a single vertex to be equal. We consider a vertex, say C, and determine the internal bisector and the external bisector corresponding to C. Using automatic deduction techniques, we will try to find necessary and sufficient conditions for these two bisectors to be equal. For this we translate (as done automatically by the program GDI, developed by one of the authors, see ([3], [4]) into algebra the construction of both bisectors as follows [Figure 1]:

(a) The circle with center C and radius CA which intersects side CB in $D(p, q)$ satisfies the properties:

$$\text{distance}(C,D) = \text{distance}(C,A) \text{ and aligned}(C,B,D)$$

$$(v-p)^2 + (w-q)^2 - (v-x)^2 - (w-y)^2 = 0 \tag{1}$$

$$Det(Mat[[v, w, 1], [z, u, 1], [p, q, 1]]) = 0 \tag{2}$$

(b) The midpoint E of the segment DA has coordinates $\frac{p+x}{2}$ and $\frac{q+y}{2}$

(c) The bisector EC intersects the side AB at a point $M(a,b)$, which verifies the properties:

$$\text{aligned}(C,E,M) \text{ and aligned}(A,B,M)$$

$$Det(Mat[[v, w, 1], [\frac{p+x}{2}, \frac{q+y}{2}, 1], [a, b, 1]]) = 0 \tag{3}$$

$$Det(Mat[[x, y, 1], [z, u, 1], [a, b, 1]]) = 0 \tag{4}$$

(d) The square of the length of the bisector CM is

$$(v - a)^2 + (w - b)^2 \tag{5}$$

(e) The algebraic conditions of the other bisector can be obtained similarly, but it is more computationally efficient the following method, based on the fact that the two bisectors are perpendicular and, therefore, the point $N(c,d)$, end of the other bisector, satisfies the following properties:

$$NC \perp CE \text{ and aligned}(A,B,N)$$

$$(c - v)(\frac{p+x}{2} - v) + (d - w)(\frac{q+y}{2} - w) = 0 \tag{6}$$

$$Det(Mat[[x, y, 1], [z, u, 1], [c, d, 1]]) = 0 \tag{7}$$

(f) The square of the length of the bisector CN is:

$$(v - c)^2 + (w - d)^2 \tag{8}$$

We use now the algorithm stated in [1] (and reformulated in [12]) for the discovery of conditions under which a given property is true. We take, as set of hypotheses for the construction, H_c, the polynomials corresponding to those equations described in (1), (2), (3), (4), (6) and (7) and as thesis, T_c, take the difference (5)-(8). We add the thesis to the hypothesis set and it is clear that $H_c + T_c \Rightarrow T_c$. Since we are interested in finding conditions on the free points (vertices) in order for T_c to be satisfied, it suffices to eliminate all variables, except those of the vertices, in the ideal $H_c + T_c$. Using CoCoA we obtain a polynomial, G_C, which factors as follows:

- $F_{1C} = v^2 + w^2 - 2vx + x^2 - 2wy + y^2$, which is a degenerate circle in the variables (v, w) or (x, y), if we set the other pair of coordinates.
- $F_{2C} = (uv - ux + wx - vy - wz + yz)^2$, which is a straight line (on the side AB, AC or BC) in the variables (v, w), or (z, u) or (x, y), respectively.
- $F_{3C} = u^2v^2 + u^3w - u^2w^2 - 3u^2vx + 4uvwx + 2u^2x^2 - v^2x^2 - 3uwx^2 + w^2x^2 + vx^3 - u^3y - 2uv^2y - u^2wy + 2uw^2y + 2uvxy - 4vwxy - ux^2y + wx^2y + 2u^2y^2 + v^2y^2 - uwy^2 - w^2y^2 + vxy^2 - uy^3 + wy^3 + u^2vz - 4uvwz - u^2xz + 2v^2xz + 2uwxz - 2w^2xz - vx^2z - x^3z + 2uvyz + 4vwyz + 2wxyz - 3vy^2z - xy^2z - v^2z^2 + uwz^2 + w^2z^2 - vxz^2 + 2x^2z^2 - uyz^2 - 3wyz^2 + 2y^2z^2 + vz^3 - xz^3$, which is a

hyperbola in (v, w), or a cubic in (x, y) or (z, u), if the other coordinates are fixed.

Analogous conditions can be obtained for the vertices A and B, in terms of some corresponding polynomials G_A and G_B, so that the existence of triangles in which at least one vertex satisfies that its internal and external bisector are equal is provided by $G_A \times G_B \times G_C = 0$. A simple geometric interpretation of this result is achieved, without loss of generality, by taking two vertices in the triangle to be the origin and unit point in the X-axis. Taking $A(0, 0)$ and $B(1, 0)$, we get:

$G_A \times G_B \times G_C = -v^{10}w^6 - 2v^8w^8 + 2v^4w^{12} + v^2w^{14} + 4v^9w^6 + 7v^7w^8 + v^5w^{10} - 3v^3w^{12} - vw^{14} - 6v^8w^6 - 8v^6w^8 - 2w^{14} + 4v^7w^6 + 3v^5w^8 - 2v^3w^{10} - vw^{12} - v6w^6 + v^2w^{10}$ with the following factors and geometric interpretation:

- w^6, condition equivalent to the fact that the point C lies in the line AB.
- $v^2 + w^2$, representing the point A.
- $v^3 + vw^2 - v^2 + w^2$, representing a cubic (a right strophoid, see [14]) describing the locus of C for the bisectors at B to be equal.
- $v^3 + vw^2 - 2v^2 - 2w^2 + v$, representing another cubic (symmetrical to the previous one) describing the locus of C for the bisectors at A to be equal.
- $v^2 - w^2 - v$, representing a hyperbola with vertices A and B, describing the locus of C so that the bisectors of C are equal.

On [Figure 2] we have displayed the different curves representing the three geometric loci of C such that the length of the internal and external bisectors in A, B or C (respectively) coincide. This image was obtained using the dynamic color property in GeoGebra following a numerical algorithm described in [6].

This original method presents several advantages versus the straightforward approach of using the implicit plot features of some mathematical software programs, such as GDI or GeoGebra (the latter includes this feature just in the beta version 4.0, released just in August 2010, while our research on this topic started much earlier). In fact:

– implicit (or parametric) plotting is, quite often, not very reliable (as reported, for instance, in [13], see also [Figure 3]).
– implicit (or parametric) plotting requires the implicit (or parametric) equations of the given curve and this prevents, if the algebraic engine is unable to provide a suitable input, the exploration of the given geometric situation.
– plotting the geometric loci with implicit plotting does not provide additional information –as the color method does– about the behavior of parts of these loci and/or of the different regions of the plane determined by the curves, concerning the given query (in our case, the equality of bisectors).

Let us clarify these points a little, referring the reader to [6] for further details. The dynamic color method displays the curves directly from its definition as loci verifying some properties. Roughly speaking, on [Figure 2], we have considered, at each point P in the plane, the absolute values of this triple of numbers

$$(eA - iA)/(eA + iA), (eB - iB)/(eB + iB), (eC - iC)/(eC + iC)$$

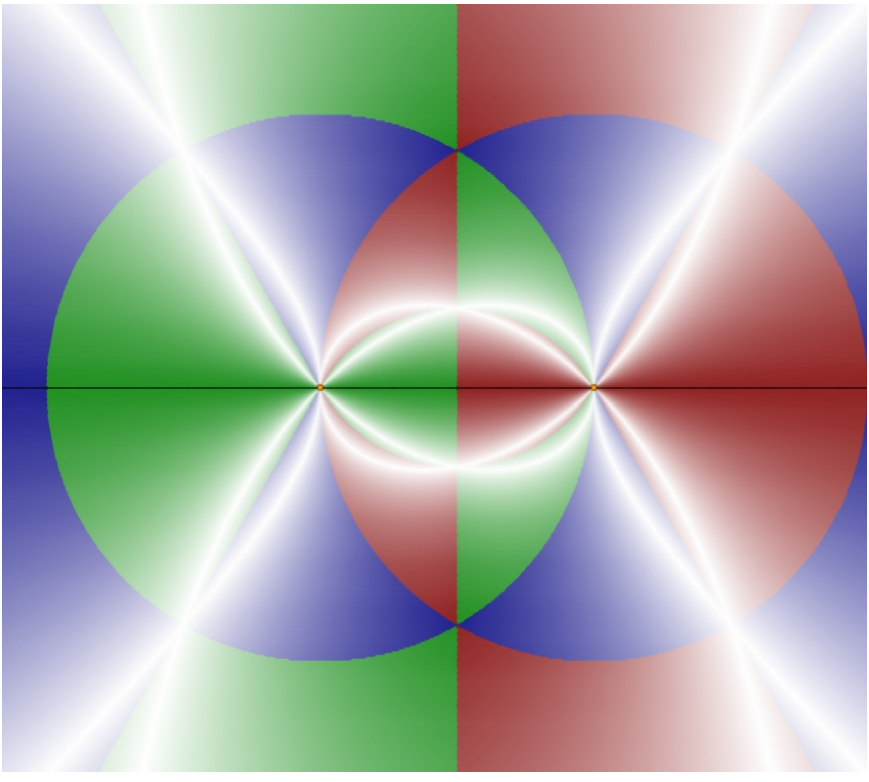


Fig. 2. Graphic result obtained for the problem of equality of internal and external bisectors at each vertex, using the procedure described in [6]. Visit http://geogebra.es/imag/fig2 for a full color figure.

Assuming that vertex C is placed at P, then eA (respectively iA) denotes the length of the external (respect. internal) bisector at A. A similar explanation applies for eB, iB, eC, iC. The three absolute values of $(eA-iA)/(eA+iA), (eB-iB)/(eB+iB), (eC-iC)/(eC+iC)$ are then compared and if the minimum is attained for $(eA-iA)/(eA+iA)$, then a color gradient related to red is displayed (the smaller the minimum, the lighter the color, so white lines correspond to the case of equality of internal and external bisectors). Same for $(eB-iB)/(eB+iB)$ (with a green color gradient) and for $(eC-iC)/(eC+iC)$ (with a blue color gradient). Thus, the red regions in the figure correspond to locations of C such that the relative difference of the lengths for external and internal bisectors at A is smaller than the relative difference for those at vertices B, C; and there are similar interpretations for the blue and green regions.

A similar procedure has been applied to produce [Figure 6], see the caption at this figure.

Notice that, in this automatic way, [Figure 2] displays (without having asked for it) two circles and one vertical line. It is easy to check that they correspond to

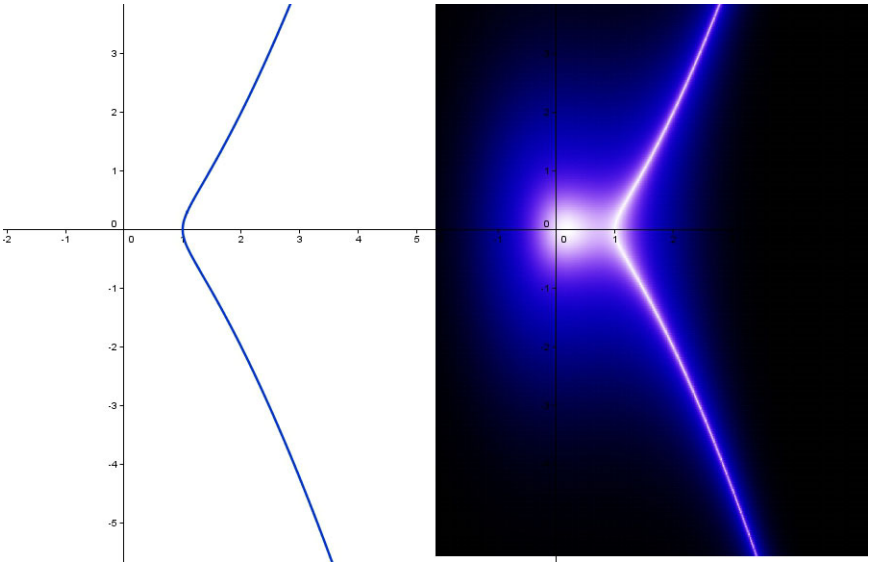


Fig. 3. Implicit plotting of the curve $x^2 + y^2 - x^3$ both with GeoGebra v.4 (left side) and with the color procedure described in [6]. Visit http://geogebra.es/imag/fig3 for a full color figure. Notice that GeoGebra misses point $(0, 0)$ in the curve.

the location of C so that the triangle is isosceles. This is a case which, as we will discover later (see remarks at the end of Section 4) by algebraic means, is quite relevant in this problem. This feature is even more crucial in other contexts, such as the one described in [11], where the searched locus is an irreducible algebraic curve, but such that different branches of this single curve correspond to diverse situations concerning equality of bisectors for different vertices. In this example the dynamic color procedure turns out to be a very singular and useful tool to visually separate the different cases.

On the other hand [Figure 4] shows the solution provided by GDI for the locus set of C such that the bisectors at A have equal length. GDI first discovers the equation from the geometric conditions (using CoCoA) and then draws its graph.

We can generalize this problem by considering, instead of the equality of internal and external bisectors at a point, the case in which the length of the bisectors satisfies a given ratio k. We just have to replace the thesis T_c by $k^2((v-a)^2 + (w-b)^2) - (v-c)^2 - (w-d)^2$ and repeat the above calculations. For instance, the result obtained by specializing $A(0, 0)$ and $C(1, 0)$ and considering the locus of B so that we have a fixed $k-$ ratio length of bisectors at C, yields $G_C = F_{1C} \times F_{2C} \times F_{3C} = 0$, with:

- $F_{1C} = u^3k^2 + uz^2k^2 - 2u^2zk - 2z^3k - 2uzk^2 - u^3 - uz^2 - 2u^2k + 2z^2k + 2uz$
- $F_{2C} = u^3k^2 + uz^2k^2 + 2u^2zk + 2z^3k - 2uzk^2 - u^3 - uz^2 + 2u^2k - 2z^2k + 2uz$
- $F_{3C} = u^2$

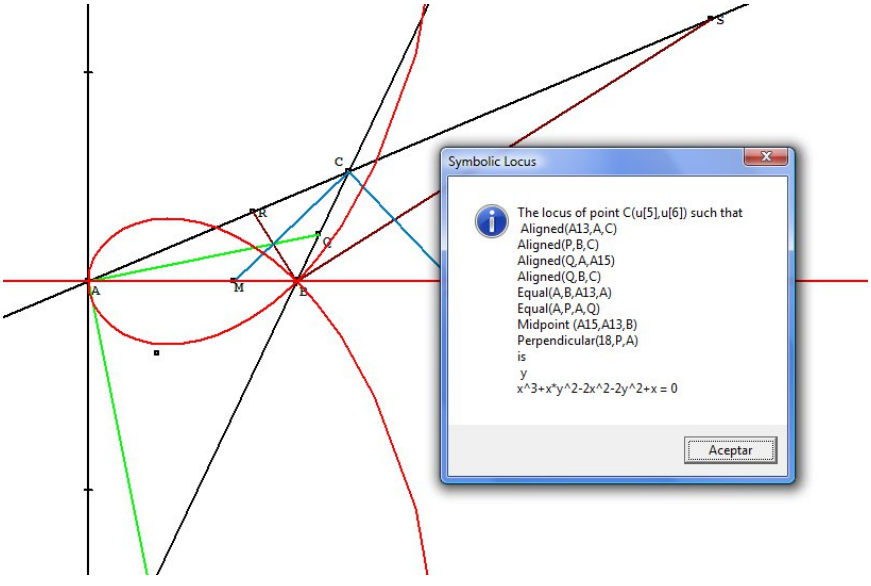


Fig. 4. Graphic and algebraic results obtained for the problem of equality of internal and external bisectors at the vertex A using GDI. Notice that the cubic curve plus a condition of degeneration ($y = 0$) are displayed. The procedure used is described in [4].

4 Equal Internal and External Bisectors at Each Vertex

The question of whether there exist triangles satisfying the equality of bisectors lengths at two and at the three vertices (i.e. such that some pair of equalities among $AQ = AP$, $BR = BS$ and $CM = CN$ hold or, even, if the three of them can simultaneously hold) arises quite naturally. Here P, Q, R, S, M, N are the feet of the corresponding bisectors.

We approach this issue by considering first the case of two vertices and, then, the full three vertices instance. Without loss of generality we assume that $A(0, 0)$ and $B(1, 0)$.

– Equal bisectors at C and equal bisectors at A

To the algebraic description of equal bisectors at C we add the algebraic description of the equality of the bisectors in A (introducing the corresponding variables, and so on, as in the previous Section II: (e, f) are the coordinates of intersection point between the circle with center A and radius 1 and the line AC; $Q = (g, h)$ and $P = (i, j)$), which is given by the polynomials:

$H_a = [e^2 + f^2 - 1,$
$Det(Mat([[v, w, 1], [0, 0, 1], [e, f, 1]])),$
$Det(Mat([[v, w, 1], [g, h, 1], [1, 0, 1]])),$
$Det(Mat([[0, 0, 1], [(e + 1)/2, f/2, 1], [g, h, 1]])),$

$Det(Mat([[i,j,1],[1,0,1],[v,w,1]]))$,
$i*(e+1)/2+j*f/2]$
$T_a = [g^2+h^2-i^2-j^2]$

Eliminating all but the indeterminates v and w, we obtain:

$$w^4 - 1/2vw^2, \quad v^2w^3 - 3/2vw^3, \quad v^3w^2 - 3/2v^2w^2$$

The intersection of these three curves gives the line $w = 0$ (side AB) and the points with coordinates $\left(\frac{3}{2}, \pm\frac{\sqrt{3}}{2}\right)$. We conclude that the only non-degenerate triangles with this property (equality of bisectors lengths at these two vertices, at least) are those with vertices $A(0,0)$, $B(1,0)$, $C\ \left(\frac{3}{2}, \pm\frac{\sqrt{3}}{2}\right)$, see [Figure 5].

It could be interesting to remark that such triangles are strictly isosceles (i.e. not equilateral). In particular, this observation yields the non-existence of triangles with equal pair of bisectors in each of its three vertices. A computational proof of this fact is also easy to obtain, see next item.

Moreover, we know, by the classical Steiner-Lehmus Theorem, that isosceles triangles have two internal bisectors of equal length. And it is easy, by symmetry, to conclude that such triangles have also two external bisectors with equal length (although different, in general, from the common length of the internal bisectors).

Thus, the obtained isosceles triangles (with equal angles at C, A), namely, those with vertices $A(0,0)$, $B(1,0)$, $C\ \left(\frac{3}{2}, \pm\frac{\sqrt{3}}{2}\right)$, must have *all four* bisectors at vertices C, A with equal lengths.

That is, we conclude there are not triangles with two vertices C, A having at each of them equal lengths l_C (respectively, l_A) for the internal and external bisectors of C (respectively, of A), but with l_A different from l_C.

A similar discussion can be carried for the case of equality of bisectors at A, B.

– Equal bisectors at C, equal bisectors at A and equal bisectors at B

It suffices to calculate, for example, which triangles have equal bisectors at B and C and check if any of them is one of the found in the previous case. This calculation leads to:

$$w^5 + 3/2vw^3, \quad v^2w^3 + 1/2vw^3, \quad vw^4 + 1/2w^4, \quad v^3w^2 - v^2w^2 + 1/2w^4$$

whose associated variety is composed of the line $w = 0$ and the points with coordinates $\left(\frac{-1}{2}, \pm\frac{\sqrt{3}}{2}\right)$.

Therefore we conclude that there is no triangle with equal pair of bisectors in each of its three vertices.

5 A Related Problem

In [11] we have determined the conditions for bisectors of different vertices to have the same length. In this paper we have studied the case of equality of bisectors for a single vertex. Considering both contributions it is easy to deduce

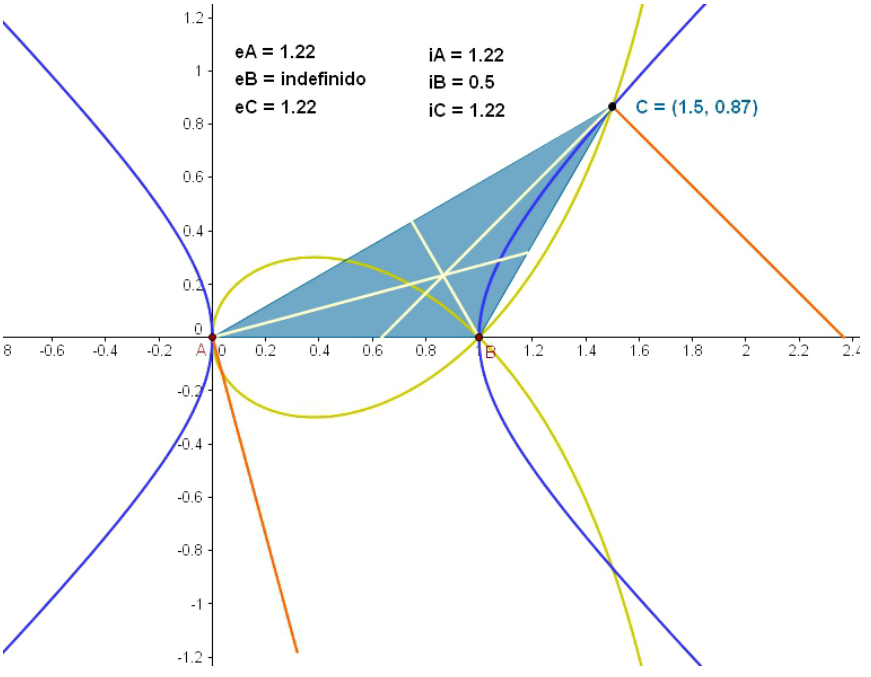


Fig. 5. A triangle with two vertices (C, A) having at each of them internal and external bisectors of equal length. As shown on top of the triangle, it happens that the four lengths are equal to 1.22. Coordinates for C are $\left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right)$.

the case of triangles which could happen to have, at two or three vertices, equal lengths for all involved bisectors. For instance, we already know that the three vertices case is impossible.

More appealing seems the following variant. Let us consider now, for each vertex of a triangle, the new (right) triangle defined by the vertex and the feet of the corresponding internal and external bisectors. Let us name such triangles the *bisector triangle* of a given vertex. Since they are always right triangles, two such triangles with coincident bisector lengths will have the same area. So, we might ask, more generally, about the triangles ABC where at least two such bisector triangles have equal area.

Taking $A(0,0)$ and $B(1,0)$ and being AP, AQ, CM, CN the bisectors corresponding to the vertices A and C, we want to find, for example, the locus set of $C(v,w)$ such that $area(A,P,Q) = area(C,M,N)$. With notations as above we take as thesis T:

$$Det(Mat([[v,w,1],[a,b,1],[c,d,1]])) - Det(Mat([[x,y,1],[i,j,1],[g,h,1]]))$$

and then, in $H_c + H_a + T$, all indeterminates but v and w are eliminated. We summarize now the results obtained following this procedure.

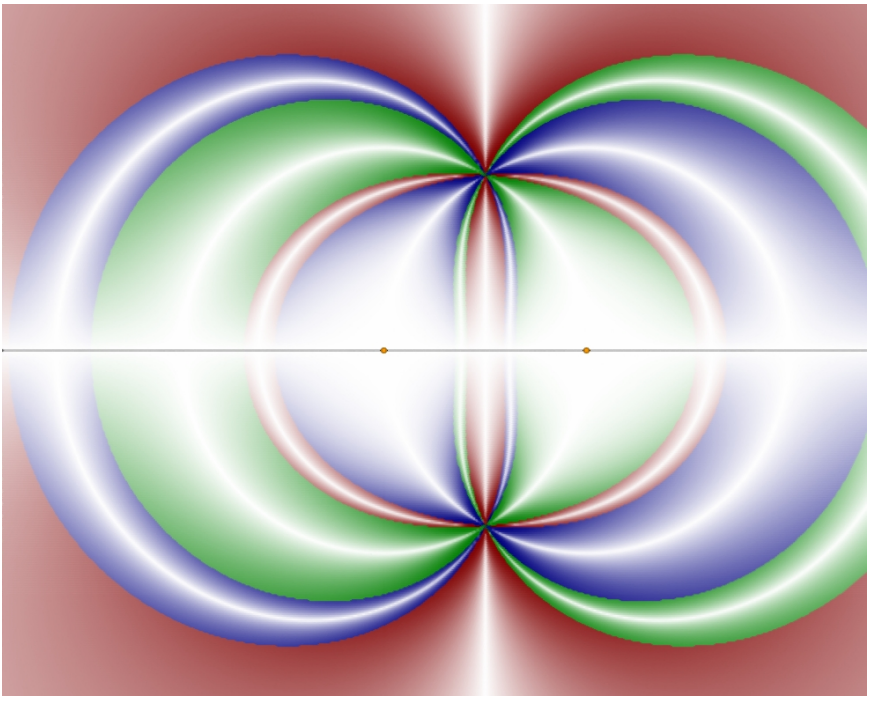


Fig. 6. Graphic result obtained for the problem of equality of the areas tA, tB, tC of triangles determined by the internal and external bisectors at vertices A, B, C. Consider, for each location of C in the plane, the minimum of the absolute values $tA - tB, tB - tC, tC - tA$. Red (gradient) color is associated to this point C when the minimum is reached at $tA - tB$, green if it is at $tB - tC$, blue if at $tC - tA$. Visit http://geogebra.es/imag/fig6 for a full color figure.

– Triangles such that $area(C, M, N) = area(A, P, Q)$

$G_{AC} = F_{1AC} \times F_{2AC} \times F_{3AC}$
$F_{1AC} = w$
$F_{2AC} = v^2 + w^2 - 2v$
$F_{3AC} = v^4 + 2v^2w^2 + w^4 - v^2 - w^2 + 2v - 1$ (Descartes' oval).

– Triangles such that $area(A, P, Q) = area(B, R, S)$

$G_{AB} = F_{1AB} \times F_{2AB} \times F_{3AB}$
$F_{1AB} = w^2$
$F_{2AB} = 2v - 1$
$F_{3AB} = v^4 + 2v^2w^2 + w^4 - 2v^3 - 2vw^2 + v^2 + w^2 - 1$ (Cassini's oval)

– Triangles such that $area(C, M, N) = area(B, R, S)$

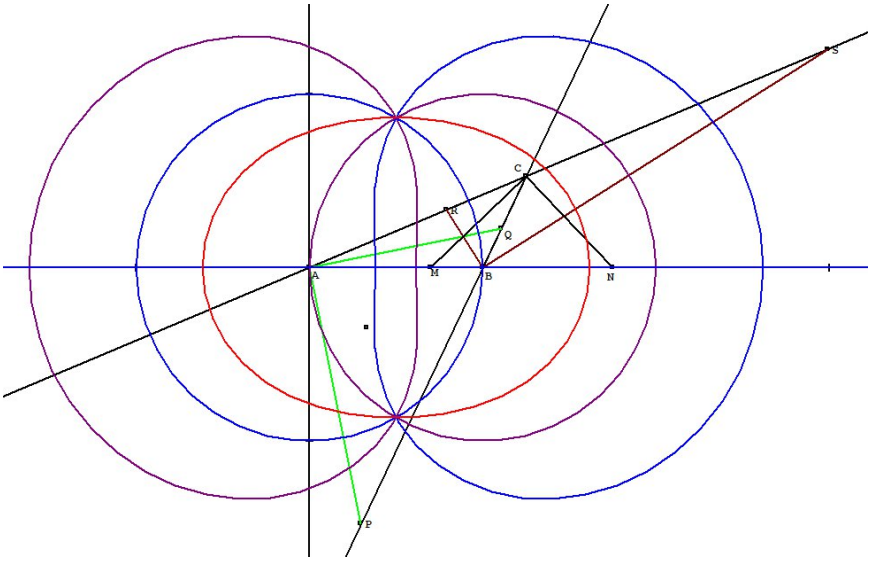


Fig. 7. Results obtained by GDI to the problem of equality of the areas of triangles determined by the internal and external bisectors at each vertex.

$G_{BC} = F_{1BC} \times F_{2BC} \times F_{3BC}$
$F_{1BC} = w^2$
$F_{2BC} = v^2 + w^2 - 1$
$F_{3BC} = v^4 + 2v^2w^2 + w^4 - 4v^3 - 4vw^2 + 5v^2 + w^2 - 4v + 1$ (Descartes' oval)

Therefore, we conclude that the triangles with base $A(0,0)$ and $B(1,0)$ and verifying the equality of at least one couple of bisector triangles areas, must have the third vertex C on the curve $G_{AC} \times G_{AB} \times G_{BC} = 0$, displayed on [Figure 6] and [Figure 7].

The observation of any of the above figures shows that there are triangles in which $area(C, M, N) = area(B, R, S) = area(A, P, Q)$. It is easy to check that the only two possible solutions are $C{=}(\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$, so there are no other than equilateral triangles with equal areas for the three bisector triangles. In fact, it is the case of infinite areas.

6 Conclusion

The usage of Dynamic Geometry programs (DGS), such as GeoGebra, has clear advantages for the study and modeling of different geometric situations. But the purely numerical approach –in practice, the only one available in these programs for the manipulation of large expressions– and the restriction to handle primitive objects (such as lines, points or conics) of these systems, presents serious drawbacks when dealing with automatic discovery of geometric facts. We think

that enhancing DGS with symbolic computation features is one way to overcome these limitations.

We have approached a collection of open questions, concerning the equality of lengths for bisectors on a given vertex of a triangle, by using automatic deduction algorithms through a dynamic geometry program (GDI), one of the few including such discovery features. The background symbolic computations have been performed in CoCoA. Moreover, the support of a specific GeoGebra package for displaying complicated locus, has been essential in the exploratory phase of our research and very useful for the graphic presentation of the output.

We believe that our contribution shows the interest, the simplicity and the power of this collaborative approach (GDI, CoCoA, GeoGebra) to the discovery of new geometric results. We have shown how the different tasks can be formulated with a Dynamic Geometry package (GDI), solved with the help of a computer algebra package (CoCoA), and represented through a graphic –numerical– tool developed within a different DGS (GeoGebra). We are working towards the full integration of all these features into a single, widely distributed and performing product that could automatize the different steps and tasks we have performed in our paper.

Acknowledgment

Second author acknowledges the support of the research grant MTM2008-04699-C03- 03 from the Spanish MEC.

References

[1] Recio, T., Vélez, P.: Automatic Discovery of Theorems in Elementary Geometry. Journal of Automated Reasoning 23, 63–82 (1999)
[2] CoCoA, http://cocoa.dima.unige.it
[3] Botana, F., Valcarce, J.L.: A dynamic-symbolic interface for geometric theorem discovery. Computers & Education 38(1-3), 21–35 (2002)
[4] Botana, F., Valcarce, J.L.: A software tool for the investigation of plane loci. Computers & Education 61(2,1), 139–152 (2003)
[5] GeoGebra, http://www.geogebra.org
[6] Losada, R.: Propiedad de color dinámico en geogebra, http://geogebra.es/color_dinamico/color_dinamico.html
[7] http://www.mathematik.uni-bielefeld.de/sillke/PUZZLES/steiner-lehmus
[8] Wu, W.-t., Lü, X.-L.: Triangles with equal bisectors. Education Press, Beijing (1985) (in chinese)
[9] Wang, D.: Elimination practice: software tools and applications. Imperial College Press, London (2004)
[10] Botana, F.: Bringing more intelligence to dynamic geometry by using symbolic computation. In: Symbolic Computation and Education, pp. 136–150. World Scientific, Singapore (2007)

[11] Losada, R., Recio, T., Valcarce, J.L.: On the automatic discovery of Steiner-Lehmus generalizations. In: Richter-Gebert, J., Schreck, P. (eds.) Proceedings ADG 2010, Munich, pp. 171–174 (2010)
[12] Dalzotto, G., Recio, T.: On protocols for the automated discovery of theorems in elementary geometry. Journal of Automated Reasoning 43, 203–236 (2009)
[13] Andradas, C., Recio, T.: Missing points and branches in real parametric curves. Journal of Applicable Algebra in Engineering, Communication and Computing (AAECC) 18(1-2), 107–126 (2007)
[14] http://www.mathcurve.com/courbes2d/strophoiddroite/strophoiddroite.shtml

On the Parametric Representation of Dynamic Geometry Constructions

Francisco Botana

Departamento de Matemática Aplicada I, Universidad de Vigo, Campus A Xunqueira, 36005 Pontevedra, Spain
fbotana@uvigo.es

Abstract. This paper describes an ongoing implementation of an open source library dealing with parametric representation of dynamic geometry constructions. We show how some current issues in standard dynamic geometry environments, such as computing envelopes of lines which are not primitive objects known by the geometric system, can be efficiently solved using the correspondence between geometry and algebra. We also propose enriching the *tool* (or *macro*) mechanism, available in some environments, with our parametric approach. Finally, some problems arising from the algebraic method considered are also studied.

Keywords: dynamic geometry, parametric representation, automatic discovery, derived curves, macroconstructions.

1 Introduction

The development, in the late eighties, of The Geometer's Sketchpad [1] and Cabri [2] marked the birth of the dynamic geometry (DG) paradigm. Since then, a myriad of dynamic geometry environments have been released (too many to be listed here, see [3]) and, some of which have been primarily used in secondary mathematical education.

Almost simultaneously, promising results were obtained in automatic reasoning in elementary geometry driven by a new class of algebraic approaches, mainly the Groebner bases method [4] and the one due to Wu [5]. Intense theoretical research on both methods was conducted (see, for instance, [6,7,8,9]), and software packages, mostly using Wu's method, were made public [8,10,11]. A further development of the automatic theorem–proving method proposed by Kapur was described in [12]. It concluded aiming for a deeper interleaving of dynamic geometry and automated discovery paradigms. Such a wish was partialy fulfilled by linking The Geometer's Sketchpad with a Maple library for parametric description of constructions in [13] using Wu's method, and with Groebner bases in GDI [14,15], a DG prototype using Mathematica and CoCoA [16] as back–end symbolic engines.

In this paper we continue pursuing the above stated goal. We describe a preliminary implementation of an open source library designed to extend current abilities of standard DG environments. Section 2 summarizes the state of the art

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 342–352, 2011.

in DG on deriving new curves and some of our previous findings on this subject. A problem concerning the generation of the envelopes of families of general lines is posed, and the different answers provided by some well known DG systems are given. In Section 3 we introduce the open source library. The rationale for using free tools is discussed and we show how to use the library for solving some problems related to curve deriving. Some problems concerning mismatchs in the correspondence between algebra and geometry are also discussed. The paper concludes pointing out the next items in our work of enhancing the DG paradigm with more powerful symbolic tools.

2 Curve Derivation: State of the Art

An informal definition of the DG paradigm is that unconstrained parts of the constructions can be moved and, as they are, all other elements automatically self–adjust, preserving dependence relationships and constraints [17]. An immediate consequence of this behavior is that it allows us to visualize the path of an object that depends on another object while this one is dragged. If the dependent object is a point, in general its trace provides a locus, whereas if it has higher dimension, the path can be used to suggest related geometric elements, such as envelopes. Loci generation has been listed as one of the five properties needed by a geometry system to be considered dynamic [18].

2.1 Finding Loci and Envelopes in Dynamic Geometry Environments

Most dynamic geometry systems implement loci generation using what we have called an *interactive approach* [19]. The basic strategy is simple: in order to compute the locus of a point depending somehow on another, that lies on a predefined path, one just needs to sample this path and register the position of the locus point for each member of the sample. The list of these positions constitutes a subset of the locus. Usually, DG systems will join, using some ad–hoc heuristics, the points on contiguous positions, returning an object similar to the other basic ones in the construction. The main exception to this joining approach is Geometry Expert [10]. So, the loci obtained in the way sketched above are just a collection of screen pixels, the system has no algebraic information about them, and sometimes they behave aberrantly when plotting them due to continuity issues (see [15] for this problem).

This loci–computing strategy is also used for dealing with envelopes of families of lines in standard DG systems. The Geometer's Sketchpad and Geometry Expert suggest envelopes by plotting a collection of family lines, while Cabri and Cinderella [20] share this approach but are sometimes able to return them as a line in the case of simple families. It must be noted that Cabri claims that "for a locus, the algorithm produces its algebraic equation if its degree is no greater than 6. For loci whose points are of very different magnitudes, numerical errors appear very rapidly as the degree increases" [21]. Although this algorithm has

not been made public by the vendor, it seems that it is very unstable, and the returned results are frequently erroneous even for simple cases [22]. For instance, computing an astroid as the envelope of a moving segment with each end on one of a pair of perpendicular axes, Cabri gives different equations when constraining the segment to the upper and lower half planes, as shown in Fig. 1. Furthermore, the degree of the equations cannot be explained as a rounding error, but a wrong algorithmic approach to finding the real sextic.

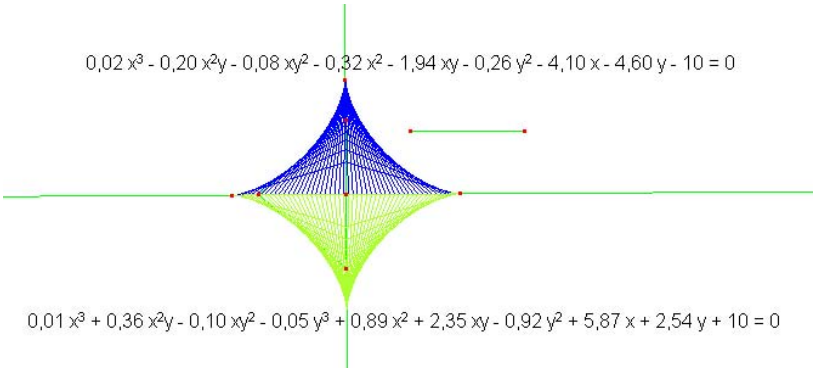


Fig. 1. Equations returned by Cabri for the upper and lower halves of an astroid

2.2 A Symbolic Approach to Curve Derivation

The application of symbolic methods for loci generation, although restricted to algebraic curves, generalizes the class of obtainable loci, returns their algebraic expressions, and behaves in a uniform way for all construction instances. Since Recio and Vélez did not deal with loci in [12], we proposed in [23] a simple extension of their automated discovery proposal and showed that it can efficiently be implemented in a DG environment. This extension has also been added to JSXGraph [24], a library for interactive geometry, and its incorporation into GeoGebra [25] is currently under development [26].

A similar strategy was used in [27] for symbolic computation of envelopes and other derived curves. Since the implementation partially used proprietary software (Mathematica) and the DG prototype just worked under Windows, we decided to rewrite the algorithms as an open source library and to develop it as an add–on for standard DG environments.

2.3 Computing Envelopes of Geometric Loci

We describe in this subsection a problem currently unsolvable in DG systems. It deals with deriving objects from non basic objects. As said above, most systems are not able to compute the equations of loci or envelopes. Consider, for instance, an offset curve of a parabola, that is, the envelope of a family of equal radius

circles centered at a point lying on the parabola. While standard DG systems show this envelope by plotting a reduced list of such circles, GDI returns it as a simple line and also provides its equation. Figure 2 illustrates these offsets in Cinderella (left) and GDI (right).

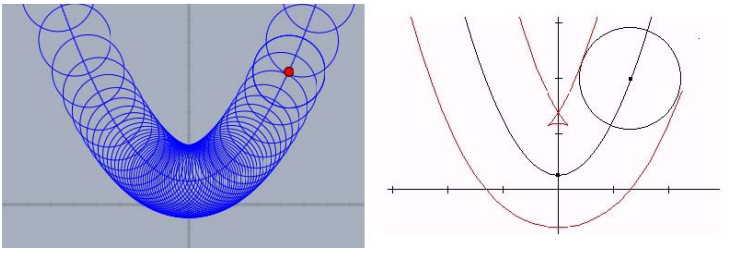


Fig. 2. The offset of a parabola as plotted by Cinderella (*left*) and GDI (*right*)

Nevertheless, plotting some family members for visualizing the envelope only works if the lines are basic objects in the environment. Replacing the circle moving along the parabola by another circle, obtained as a locus, no standard system, as far as we know, can compute the envelope. There is just a solution: moving the circle, with its trace activated. But note that not all systems can trace loci. For instance, current version of GeoGebra cannot trace them (although GeoGebra 4.0 will). Besides that, using the trace option is hard in order to get a descriptive picture of the envelope.

Consider an ellipse built as a locus (following the gardener's method) and let A, B be its foci, where A is a fixed point and B lies on a line. Figure 3 shows the envelope of these ellipses, when B moves along its parent line, in Cinderella (left) and GeoGebra 4.0 Beta (right).

In the above situation, GDI would also fail, since it could not return this ellipse as a locus. Note that the ellipse we are trying to build is the locus of all points X in the plane such that distance(A, X) + distance(B, X) is a constant. Being B a semifree point, the algebraic answer should be, as GDI returns, the whole plane (or a bidimensional subset). And asking for the envelope of non linear elements is forbidden in GDI.

In order to overcome the above situation, we roughly proceed as follows. We compute the ellipse as a locus in a 4–dimensional space (where two variables are the locus ones, and the other pair comes from B), and project it over the space of the first pair of variables (see next Section for a more detailed description).

3 The Open Source Library

There are two approaches for extending DG systems with new symbolic, algebraic related, abilities. The first one consists of incorporating the algorithms in

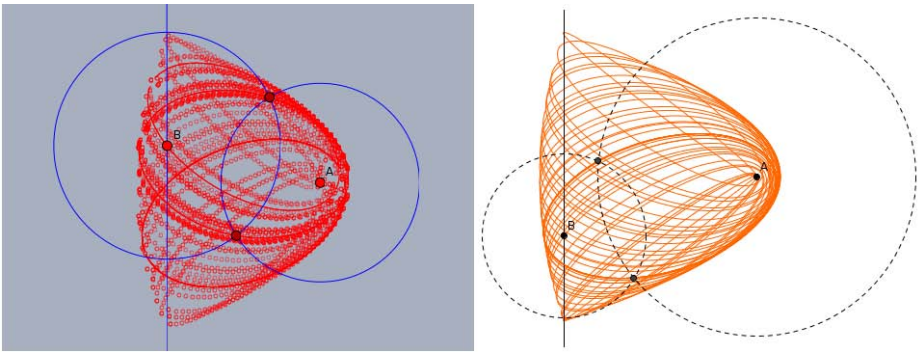


Fig. 3. The traces of a moving ellipse obtained as locus in Cinderella (*left*) and GeoGebra 4.0 Beta (*right*)

the heart of the system, while the second one uses preexistent software (mainly computer algebra systems, CAS) and connects them somehow with the DG environment. Apart from historical reasons concerning the places where Wu and Groebner methods first appeared, we have no doubt about the importance of the developers milieu when making such a decision. So, Chinese systems mainly use the first approach, while academic proposals coming from the occidental world use external CAS, since their cost is a minor point when distributing the system. Paradigmatic examples are Geometry Expert and Geometry Expressions [28]. Nevertheless, the global systemic crisis, with its present and upcoming cuts in educational and non–profit research budgets, the globalization of the information and some centrifugal tendencies in academy, help to explain a renewed interest in providing free access to DG systems and related tools. The growing use of GeoGebra and its probably settlement as the de facto standard in secondary mathematical education is a vivid example of it. Apart from its free character, the open source model followed by GeoGebra also helps explain its success, since it involves a bigger part of the educational community and reacts faster to user requirements and updates than other proprietary DG software does. Same reasons apply to a recent CAS, Sage [29], a free open source mathematics software system licensed under the GPL, whose declared mission is "*creating a viable free open source alternative to Magma, Maple, Mathematica and Matlab*". So, using Sage as development platform, we decided to rewrite from scratch our previous algorithms related to automatic discovery in geometry.

3.1 The Structure of the Library: Examples

Currently, there are about a dozen of basic geometric predicates, and some internal functions, needed for internal work or consistency checkings. The interested reader can download the library as a Sage worksheet or text from [30]. The constructive predicates are: `FreePoint`, `Line`, `Circle`, `MidPoint`, `PointOnObject`, `ParallelLine`, `PerpendicularLine`, `TangentLine`,

`IntersectionObjectObject`, `Locus`, `Locus2` and `Envelope`. When invoked, each predicate adds to a dictionary, called `Todo`, the corresponding geometric element together with some relevant information. The elementary action of adding a point, for instance, is performed through the function

```
def FreePoint(pnt,absc,orde):
    """Adds the point $pnt$ with coordinates $(absc,orde)$
    to the geometric construction."""
    Todo.update({pnt:{'coords':(absc,orde),
                      'parents':Set([]),
                      'type':'FreePoint',
                      'hist':['FreePoint',pnt,absc,orde],
                      'eq':Set([])}})
```

where partial indenting has been done for legilibility. So, the evaluation of the command `FreePoint('P',2,-1)` defines the point $P(2,-1)$, with type `'FreePoint'`, and without parents or equation, being the remaining keyword for future use.

Defining the midpoint of a pair of points is done through the function

```
def MidPoint(n,p,q):
    """Constructs the midpoint $n$ of points $p$ and $q$."""
    if n in Todo.keys():
        eq=Todo[n]['eq']
        parents=Todo[n]['parents']
        temp=Todo[n]['coords']
    else:
        eq=Set([])
        parents=Set([])
        temp=(BoVar.pop(),BoVar.pop())
    eq=eq.union(Set([temp[0]-1/2*(x(p)+x(q)),
                     temp[1]-1/2*(y(p)+y(q))]))
    parents=parents.union(Set([p,q]))
    Todo.update({n:{'coords':temp,
                    'type':'BoundedPoint',
                    'eq':eq,
                    'parents':parents,
                    'hist':['MidPoint',n,p,q]}})
```

where it should be noted that multiple definition for points is allowed. This function also introduces the second type for 0–dimensional objects, `BoundedPoint`. There are other two general types for objects in the library, `Line`, for 1–dimensional objects, and `Plane`, for any geometric object with dimension 2.

3.2 Finding Envelopes of Loci

Recalling the unsolved envelope problem in 2.3 we define the ellipse as follows:

```
FreePoint('A',4,0)
FreePoint('P1',0,0)
FreePoint('P2',0,1)
Line('y','P1','P2')
PointOnObject('B','y')
FreePoint('M',2,2)
FreePoint('N',2,7)
Line('MN','M','N')
PointOnObject('P','MN')
Circle('c1','A','M','P')
Circle('c2','B','N','P')
IntersectionObjectObject('X','c1','c2')
Locus2('loc','X','B','P')
```

Asking for the definition of the special object `loc` we get

```
{'type': 'Locus2',
 'tracer': 'X',
 'hist': ['Locus2', 'loc', 'X', 'B', 'P'],
 'parents': {'X', 'P'},
 'mover': 'P',
 'eq': {x1, 4*orde^2*x2^2 - 4*orde*x2^3 - 36*absc^2 - 100*orde^2
 + x2^4 - 32*absc*orde*x2 + 16*absc*x2^2 + 164*orde*x2 - 82*x2^2
 + 144*absc + 81},
 'implicit': False}
```

where the locus is the zero set of two polynomials in $\mathbb{Q}[absc, orde, x_1, x_2]$. The evaluation of `Envelope('env','loc','B')` just carries out the elementary computation for plane envelopes, returning

```
{'type': 'Line',
 'hist': ['Envelope', 'env', 'loc', 'B'],
 'parents': {'loc', 'B'},
 'mover': 'B',
 'eq': {absc^2*orde^4 + orde^6 - 16*absc^3*orde^2 - 24*absc*orde^4
 - 36*absc^4 + 74*absc^2*orde^2 - 2*orde^4 + 432*absc^3
 + 32*absc*orde^2 - 1647*absc^2 - 207*orde^2 + 1656*absc + 1296},
 'tracer': 'loc'}
```

Since the library provides the equation of the envelope, plotting it in a DG environment would be a simple matter. Figure 4 shows a plot of this equation. Nevertheless, special care must be taken here. Factoring the polynomial we get the shown parabolas, $-orde^2 + 18 * absc + 9 = 0$ and $orde^2 + 2 * absc - 9 = 0$, plus an extraneous factor $-absc^2 - orde^2 + 8 * absc - 16$, that is, the focus A. The problem of such extraneous factors deserves special consideration and its relation with the application of this library to DG environments will be discussed in a future note.

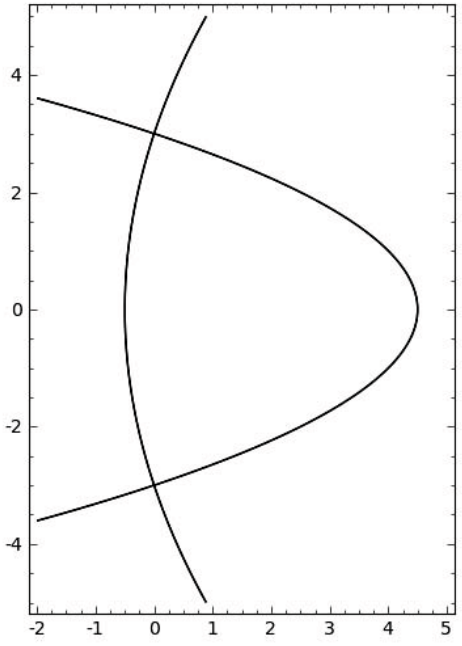


Fig. 4. A Sage plot of the moving ellipse envelope

3.3 Macro Definitions

Even with the small set of library basic functions listed in 3.1, new curves can be easily derived following an approach that resembles the well known mechanism of macros or tools in standard DG environments. Consider computing the pedal curve of a line with respect to a given point. The following function solves the problem:

```
def Pedal(n,l,p):
    """Computes the pedal line $n$ of line $l$ with respect to
    point $p$."""
    TangentLine('tan',l,'ptemp')
    PerpendicularLine('perp',p,'tan')
    IntersectionObjectObject('x','tan','perp')
    Locus(n,'x','ptemp')
```

If the line `l` is the ellipse with foci in $(0,0)$ and $(4,0)$, and passing through $(9/2,1/2)$ (so having as equation $36x^2+100y^2-144x-81=0$), its pedal curve with respect to the point $(2,0)$, computed with the above function, is the quartic $4x^4+8x^2y^2+4y^4-32x^3-32xy^2+71x^2+23y^2-28x-36=0$.

The above protocol would be natural in any DG system, assuming the system can compute tangent lines. If using GeoGebra, the procedure can be easily defined as a tool, since all involved operations are defined. We propose enhancing

the definition of tools with our library. It must be noted that this connection will require fine tuning. For instance, GeoGebra (and many other DG systems) deal with straight lines, conics and greater degree curves as essentially different objects, while they are just `Line` objects in the library. Defining a parent class for 1–dimensional objects would solve this point.

Another illustration of this technique is the computation of catacaustics, the reflective case of caustics. These caustics envelope a family of reflected rays, so, given a fixed point p and a line l, the procedure

```
def Caustic(n,l,p):
    """Computes the catacaustic $n$ of line $l$ with radiant $p$."""
    TangentLine('tan',l,'ptemp')
    PerpendicularLine('perp','ptemp','tan')
    Symmetrical('q',p,'perp')
    Line('ref_ray','q','ptemp')
    Envelope(n,'ref_ray','ptemp')
```

returns the caustic.

3.4 Using the Library: Caveat Emptor!

Some considerations must be made about using the library. The first one is an advice for non expert users when mimicking the macro approach described above. Both procedures use some intermediate objects that share their names (`tan`, `perp`, `ptemp`,...) and, since accessing elements is done by name, a previously constructed element could be used. Multiple constraining is allowed for some elements (`BoundedPoint`, for example). So, a careless user can introduce undesired constraints for construction elements. The library code contains two alternative definitions for computing pedals and caustics, with appropriate names for intermediate objects and where the dictionary values of the final curves contain more specific information about them.

The second consideration involves the relation between varieties and ideals, so allowing a correspondence between geometry and algebra. The problem of extraneous factors, illustrated in Section 3.2, can be seen as one of an algorithmic nature. It is planned that new library versions will add alternative algorithms (different types of resultants) and heuristic strategies for dealing with it. Nevertheless, there is another source or imprecision, coming from the elimination approach taken in the library. Although algebraic elimination has been proved as succesfull when introducing automated deduction in DG systems, its findings must be critically examined in this environment. For instance, once computed the pedal in Section 3.3, one should note that the zero set of the polynomial wrongly includes the pedal point. So, we cannot rely on the pedal object for a posteriori computations. This behavior can be explained as follows: when we eliminate variables, we do not get just a projection, but its Zariski closure. That is, there can exist spurious points. As GDI warns, when finding a locus, "*The locus is (or is contained in)...*". There is an ongoing theoretical work on this subject. Future versions of the library will incorporate these new developments.

4 Conclusions and Future Work

We report a free open source implementation of a Sage based library for computations related to plane parametric geometric contructions. The library solves the problem of finding the algebraic description of objects in a virtual ruler and compass environment. Although it is restricted to a purely algebraic realm and it does not give a complete solution to the translation between geometry and algebra (if there is one!), it is an efficient solver for handy computations and can be easily integrated as an add–on to general dynamic geometry environments. Future work concerning library development involves automatic theorem proving and discovery.

Acknowledgements

The author was partially supported by research grant MTM2008-04699-C03-03/MTM from the Spanish MICINN.

References

1. Jackiw, N.: The Geometer's Sketchpad, vol. 4.0. Key Curriculum Press, Berkeley (2002)
2. Laborde, J.M., Bellemain, F.: Cabri Geometry II. Texas Instruments, Dallas (1998)
3. List of interactive geometry software, `http://en.wikipedia.org/wiki/List_of_interactive_geometry_software`
4. Buchberger, B.: Groebner Bases: An Algorithmic Method in Polynomial Ideal Theory. In: Bose, N.K. (ed.) Multidimensional Systems Theory, pp. 184–231. Reidel, Dordrecht (1985)
5. Wu, W.T.: Mechanical Theorem Proving in Geometries. Springer, Vienna (1994)
6. Kapur, D.: A Refutational Approach to Geometry Theorem Proving. Artif. Intell. 37, 61–94 (1988)
7. Kapur, D.: Using Groebner Bases to Reason about Geometry Problems. J. Symb. Comput. 2, 399–408 (1986)
8. Chou, S.C.: Mechanical Geometry Theorem Proving. Reidel, Dordrecht (1988)
9. Chou, S.C.: Proving Elementary Geometry Theorems Using Wu's Algorithm. In: Bledsoe, W.W., Loveland, D.W. (eds.) Automated Theorem Proving: After 25 years. Contemporary Mathematics, vol. 29, pp. 243–286. AMS, Providence (1984)
10. Gao, X.S., Zhang, J.Z., Chou, S.C.: Geometry Expert. Nine Chapters, Taiwan (1998)
11. Wang, D.: GEOTHER: A Geometry Theorem Prover. In: McRobbie, M.A., Slaney, J.K. (eds.) CADE 1996. LNCS, vol. 1104, Springer, Heidelberg (1996)
12. Recio, T., Vélez, M.P.: Automatic Discovery of Theorems in Elementary Geometry. J. Autom. Reasoning 23, 63–82 (1999)
13. Roanes–Lozano, E., Roanes–Macías, E., Villar, M.: A Bridge between Dynamic Geometry and Computer Algebra. Math. Comput. Model. 37(9–10), 1005–1028 (2003)
14. Botana, F., Valcarce, J.L.: A Dynamic–Symbolic Interface for Geometric Theorem Discovery. Comput. Educ. 38(1–3), 21–35 (2002)

15. Botana, F., Recio, T.: Towards solving the dynamic geometry bottleneck via a symbolic approach. In: Hong, H., Wang, D. (eds.) ADG 2004. LNCS (LNAI), vol. 3763, pp. 92–110. Springer, Heidelberg (2006)
16. Capani, A., Niesi, G., Robbiano, L.: CoCoA, a System for Doing Computations in Commutative Algebra, http://cocoa.dima.unige.it
17. King, J., Schattschneider, D.: Geometry Turned On. MAA, Washington (1997)
18. Gao, X.-S.: Automated geometry diagram construction and engineering geometry. In: Wang, D., Yang, L., Gao, X.-S. (eds.) ADG 1998. LNCS (LNAI), vol. 1669, p. 232. Springer, Heidelberg (1999)
19. Botana, F.: Interactive versus Symbolic Approaches to Plane Loci Generation in Dynamic Geometry Environments. In: Sloot, P.M.A., Tan, C.J.K., Dongarra, J.J., Hoekstra, A.G. (eds.) ICCS 2002. LNCS, vol. 2657, pp. 801–810. Springer, Heidelberg (2003)
20. Richter–Gebert, J., Kortenkamp, U.: The Interactive Geometry Software Cinderella. Springer, Berlin (1999)
21. http://download.cabri.com/data/pdfs/manuals/cabri2plus140/Man_uk_PDF3.pdf
22. Botana, F., Abánades, M.A., Escribano, J.: Computing locus equations for standard dynamic geometry environments. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2007. LNCS, vol. 4488, pp. 227–234. Springer, Heidelberg (2007)
23. Botana, F., Valcarce, J.L.: A Software Tool for the Investigation of Plane Loci. Math. Comput. Simul. 61(2), 139–152 (2003)
24. JSXGraph, http://jsxgraph.uni-bayreuth.de
25. GeoGebra, http://www.geogebra.at
26. GeoGebra Locus Line Equation, http://www.geogebra.org/trac/wiki/LocusLineEquation
27. Botana, F., Valcarce, J.L.: Automatic Determination of Envelopes and Other Derived Curves within a Graphic Environment. Math. Comput. Simul. 67(1–2), 3–13 (2004)
28. Geometry Expressions, http://www.geometryexpressions.com
29. Stein, W.A., et al.: Sage Mathematics Software (Version 4.6.0). The Sage Development Team (2010), http://www.sagemath.org
30. Automatic Discovery Sage Library, http://webs.uvigo.es/fbotana/AutDiscLib.sws.txt

Using Free Open Source Software for Intelligent Geometric Computing

Miguel A. Abánades[1], Francisco Botana[2], Jesús Escribano[3], and José L. Valcarce[4]

[1] CES Felipe II, Universidad Complutense de Madrid, Campus de Aranjuez, 28300 Aranjuez, Spain
abanades@ajz.ucm.es
[2] Departamento de Matemática Aplicada I, Universidad de Vigo, Campus A Xunqueira, 36005 Pontevedra, Spain
fbotana@uvigo.es
[3] F. Informática, Universidad Complutense de Madrid, 28040 Madrid, Spain
jesus_escribano@mat.ucm.es
[4] IES Pontepedriña, 15701 Santiago de Compostela, Spain
jvalcarce@edu.xunta.es

Abstract. This paper describes some examples of fruitful cooperation between geometric software tools (in particular, GeoGebra) and a free open source computer algebra system, Sage (Software for Algebra and Geometry Experimentation). We collect some of our efforts for enhancing mathematics education via technologically rich environments. We show that a math teacher with no specialized programming knowledge can mix widespread resources to get motivating new teaching tools. Furthermore, we explore an exciting but barely used (even known!) characteristic of Sage: its use as a remote kernel. We test it by computing symbolic tasks in a dynamic geometry prototype that are currently out of scope of standard dynamic geometry software. Finally, we illustrate the development of web–based geometric resources by communicating GeoGebra and Sage through Javascript.

Keywords: learning resources, dynamic geometry, computer algebra systems, Sage, Internet accessible mathematical computation, GeoGebra.

1 Introduction

In the last years, the use of new technologies in education is becoming more and more important. Every day we discover a new tool: web pages, wikis, (simple or advanced) math applications, different ways of interaction with students or with other colleagues, etc. The development of these tools is having a great impact on teaching and also on research.

One very well known paradigm of a new technology applied to education is the case of Dynamic Geometry Systems (DGS). Using a DGS, you can draw many interesting geometric configurations and, what is more important, you

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 353–367, 2011.

can also *interact* with the configuration. If, for example, you move a point in the configuration, all the elements depending on this point (say, a line passing through the point, or a circle whose center is that point) move accordingly. There are many examples of DGS, out of which The Geometer's Sketchpad [1], Geometry Expert [2], Cabri [3], Cinderella [4] and GeoGebra[5] are the most popular.

DGS produce graphical objects and from the begining, we can find in many authors the idea of adding capabilities to DGS, mainly computation capabilities, in order to expand the possibilities of use of the DGS. Different solutions have been proposed, from producing DGS with own incorporated computation habilities, to connecting DGS to other computer systems. In particular the connection of DGS with Computer Algebra Systems (such that Mathematica or Maple), has proved to be quite fruitful [6], [7].

The authors have worked in this line [8]. There exist good DGS and there exist good CAS, so it should not be neccesary to "re-invent the wheel", we just have to connect the different systems.

However there is one important consideration. If you connect one DGS with one CAS, but users have problems accessing any of them (because it is expensive or difficult to install), the effort is useless. From the premise that allowing free access to the resources is fundamental, the authors have worked in two directions: constructing web applications (e.g. [9]) and promoting the use of free open source software [10].

In this note we will focus on two open source applications. One is the DGS GeoGebra and the other is the CAS Sage [11]. Sage is a powerful CAS, developed by W. Stein.

Besides (freely) downloading and installing Sage locally in your computer, one can easily access Sage remotely through its web interface *notebook* from a Sage server. This makes any installation of software unnecessary, and thus allowing the access to Sage to a much wider range of users.

In this paper we show three examples of interaction between DGS and CAS. In section 2 we show the posibilities of using applets inside a Sage worksheet illustrated by two problems: the computation of Lagrange multipliers and the finding of equations for envelopes of families of algebraic curves. In Section 3 we connect the prototype for a DGS, called GDI and developed by the second and fourth authors ([12], [13]), and Sage, to add algebraic computation to GDI. Finally, in Section 4 we show a different approach: connecting Sage and GeoGebra in a web page, using Javascript.

2 Embedding Applets in Sage Worksheets

The most common and simplest way to interact with Sage, mentioned in Section 1 is through the notebook, which is a graphical user interface for the Sage software. Since the notebook is a web application that can be remotely or locally accessed, it is an ideal support for distributing and sharing material with students.

Even more, with only some notions on web servers, setting up our own Sage notebook server is not a very difficult task [14]. Despite some security issues, appearing when allowing external users to remotely access an entire CAS, a Sage server can replace expensive proprietary software and solve most problems in the day to day undergraduate math teaching.

In this section we sketch two important characteristics of the notebook: the *interact* function and the easy integration of external applets in Sage worksheets. The first example deals with Lagrange multipliers, where an applet for approximate computation of extrema gives numerical results while the exact symbolic approach is also used. Furthermore, the symbolic part is encapsulated under the interact function, hence allowing the use of the worksheet as an immediate graphic–symbolic calculator. Nevertheless, the mathematical code behind the symbolic part remains accessible for examination or modification. We must underline that this approach of allowing code access should be, in our opinion, the only one accepted when teaching and learning mathematics. Regarding the second example, we include a GeoGebra construction into the Sage worksheet and show how the proposed problem gets an analytical solution currently unavailable within the dynamic geometry environment.

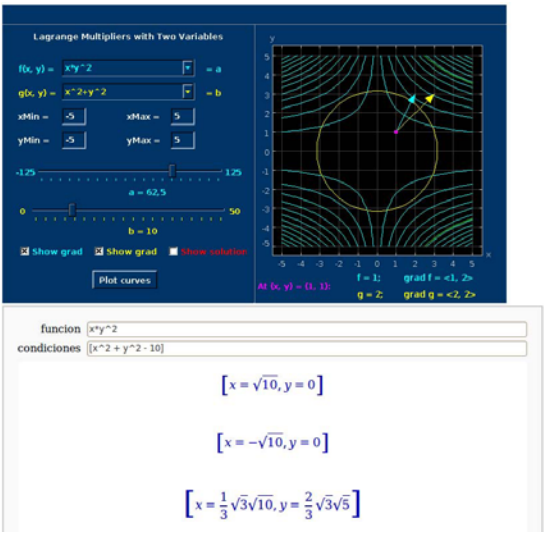


Fig. 1. The Lagrange worksheet showing the applet and part of the symbolic solutions

2.1 Lagrange Multipliers

In [15] an applet for approximate computing of Lagrange multipliers for functions of two variables is offered. Reading the text form of the web page we can find the URLs of the applet files and then use it embedded in the worksheet, as shown in the following code:

```
html('<applet code="LagrangeMultipliersTwoVariables" archive="http:
//ocw.mit.edu/ans7870/18/18.02/f07/tools/lagrangeMultipliersTwoVari
ables.jar, http://ocw.mit.edu/ans7870/18/18.02/f07/tools/mk_lib.jar
, http://ocw.mit.edu/ans7870/18/18.02/f07/tools/parser_math.jar, ht
tp://ocw.mit.edu/ans7870/18/18.02/f07/tools/jcbwt363.jar" width=760
 height=450></applet>')
```

Adding a simple code for the symbolic computation of the extrema candidates, the worksheet (Figure 1) allows a graphic and symbolic approach to learning about this topic. The interested reader can download the worksheet, called "Lagrange-multipliers-applet-interact" in [16], and experiment with it once granted access.

2.2 Finding the Equation of an Envelope

The well known problem of the sliding ladder (that is, considered a ladder supported on a wall, take a point on the ladder and then study the trajectory of the point when the ladder falls down to the floor) has been recently generalized in [17], allowing the ladder to have variable length. The worksheet to deal with this learning situation integrates GeoGebra for drawing the generalized ladder (Figure 2, the lenght of the ladder is the distance from the origin to a point on the curve) and uses the Sage interface for Singular [18] for performing ideal elimination and saturation (see the code in the worksheet) [19].

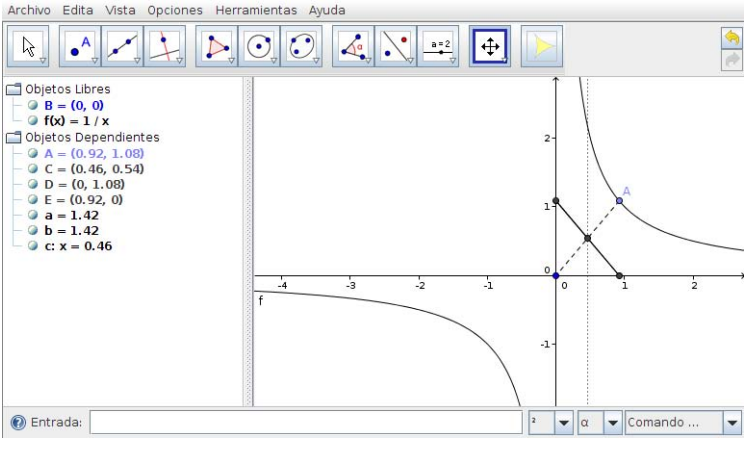


Fig. 2. A flexible ladder constructed in GeoGebra. The point A moves on the hyperbola $y = 1/x$ and the goal is to find the equation of the envelope of the family of (solid) segments.

The equation for the envelope returned by Singular when the ruling curve is the hyperbola $y = 1/x$ is again a hyperbola, $4xy - 1 = 0$. Figure 3 shows the trace of segments in GeoGebra, just suggesting the envelope. Note that since the

last beta version of GeoGebra can plot graphs of general equations, it would be possible for GeoGebra to read the external result and add it to its knowledge base. This possibility, although involves deeper programming skills, is discussed in section 4.

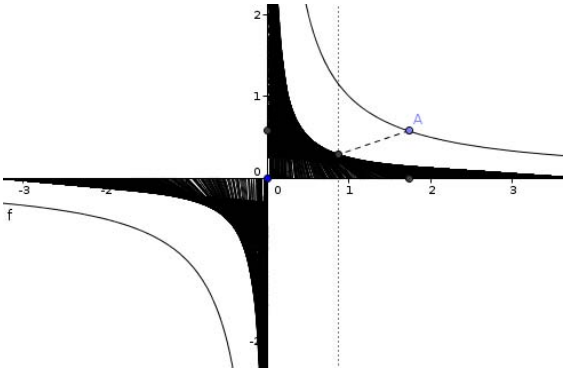


Fig. 3. The envelope returned by GeoGebra

3 Computation of Envelopes and Pedals with GDI Using a Sage Server

Some DGS incorporate, among other functionalities, the computation by numeric methods of graphs of some curves, such that geometric locus, or envelopes of plane curves can be calculated. In particular, many DGS can calculate envelopes of some families of plane curves, including pedals.

In this paper, a family of curves F supported on the guide curve G will be the set of algebraic curves depending on a point x in G. In other words, for each point $x \in G$ there exists a curve in the family F. The guide curve G will typically be an algebraic curve predefined by the DGS, usually a line or a conic.

In general, a family F can be expressed by a polinomial equation, depending on r parameters, $F(x, y, a_1, a_2, ..., a_r)$, where the parameters a_i represent symbolic coordinates of the points in the geometric configuration.

3.1 Envelopes

Following the general set up above, let us consider the following example:

- the cissoid $G(x, y) = -1 + 3x - 3x^2 + x^3 + (1 + x)y^2$, as guide curve G.
- the circle of center $x = (a, b)$, in G, and diameter 1, as the generic curve of the family F. The polynomial equation is $F(x, y, a, b) = (x-a)^2 + (y-b)^2 - 1/4$.

The envelope of the family of plane curves $F(x, y, a, b)$, where there exists a dependency relation between the parameters defined by the polynomial $G(a, b)$, is the set of points $(x, y) \in \mathbb{R}^2$ such that:

- $F(x, y, a, b) = 0$
- $G(a, b) = 0$
- $\frac{\partial}{\partial a}F(x, y, a, b)\frac{\partial}{\partial b}G(a, b) - \frac{\partial}{\partial b}F(x, y, a, b)\frac{\partial}{\partial a}G(a, b) = 0$

where the first equation defines the family of curves, the second equation defines the relation between the two parameters (in our situation, this is the equation of the guide curve, and the parameters are the generic coordinates (a, b) of a point in this curve) and the third one is the tangency condition of the envelope with the lines in the family. By eliminating the parameters a and b, using the three equations, we obtain the equation of the envelope.

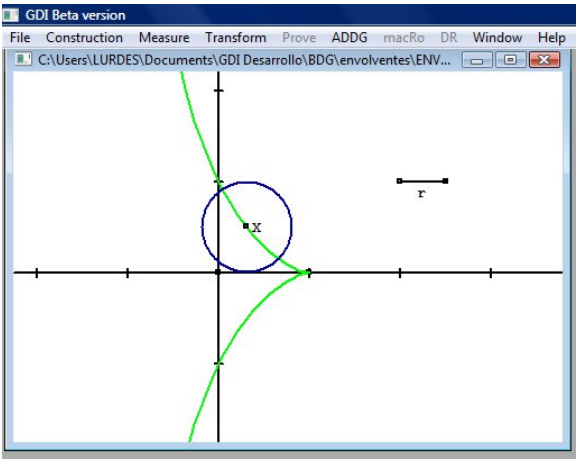


Fig. 4. The cissoid and a circle with center X on the cissoid, constructed using GDI

Using CoCoA [20], we get the polynomial of the envelope of the family $E = E_1 \cdot E_2$, where:

- $E_1 = 256x^8+768x^6y^2+768x^4y^4+256x^2y^6+2048x^7+4608x^5y^2+3072x^3y^4+512xy^6-5376x^6+14016x^4y^2+19584x^2y^4+192y^6-28160x^5-56576x^3y^2+28928xy^4+125792x^4+12816x^2y^2+7856y^4-182912x^3+56864xy^2+106032x^2-64508y^2-13984x+529$
- $E_2 = (4x^2 + 4y^2 - 8x + 3)^3$

We show now an easy procedure to perform the calculation above using the dynamic geometry system GDI. Although GDI can independently calculate envelopes and other curves [21], here we see how we can solve this problem using a remote Sage server. The point being that this technique can be incorporated to any other DGS. The main idea is that the user, to obtain an equation, does not have to know anything about the internal operations in the CAS, or the algebraic operations involved.

The procedure goes as follows. The user produces a geometric construction with, at least, a guide curve G, a point $x \in G$ and a curve f of the family

depending on x (Figure 4). To calculate the envelope of the family F the user must select the curve f and the point x, and click on the corresponding command in the menu.

This action triggers on one hand the production of a file describing the construction along with the computational task (the calculation of the envelope) in the syntax of Singular and at the same time the request to open a SAGE server session. The file is then sent to the server where the code is executed and we obtain the following result, that can be read in the web browser:

```
{ "status": "done", "files": [], "cell_id": 1 } ___S_A_G_E___
(4*x^2 + 4*y^2 - 8*x + 3)^3*(256*x^8 + 768*x^6*y^2 +
768*x^4*y^4 + 256*x^2*y^6 + 2048*x^7 + 4608*x^5*y^2 +
3072*x^3*y^4 + 512*x*y^6 - 5376*x^6 + 14016*x^4*y^2 +
19584*x^2*y^4 - 28160*x^5 + 125792*x^4 - 182912*x^3 +
106032*x^2 + 192*y^6 - 56576*x^3*y^2 + 28928*x*y^4 +
12816*x^2*y^2 + 7856*y^4 + 56864*x*y^2 - 64508*y^2 -
13984*x + 529)
```

What about the graph of the curve? In the current state of application development, the Sage server does not show the graph of the curve, but this is not a relevant issue for two reasons. On one hand, there exist available specialized programs to obtain graphs from equations. On the other hand, the DGS can produce graphs of envelopes of families of curves of this kind, although sometimes they are just approximated versions. To obtain the approximated graph is always possible, while obtaining the exact graph by algebraic methods is difficult sometimes, depending on the number of variables and polynomials describing the problem.

A better alternative to both above situations would be the integration of the new object (the envelope) into the DGS, that is, returning back from the CAS to the DGS. This kind of integration would allow to drag the envelope (this is a defining characteristic of the DGS), or to use it to construct new objects. For example, figure 5 shows the envelope produced by GDI, using CoCoA and

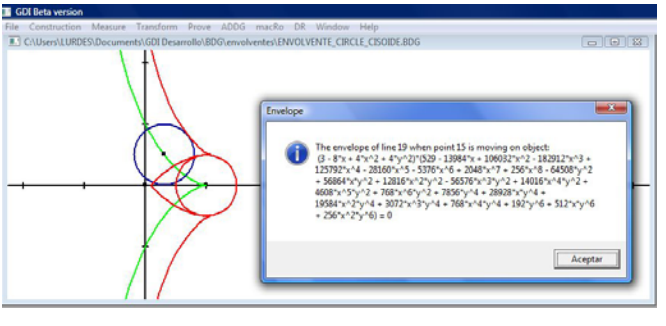


Fig. 5. Envelope of family of circles with center on a point on the cissoid and diameter 1, provided by GDI

Mathematica [22] in local mode to obtain the equation and the graph. We note that the guide curve G (cissoid) is not a primitive object in the system but it has been constructed as a locus using algebraic methods. This makes its equation available to the system so new objects (such as this envelope) can be constructed from it. This possibility is not available in any standard DGS as far as we know.

3.2 Pedal Curves

Pedal curves are planar curves that can be generated from another plane curve and a point on its plane. They were first studied by Colin Maclaurin (1718) and Quetelet (1822) [23].

Given a plane curve G with tangent at each of its points (except perhaps a finite number) and a point P in the plane, called foot point (or pedal point), the pedal of P with respect to G (or pedal of G with respect to P) is the locus set of the orthogonal projections Q of P onto the different tangent lines to the curve G.

Another equivalent way of defining the pedal of P with respect to G is the envelope of the family of curves defined by the circles of diameter PX as the point X moves along the curve G. This is the considered in the following example.

Consider, for example, a circle G defined by its center $O(0,0)$ and a point $A(1,0)$. To compute the pedal of A with respect to the circumference G we take the family of curves F equal to all the circles with diameter AX, where $X(a,b)$ is a point in G (Figure 4). We have:

- $G(x,y) = x^2 + y^2 - 1$
- $F(x,y,a,b) = (x-(a+1)/2)^2 + (y-b/2)^2 - ((a-1)^2 + b^2)/4$

And, as before, it is easy to check with a CAS that the pedal is given by the polynomial $P(x) = x^4 + 2x^2y^2 + y^4 - 2x^3 - 2xy^2 - y^2 + 2x - 1$. We obtain the result by GDI using the Sage server in this format:

```
{ "status": "done", "files": [], "cell_id": 1 } ___S_A_G_E___
x^4 + 2*x^2*y^2 - 2*x^3 + y^4 - 2*x*y^2 - y^2 + 2*x - 1
```

4 Web Resources Based on GeoGebra and Remote Access to CAS

In section 2 above, it was shown how one can enrich a Sage worksheet by inserting applets. In particular, a GeoGebra applet was used to illustrate the problem of finding the envelope of a family of lines associated to the generalization of the sliding ladder problem for a flexible ladder.

In this section we show how one can somehow reverse that approach and bring the CAS to the applet instead of the applet to the CAS. More precisely, we show two examples of web based resources in which a GeoGebra applet is complemented with access to a CAS, allowing the automatic solving of involved geometric tasks not possible in standard dynamic geometry software. In both

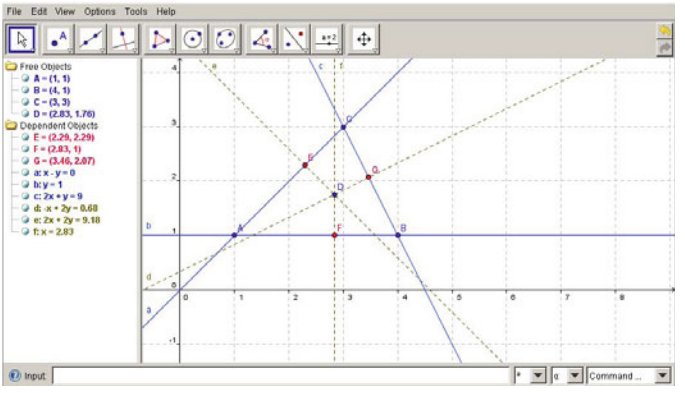


Fig. 6. When are E, F, G aligned?

cases the resources offer the general framework for a particular geometric task in which the different elements can be changed by the user. For this reason, we refer to them as *templates*.

4.1 Template 1: When Are Three Points Aligned?

Given a triangle ABC and a free point D we have the orthogonal projections E, F and G of D onto the three sides of the triangle (Figure 6). When are these three points aligned? Of course, not in general, but the interesting question is: what is the locus set of points D such that points A, B and C are aligned? This is the question proposed to the student in the GeoGebra applet in this template [24]. The idea, as in most dynamic geometry applets, is that the student will explore the changes in the construction for the different positions of some elements (mainly D in this case) in order to generate a conjecture.

In this exercise, the student can find the answer to the question by just pressing a button, but unlike most interactive exercises that provide one solution to one problem, here, by pressing the button, a symbolic computer algebra task is sent to the server so the computational problem is remotely solved in real time for the particular position of the construction in the applet at that moment. This is due to the fact that the application is programmed using JavaScript which allows us to directly read the different elements in the geometric construction using the available GeoGebra JavaScript methods [25].

So, for instance, for the default problem stated above, by pressing the *Show equation of locus* button, we obtain the equation of the circumcircle of triangle ABC, as stated by the Wallace-Simson theorem. By performing the task in real time we are able to obtain the different circumcircle equations corresponding to the different triangles ABC when moved in the applet. For instance, we obtain the equation $6 - 5x + x^2 - 3y + y^2 = 0$ for the default triangle in Figure 6 and $36 - 19x + 3x^2 - 11y + 3y^2 = 0$ if we move point A from $(1, 1)$ to $(2, 2)$ as shown in Figure 7.

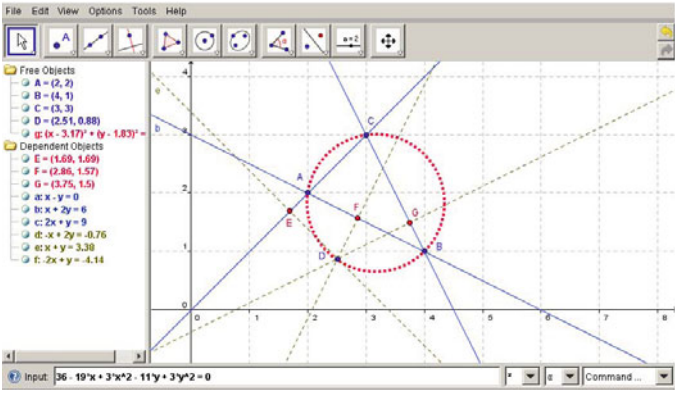


Fig. 7. Locus set of points D such that E, F, G are aligned

But the main point of this approach is that we can completely change the construction in the applet and hence generate different questions. So creating exercises becomes the real exercise for the student. This can be done by making a new construction directly in the applet or loading a previously designed GeoGebra construction. For instance, starting with a triangle ABC and a free point D we can consider the points E, F and G defined, respectively, as the midpoint of DC, midpoint of DA and the orthogonal projection of D onto the orthogonal line to AB through C (see Figure 8). So although the question now is the same, namely *what is the locus set of points D such that E, F and G are aligned?*, the problem is completely different.

In this case, the answer is the line reflection of line AC with respect to the line CF as shown in Figure 9.

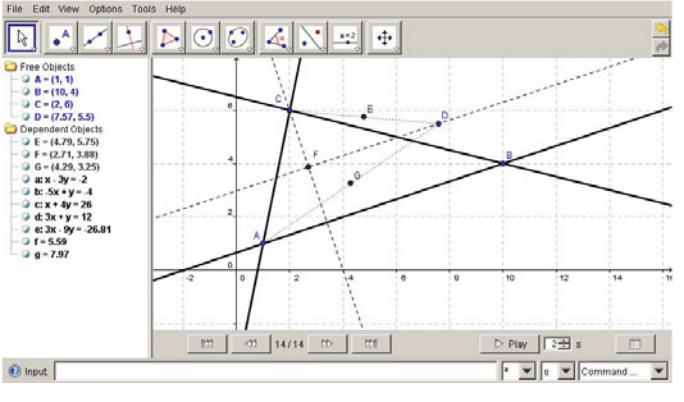


Fig. 8. When are E, F, G aligned?

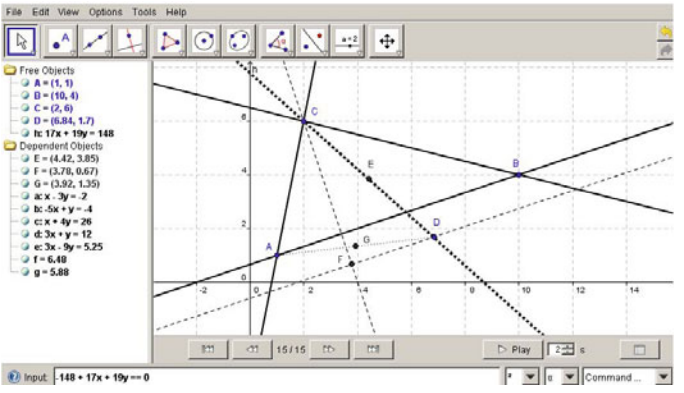


Fig. 9. Locus set of points D such that E, F, G are aligned

The computational kernel of this resource is based on the symbolic capabilities of Mathematica and CoCoA that are used to perform elimination tasks over polynomial ideals using Groebner bases. More precisely, to produce the answer to the requested task from the corresponding GeoGebra configuration, an ad-hoc encoding of the task in terms of points and point-constraints is generated by a JAVA application. Mathematica is then launched initializing variables according to the specifications in the description of the task. An initialization file for CoCoA containing the ideal generated by the appropriate defining polynomials is also written out, and CoCoA, launched by Mathematica, computes a Groebner basis for this ideal. Each generator is factored (a task also done by CoCoA), and a process of logical expansion is performed on the conjunction of the generators in order to remove repeated factors. A standard interpretation of the final polynomial generators directly provides the final answer.

This web application could be easily modified to make Sage its computational kernel so it would be completely independent of the proprietary software Mathematica. However, we have considered it interesting in its current form as an example of application that could benefit from the migration to the open source software Sage that we try to promote with this note.

4.2 Template 2: Flexible Sliding Ladder

In the template 1 above, symbolic capabilities have been added to a GeoGebra applet making possible the solving of a wide variety of advanced discovery tasks. The development of that web application involves advanced programming skills and consequently it is hard to replicate. Moreover, the template 1 makes use of the fact that both the main web page and the CAS are hosted in the same server for which we have administrator rights. The fact that both systems are installed in the same server makes interconnection easy.

In contrast, we present now a web application (template 2 [26]) whose development, besides access to a standard web page, only requires basic knowledge of JavaScript and HTML.

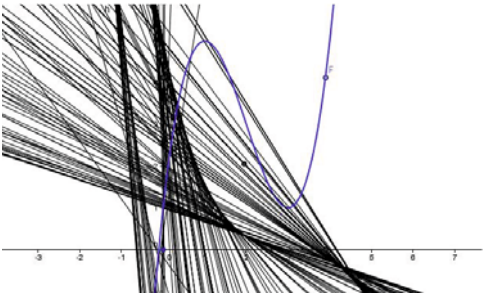


Fig. 10. Family of lines determined by flexible sliding ladder

For the sake of clarity and in order to show how the different elements presented in this paper can be combined in different ways to tackle the same problems, this teaching resource revisits a problem already treated in section 2 above. Namely, the problem consists of finding the envelope of the family of lines determined by the different positions of a flexible ladder defined by the segment joinning the origin and a point in the graph of a general function.

In section 2, a GeoGebra applet with the geometric construction was inserted in a Sage worksheet where all computational tools were readily available. Solving the problem became then an application of calculus and commutative algebra techniques for which the application Singular (included in Sage) was used.

In this template 2, the key point is to make available the same computational capabilities to a GeoGebra applet within a simple web page without installing our own Sage server. This is done interacting with a Sage session over HTTP through what is known as a simple Sage server [27]. Notice that this allows a teacher to emulate this template without the effort of mantaining a Sage server by remotely using a Sage server, ideally one provided by his/her educational institution.

More precisely, besides a GeoGebra applet in which the flexible slidding ladder is constructed, this template includes a text area and three text fields accompanied by instructions. For security reasons, although we can retrieve and show in our web page the information provided by the remote Sage server, we can not have direct access to the data from the web page. A text area is used to make the different answers of Sage to the different requests available to the user. In particular, the final equation for the envelope needs to be manually copied and pasted from the text area in order for it to be displayed in the applet.

The three text fields are used to "talk" with Sage and the applet. First there is a technical field in which the user has to paste the Sage session number as provided in the text area. This is neccessary for subsequent requests to the server, again by security reasons. In the second text field, the user has to indicate the function describing the flexible ladder. Once this is done, the user just has to press a button for Sage to provide the equation of the sought envelope in the text area. Once the equation is shown in the text area, a proccess that takes a few seconds, one just has to copy/paste it in the last text field for its graph to

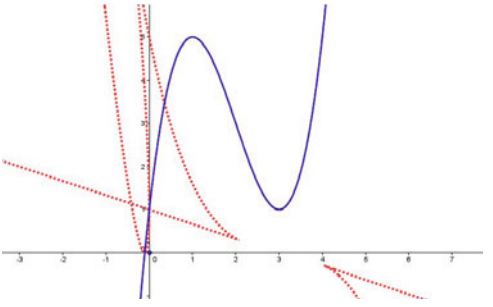


Fig. 11. Envelope of family of lines

be displayed in the GeoGebra applet. Behind all this seemingly simple process there are of course involved computations remotely performed by Sage (Singular in particular).

As an example, the envelope for the family of lines (in Figure 10) determined by the function $f(x) = (x-1)^3 - 3(x-1)^2 + 5$ is given by the following equation of degree 6:

$$\begin{aligned} 135x^6 + 432x^5y + 72x^4y^2 + 4x^3y^3 - 1620x^5 - 3942x^4y - 666x^3y^2 - 36x^2y^3 + \\ 7290x^4 + 13176x^3y + 5031x^2y^2 + 594xy^3 + 27y^4 - 14310x^3 - 15606x^2y - \\ 2646xy^2 - 216y^3 + 9315x^2 - 378xy + 486y^2 + 2430x - 432y + 135 = 0 \end{aligned}$$

whose graph is shown in Figure 11 together with the graph defining the flexible ladder.

5 Conclusions

In this work we show several examples of interaction between applications of quite different nature (DGS and CAS). This conection has a potential importance in educational as well as in research situations.

Moreover, the use of free software tools provides important advantages, among which we can highlight universal access at no cost to the software as well as to documentation.

However, there are still issues. Installing our own Sage server is not very difficult but, as any technology depending on connection, it is a process subject to constant updates and maintenance. On the other hand, due to security reasons, there exist limitations to the connection between web pages and Sage servers, using Javascript. The authors think that the idea of simple Sage server will help in this connection.

In the authors opinion, in order to integrate different systems, it is neccesary to clarify the input/output interfaces. It would be interesting that the systems could write/read information written in some common language. OpenMath [28] would be a good candidate for that. Hence, better than developing big systems, using OpenMath we could obtain great results by combining the power of smaller systems.

Acknowledgements

The first, second and third authors were partially supported by research grant MTM2008-04699-C03-03/MTM from the Spanish MEC.

References

1. Jackiw, N.: The Geometer's Sketchpad v 4.0. Key Curriculum Press, Berkeley (2002)
2. Gao, X.S., Zhang, J.Z., Chou, S.C.: Geometry Expert. Nine Chapters, Taiwan (1998)
3. Laborde, J.M., Bellemain, F.: Cabri Geometry II. Texas Instruments, Dallas (1998)
4. Richter-Gebert, J., Kortenkamp, U.: The Interactive Geometry Software Cinderella. Springer, Berlin (1999)
5. GeoGebra, http://www.geogebra.org
6. Roanes-Lozano, E., Roanes-Macías, E., Villar, M.: A Bridge between Dynamic Geometry and Computer Algebra. Math. Comput. Model. 37(9-10), 1005–1028 (2003)
7. Botana, F., Abánades, M., Escribano, J.: Computing Locus Equations for Standard Dynamic Geometry Environments. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2007. LNCS, vol. 4488, pp. 227–234. Springer, Heidelberg (2007)
8. Abánades, M.A., Escribano, J., Botana, F.: First Steps on Using OpenMath to Add Proving Capabilities to Standard Dynamic Geometry Systems. In: Kauers, M., Kerber, M., Miner, R., Windsteiger, W. (eds.) MKM/CALCULEMUS 2007. LNCS (LNAI), vol. 4573, pp. 131–145. Springer, Heidelberg (2007)
9. Escribano, J., Botana, F., Abánades, M.A.: Adding Remote Computational Capabilities to Dynamic Geometry Systems. Math. Comput. Simulat. 80, 1177–1184 (2010)
10. Abánades, M.A., Botana, F., Escribano, J., Tabera, L.F.: Software matemático libre. Gaceta RSME 12(2), 325–346 (2009)
11. Stein, W.A., et al.: Sage Mathematics Software (Version 4.6.0), The Sage Development Team (2010), http://www.sagemath.org
12. Botana, F., Valcarce, J.L.: A dynamic-symbolic interface for geometric theorem discovery. Comput. Educ. 38(1-3), 21–35 (2002)
13. Botana, F., Valcarce, J.L.: A software tool for the investigation of plane loci. Math. Comput. Simulat. 61(2,1), 139–152 (2003)
14. Sage server, http://wiki.sagemath.org/SageServer
15. Lagrange Multipliers (Two Variables), http://ocw.mit.edu/ans7870/18/18.02/f07/tools/LagrangeMultipliersTwoVariables.html
16. Lagrange-multipliers-applet-interact, https://193.146.36.46:9000/home/pub/4
17. Apostol, T.M., Mnatsakanian, M.A.: A New Look at the So-Called Trammel of Archimedes. Am. Math. Month. 116(2), 115–133 (2009)
18. Decker, W., Greuel, G.M., Pfister, G., Schönemann, H.: Singular 3-1-1 — A computer algebra system for polynomial computations (2010), http://www.singular.uni-kl.de
19. EscaleraFlexibleSaturada, https://193.146.36.46:9000/home/pub/5

20. Capani, A., Niesi, G., Robbiano, L.: CoCoA, a system for doing Computations in Commutative Algebra (1991), http://cocoa.dima.unige.it
21. Botana, F., Valcarce, J.L.: Automatic determination of envelopes and other derived curves within a graphic environment. Math. Comput. Simulat. 67(1-2), 3–13 (2004)
22. Wolfram, S.: Mathematica: a system for doing mathematics by computer, 2nd edn. Addison Wesley Longman Publishing Co., Inc., Redwood City (1991)
23. Lockwood, E.H.: A Book of Curves. Cambridge University Press, Cambridge (1971)
24. Template 1, http://nash.sip.ucm.es/Ggb-direct/template1/Ggb-direct-LAD-3PointsAligned.html
25. GeoGebra JavaScript methods, http://www.geogebra.org/en/wiki/index.php/GeoGebra_Applet_Methods
26. Template 2, http://nash.sip.ucm.es/Ggb-directPlus/Ggb-directPlus-template-flexible-ladder.html
27. Simple Sage server, http://www.sagemath.org/doc/reference/sagenb/simple/twist.html
28. http://www.openmath.org

A Coq-Based Library for Interactive and Automated Theorem Proving in Plane Geometry*

Tuan-Minh Pham[1], Yves Bertot[1], and Julien Narboux[2]

[1] INRIA Sophia Antipolis
[2] LSIIT, Université de Strasbourg, CNRS

Abstract. In this article, we present the development of a library of formal proofs for theorem proving in plane geometry in a pedagogical context. We use the Coq proof assistant [4]. This library includes the basic geometric notions to state theorems and provides a database of theorems to construct interactive proofs more easily. It is an extension of the library of F. Guilhot for interactive theorem proving at the level of high-school geometry [7], where we eliminate redundant axioms and give formalizations for the geometric concepts using a vector approach. We also enrich this library by offering an automated deduction method which can be used as a complement to interactive proof. For that purpose, we integrate the formalization of the area method [3] which was developed by J. Narboux in Coq [12,10].

Keywords: formalization, automation, geometry, Coq, teaching.

1 Introduction

Technological tools are widely used to teach mathematics in schools. Dynamic Geometry Software (DGS) and Computer Algebra Software (CAS) are the two families of software that are well represented. These tools are widely used to explore, experiment, visualize, calculate, measure, find counter examples, conjectures... but most of them can not be used directly to build or check a proof. Proof is a crucial aspect of mathematics and therefore should be integrated more into the mathematical tools. The exploration and proof activities are interlaced. We think that these two activities could be better interlaced if they were both conducted using the computer.

Dynamic geometry systems (DGS) are used more and more to teach geometry in school. They not only allow students to understand construction steps that lead to final drawings, but also provide access to geometric objects, allow students to move free points and see the influence on the rest. Then, students can find out new properties from the drawings. To justify conjectures, some among the numerous DGS provide proving features. These systems are almost

* This work is partially supported by the french ANR project Galapagos.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 368–383, 2011.

all based on algebraic (coordinate-based) methods or semi-algebraic (coordinate-free) methods. Most DGS do not allow users to interactively build traditional proofs. The few DGS which provide interactive theorem proving features are either based on databases of proofs known in advance or on adhoc theorem provers which can not be easily extended. In [13], we give a more precise description of the different systems. We believe that deductive proofs should continue to be an essential part of the high-school curriculum for a geometry course. We also think that the use of a proof assistant to interactively build proofs could help students understand the concepts of deduction. This use has the following advantages:

- It gives clear logical view about the geometric problems. The system makes clear what are the hypotheses and the conclusions. Student can understand the logical inferences used in each reasoning step by observing the change of the proof environment.
- Reasoning steps are verified by the proof assistant, thus constructed proofs have very high level of confidence.
- It allows to combine purely geometric arguments with other kinds of proofs (using complex numbers for instance).

This leads to the necessity of developing a geometry proving tool for high school students which they can use to interactively construct traditional geometry proofs with the support of a proof assistant.

The first step in this direction was done by F. Guilhot, a high school teacher who developed a library in the Coq proof assistant for interactive theorem proving in geometry at the level of high-school [7]. This development is based on a specific axiom system which is adapted to the knowledge of high school students. It covers a large portion of basic notions of plane geometry, properties and theorems. Its proofs and its geometry reasoning are close to what students learn in high school. Some classical theorems are proved in this library, such as Menelaus, Ceva, Desargues, Pythagoras, Simsons' line etc. Then, a tool called GeoView was developped in order to visualize graphically geometry statements using a DGS [2].

Another tool using Coq that was developed by J. Narboux is GeoProof. This tool works in the opposite direction: the DGS is used to generate statements and tactics for Coq. Its proving feature is based on the automated deduction method - the area method. It is not an interactive proving tool [14].

Recently, the first author developed a tool for interactive proof using GeoGebra [6] and Coq [16]. A proof window is added in GeoGebra, it allows to communicate with Coq. The user can draw figures and state conjectures, geometric constructions and conjectures are translated into relations between geometric objects and sent to Coq. A geometric problem is expressed by a set of hypotheses and conclusions. The reasoning steps are interactively built by mouse clicks over geometric objects and hypotheses using appropriate proving methods. The reasoning steps are translated into commands and executed by Coq. The new state is sent back to the user. The changes of hypotheses and conclusions allows the user to continue the proof. We refer readers to [16] for more information about the graphical user interface for this tool.

A crucial part of this tool lies in its library for synthetic geometry in Coq. The library contains formalizations of geometric notions, and of geometric theorems that allow the user to manipulate geometric objects and to produce proofs. Developing a library in the style of the synthetic geometry system is appropriate. In the synthetic geometry, we postulate the existence of geometric objects (e.g. points, lines,...) and geometric relations as primitive notions and build up the geometry from an axiom system that deals with these primitive notions. The geometry proofs are constructed by making use of axioms, theorems and logical arguments. This approach is pedagogically suitable and satisfactory from a logical point of view. In this paper, we focus on this synthetic library and we present why and how we extended [7].

The development by F.Guilhot has three main drawbacks: first it is based on many axioms which are not really needed, second it lacks constructive definitions for the existence of geometric objects, third it lacks automation. Indeed, in this library, each geometric notion is defined by axioms that assert its properties. This leads to an explosion of the number of axioms. Moreover, the library does not provides functions to construct points and lines and compound geometric constructions, it only provides axioms which states the existence of the compound objects. We will detail these drawbacks in section 2. Finally, the library does not include any automatic theorem proving methods. We think that having an automatic theorem proving method is important because:

- During an interactive proof, students can meet minor proof obligations. It they need to solve these proof obligations interactively this can lead to very technical proofs. This is not adapted to their level of abstraction.
- It may be useful to be able to check if a statement is correct in order to help the student or to give better error messages when the student make a mistake.

This motivates us to improve this library. We reduce the number of axioms by providing a more compact axiom system and constructively building up the geometric notions from this system. We also enrich the library with an automated deduction method - the area method [3]. This method gives readable proofs, based on the notions: signed area, ratio of directed segment and Pythagoras difference which are also very close to high school knowledge. For that purpose, we integrate the formalization of the area method in Coq by J.Narboux [12,10].

This article is organized as follows. After giving some discussion about the formalization of F. Guilhot in Section 2, we present in Section 3 our developments to eliminate its redundant axioms and to redefine geometric notions. Section 4 deals with the integration of the area method. The last section is for the conclusion and our perspectives.

Related works. Several axiom systems for synthetic geometry were produced, Hilbert proposed his axiom system [8] with 20 axioms relating two primitive notions: point and line. Another was proposed by Tarski [18] which is simpler, relies on first order logic, and contains only eleven axioms, two predicates, and

one primitive notion: point. The axiom system of Hilbert was formalized in Coq by C. Dehlinger [5] and in Isabelle/Isar by L. Meikle [11]. The one of Tarski was formalized in Coq by J. Narboux [15]. However, systems such as Hilbert's system or Tarski's system have starting points that are too low, hence they can not be used in a pedagogical context. The work closest to ours is the one by P. Scott and J. Fleuriot [17]. This work present the integration of an automatic theorem proving tool in the user interface of a proof assistant. The difference with our work is that the automatic theorem proving tool they integrate is not specialized in geometry. Hence, it can be used in many contexts as for example a low level development about Hilbert axiom system but it may be less efficient for complex geometry theorems.

2 Formalization of High-School Geometry

In this section, we present the development of F.Guilhot and what can be improved. As mentioned, F. Guilhot does not try to provide a system with a minimal number of axioms, nor to provide an automated tool for theorem proving. She only tries to provide a system in which definitions of geometric notions, theorems and geometry reasonings are described as they are taught in high school. She does not build up the whole of Euclidean geometry from a fundamental axiomatic system (such as the systems mentioned above). She uses an alternative approach to the same geometrical results. She first constructs a vector space attached to the affine geometry, then she construct an Euclidean space by adding the notion of scalar product to the affine geometry.

We use small letters a, b, $c\ldots$ to denote real numbers; capital letters A, B, $C\ldots$ to denote points; pairs of a real number and a point in the form aA to denote mass points.

The key technique in her formalization lies in using the universal space proposed by M. Berger [1]. The vector space is extended to the universal space which is a union of points with a given non-zero mass and the vectors in the vector space. The rule $aB - aA = a\overrightarrow{AB}$ is used to convert vectors to the representation in mass points. The vector space property of the universal space is preserved and defined by the following axioms:

- Axiom 1 (Definition of addition) : With $(m+n \neq 0)$, there is a unique point R (called barycenter) on PQ such that $nP + mQ = (m+n)R$.
- Axiom 2 (Definition of vector): With $(m+n=0)$, the sum of $(nP+mQ)$ is a vector $m\overrightarrow{PQ}$.
- Axiom 3 (Idempotent): $nP + mP = (m+n)P$.
- Axiom 4 (Commutative): $nP + mQ = mQ + nP$.
- Axiom 5 (Associative): $nP + (mQ + kR) = (nP + mQ) + kR$.
- Axiom 6 (Definition of scalar multiplication): $k(nP) = (k * n)P$.
- Axiom 7 (Distributivity) $k(nP + mQ) = knP + kmQ$.

These properties form an extension of the theory of mass points (where mass is not limited to a positive real number). So we can completely perform calculations in mass points as we can do with problem-solving technique of mass point

theory. Furthermore, these calculations are taught in high school courses and straightforward.

In Coq, the type of mass points is declared as a record composed of a real number and a point. A variable *add_MP* which takes two arguments of type mass point and gives a mass point is declared for addition operator and a variable *mult_MP* which takes a real number and a mass point as arguments and gives a mass point is declared for scalar multiplication operator.

```
Record MassPoint:Type := cons{number: R; point:Point}.
Variable add_MP  : MassPoint -> MassPoint -> MassPoint.
Variable mult_MP : Real      -> MassPoint -> MassPoint.
```

The system of axioms which defines the universal space is introduced. It allows us to manipulate mass points. A good idea for calculating mass points in Coq is mapping this space to an abstract field structure. This enables us to simplify equations of mass points using automated tactics in Coq library for field (such as `ring_simplify`, `field_simplify`...). It makes calculations easier.

Plane Euclidean geometry can be obtained by equipping affine geometry with the notion of scalar product. The other geometric notions are added step by step using existing formalized notions and their relations in the library.

However, the geometric objects are almost all declared as abstract functions that take points as input, and axioms allow to manipulate these notions. For example, the orthogonal projection of a point C onto the line AB is defined as follows.

```
Variable orthogonalProjection : Point->Point->Point->Point.
Axiom def_orthogonalProjection :
 forall A B C H : Point,
 A <> B ->
 collinear A B H ->
 orthogonal (vect A B) (vect H C) ->
 H = orthogonalProjection A B C.
Axiom def_orthogonalProjection2 :
 forall A B C H : Point,
 A <> B ->
 H = orthogonalProjection A B C ->
 collinear A B H /\ orthogonal (vect A B) (vect H C).
```

Where a variable expresses the orthogonal projection function. The axioms say that if we have collinearity of A, B and H and orthogonality of $\overrightarrow{AB}$ and $\overrightarrow{CH}$ then H is the orthogonal projection of C onto the line AB and vice versa.

Using axioms asserting properties of objects to define them is a usual manner in high school. However, this leads to an explosion of the number of axiom and makes the axiomatic system redundant. Moreover, for compound objects, this kind of definitions are not constructive. The existence of the compound objects is stated by axioms and constructing these objects from existing simpler objects is not clear.

In the next section, we introduce a more compact axiom system, we redefine geometric notions and build up a new geometry system. Redundant axioms in

the older system are put in the form of theorems to be proved. Theorems and properties that are proved in the older system are maintained. New properties corresponding to new definition of notions are introduced.

3 From Affine to 2D-Euclidean Geometry: A Vector Approach

3.1 Axiomatic System

The use of mass points is a good approach for computation, similar to what students know. However, students may feel bored in these computations since they can not visualize objects that they are manipulating. So, we use the notion of vector as the key notion in our development. We will define other geometric notions by relations of vectors.

In the last section, the vector is defined as a sum of two mass points in the special case where the sum of mass equals zero. We use a sub-type of mass point to define it in Coq.

```
Definition isVector (v:MassPoint):= exists A, B :Point,
                                v = add_MP ((-1) A) (1 B).
Record Vector : Type :=
          vecCons { mpOf :MassPoint; proof:isVector mpOf}.
```

The vector data type is constructed from an element *mpOf* having the mass point type with a proof showing that this element can be expressed by the sum of mass point of two certain points A and B with mass values (-1) and 1 respectively. After proving the fact that a linear combination of two mass points satisfying *isVector* predicate also satisfies this predicate, we can easily define the addition operator and the scalar multiplication operator for the vector type. It is clear that we can perform computation over vectors by reusing computations over mass points.

To build up the Euclidean space, we introduce a system of axioms for the scalar product (also called the dot product or the inner product).

Axiom 8_positivity: $\forall\ \overrightarrow{v}$:Vector, $\overrightarrow{v}\cdot\overrightarrow{v} \geqslant 0$
Axiom 9_positivity2: $\forall\ \overrightarrow{v}$:Vector, $\overrightarrow{v}\cdot\overrightarrow{v} = 0$ -> $\overrightarrow{v} = \overrightarrow{0}$
Axiom 10_symmetry: $\forall\ \overrightarrow{u}\ \overrightarrow{v}$:Vector, $\overrightarrow{u}\cdot\overrightarrow{v} = \overrightarrow{v}\cdot\overrightarrow{u}$
Axiom 11_distributivity: $\forall\ \overrightarrow{v_1}\ \overrightarrow{v_2}\ \overrightarrow{v_3}$:Vector, $(\overrightarrow{v_1}+\overrightarrow{v_2})\cdot\overrightarrow{v_3} = \overrightarrow{v_1}\cdot\overrightarrow{v_3}+\overrightarrow{v_2}\cdot\overrightarrow{v_3}$
Axiom 12_homogeneity: $\forall$ (k : R)($\overrightarrow{u}\ \overrightarrow{v}$:Vector), $(k\times\overrightarrow{u})\cdot\overrightarrow{v} = k\times(\overrightarrow{u}\cdot\overrightarrow{v})$

To define the Euclidean plane, we introduce an axiom about the existence of three distinct and non-collinear points and another about co-planarity any fourth point with them. These axioms not only ensure that all given points lie on the same plane, but also allow us to define the orientation for planes. The concept of orientation is important because of its use in the definition of trigonometric functions.

Variable O O1 O2 :Point.
Axiom 13_exist_3_distinct_notcol_points:
$O \neq O1 \wedge O \neq O2 \wedge O \neq O3 \wedge \neg collinear\ OO1O2$.
Axiom 14_coplanarity: $\forall$ (M :Point), $\exists$ k1 k2 :R,
$\overrightarrow{OM} = (k1 \times \overrightarrow{OO1}) + (k2 \times \overrightarrow{OO2})$.

3.2 Overview of the Structure of the Formalization

We construct geometric notions from our primitive notions and prove their properties using our axiom system. Properties and theorems in the library of F. Guilhot are either preserved or reformalized with new definitions of geometric notions. Let's take a look at the Fig. 1 and 2 to have a overview of our formalization.

Figure 1 shows the dependency between concepts in our formalization for affine geometry. Notions in affine geometry are easily formalized from vector. Alignment of three points A, B and C is defined by collinearity of $\overrightarrow{AC}$ and $\overrightarrow{AB}$, parallelism of two vectors is defined by collinearity of them. The other ones are also covered such as midpoint, center of gravity of triangle, parallelogram, ...

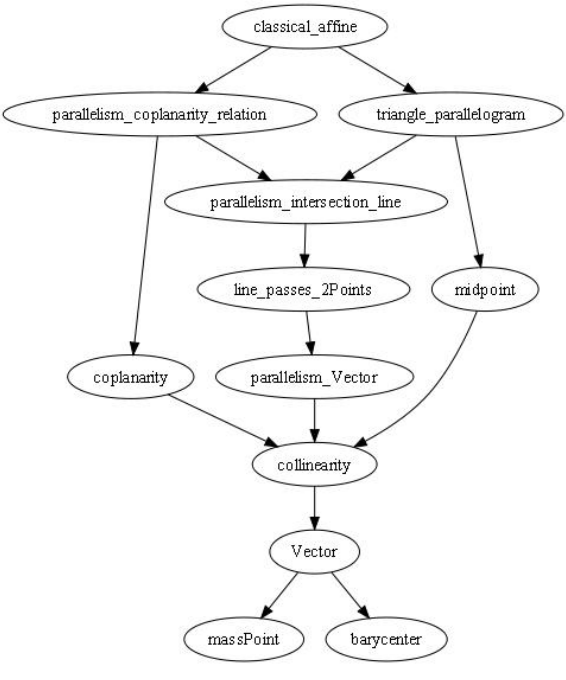


Fig. 1. Formalization of notions in affine geometry

The formalization of plane geometry is presented in Fig. 2, we start with definition of orthogonality of vector and Euclidean distance. Two vectors are

orthogonal if their scalar product equals zero. Euclidean distance of two points A and B is expressed by root square of the scalar product of $\overrightarrow{AB}$ with itself.

```
Definition orthogonal (vec1 vec2 :Vector) :=
  scalarProduct vec1 vec2 = 0.
Definition distance (A B :Point) :=
  sqrt (scalarProduct (vector A B) (vector A B)).
```

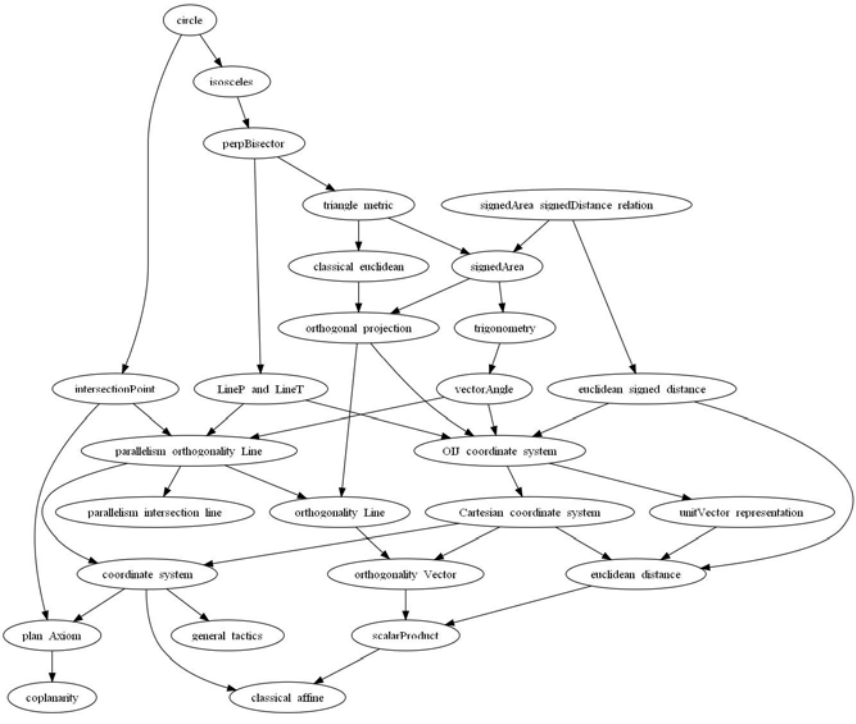


Fig. 2. Formalization of plane geometry

With notions of orthogonality and parallelism of vectors, we can define complete straight line including: line passing through two points, parallel and perpendicular lines. Orthogonal projection and the orthocenter of triangle are also formalized. The Euclidean distance allows us to define unit representation of vector. With the support of orthogonality, we can construct the Cartesian coordinate system. Trigonometric functions and oriented angle are formalized. The signed area and determinant of vectors are also expressed thanks to the trigonometric functions and the angle. The equality of distances allows to define perpendicular bisector, isosceles triangle, circle ...

Our formalization covers a large part of the geometry curriculum in French high school. Many notions are formalized, their properties are verified. They are considered as basic concepts to produce more complex geometric proofs.

3.3 A Constructive Geometry Library

In this section and the following one, we use these notations:

- $\overrightarrow{v_1} \cdot \overrightarrow{v_2}$ for the scalar product of $\overrightarrow{v_1}\overrightarrow{v_2}$
- $|AB|$ for the Euclidean distance from A to B
- $\overrightarrow{v}^{\perp}$ for the orthogonal vector of $\overrightarrow{v}$
- $\overline{AB}$ for the signed distance from A to B
- $\mathcal{S}_{ABC}$ for the signed area of $\triangle ABC$
- $|\overrightarrow{v}|$ for the measure of $\overrightarrow{v}$

In GeoGebra or other DGS, the user can create and manipulate geometric constructions. The user starts with points and simple constructions, then constructs new geometric object from existing ones. So, to match our library with constructions in DGS, we try to provide primitive constructions which are elementary constructions by rules and compass and mentioned in [9]. They include: the point lying on a given line, the midpoint of two given points, the line passing through two given points, the line passing through a given point and perpendicular to a given line; the line passing through a given point and parallel to a given line, the intersection point of two lines, the circle with a given center passing through a given point, the circle with a given diameter,...

These constructions one by one are defined by vector or by relations over existing notions. The line AB is formalized by A and $\overrightarrow{AB}$. The midpoint I of two points A and B is formalized by constructing $\overrightarrow{AI}$ such that $\overrightarrow{AI} = \frac{1}{2} \times \overrightarrow{AB}$. The circle with center O that passes through A is formalized as the set of points M which distance OM equals distance OA...

One interesting formalization here is the one of line. We look for a data structure to express all sorts of line. In Coq, we use a record with a root point and a direction vector to define the line. To ensure existence of line, an element is added to record, this is a proof to show that this direction vector is not the null vector.

```
Record Line:Type := lineCons {rootOf: Point; vecOf: Vector;
 proof: isNotZeroVec vecOf }.
```

The line passing through two points A and B is expressed by a null Line in the case $A = B$, and by the construction of A and $\overrightarrow{AB}$ in the other case. The line which goes through a given point A and which is parallel to another line a (denoted *lineP A a*) is expressed by the construction of A and $\overrightarrow{AB}$ where B is constructed such that $\overrightarrow{AB}$ equals the direction vector of line a. The perpendicular line (denoted *lineT*) is formalized in the same way.

We define also equality of lines by collinearity of their direction vectors for the case that two lines have the same root point, and by collinearity of their direction

vector with the vector constructed by two root points for the case where these points are distinct. Properties concerning equality of lines are verified with this definition, such as:

```
Lemma align_line:
 forall A B C : Point,
 A <> B -> A <> C ->
 collinear A B C -> line A B = line A C.
Lemma liesOnLine_eqLine:
 forall A B C D: Point,
 A <> B -> C <> D ->
 liesOnLine C (line A B) -> liesOnLine D (line A B) ->
 line A B = line C D.
```

For example, to prove the first lemma, we start with the hypotheses $A \neq B$ and $A \neq C$ we have that $\overrightarrow{AB} \neq \overrightarrow{0}$ and $\overrightarrow{AC} \neq \overrightarrow{0}$. By the definition of line $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are direction vectors of line AB and line AC respectively. By the definition of collinearity we have that $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are collinear. So $line\ AB = line\ AC$ by the definition of equality of lines for the case that lines have the same root point.

For compound constructions, we formalize them by the way that they are built from primitive constructions. For example, instead of defining the orthogonal projection with three given points by axioms as mentioned in Section 3, we define an orthogonal projection of a given point C onto a given line a. We construct the line that passes through C and which is perpendicular with a, we prove that this line and a are not parallel, then we get the intersection point of them.

```
Definition orthogonalProjection(C :Point) (a :Line) :=
 intersectionPoint (lineT C a) a.
```

As we said above, students are familiar with definition of geometric objects by axioms about their properties. So, to avoid losing the pedagogical meaning of the library and to verify if objects are well formalized, we keep the former axioms in the form of theorems, and prove them. Each object is accompanied by a pair of theorems. One allows us to get properties from its definition, the other is to get a definition from properties.

```
Lemma orthogonalProjection_Properties :
 forall A B C H :Point,
 A <> B->
 H = orthogonalProjection C (line A B) ->
 collinear A B H /\ orthogonal (vect A B) (vect H C).
Lemma properties_orthogonalProjection :
 forall A B C H : Point,
 A <> B ->
 collinear A B H ->
 orthogonal (vect A B) (vect H C) ->
 H = orthogonalProjection C (line A B).
```

Proving theorems of the first kind is simpler than proving theorems of the second kind. The latter usually lead to the proof of uniqueness property of objects. Many compound constructions are introduced: the perpendicular bisector of two given points, orthogonal projection of a point onto a line, the circumcircle of three given points, the center of circumcircle, the orthocenter, the center of gravity,...

Reformalizing the geometric objects allows us to eliminate a lot of redundant axioms. Remaining axioms are used to introduce other geometric notions such as: the parallelism of lines, the perpendicularity of lines, the signed area of vectors or of triangle, the vector angles, the trigonometric functions,... These notions are essential for the library manipulating geometric objects and stating geometric problems. Except for vector angles and the trigonometric functions, the rest are formalized without much effort, and many axioms are eliminated.

4 Integration with the Area Method

Our library contains a large number of geometric notions and propositions. This allows us to prove many geometry theorems. Constructing a traditional proof consists in finding a sequence in logical steps to the conclusion. But in the context of education, a drawback of constructing fully traditional proofs is that, as mentioned in the introduction section, there are minor goals in interactive proving which are necessary to complete the formal proof but which lead to tedious steps and are not adapted to the level of abstraction at which we usually work with students. Hence we integrate a coordinate-free automatic deduction method with a library for interactive proof. This improves the power and does not decrease pedagogical meaning of the library. The coordinate-free automatic deduction method we chose is the area method of Chou, Gao and Zhang. This method consists in expressing the goal to be proved using three geometric quantities (the signed area of a triangle, the signed distance and the Pythagoras difference), and eliminating the points from the goal in the reverse order of their construction using some elimination lemmas.

To integrate the area method in our development, we need first to ensure the correctness of all the elimination lemmas of the area method and second we need to create a mapping between the definitions of the geometric constructions of the two systems.

4.1 Correctness of the Area Method

The first work in the process of integration is to ensures the correctness of this method in our library. Because the area method is constructed based on its own axiom system mentioned in Table 1, we need only to verify these axioms. Before proving them, we have to make a mapping of primitive notions of this method into our library. In fact, this method only has two primitive notions, being the signed area and the signed distance. Others notions such as the Pythagoras difference, the parallelism, the perpendicularity, the collinearity..., as well as geometric constructions, are defined from these two primitive notions.

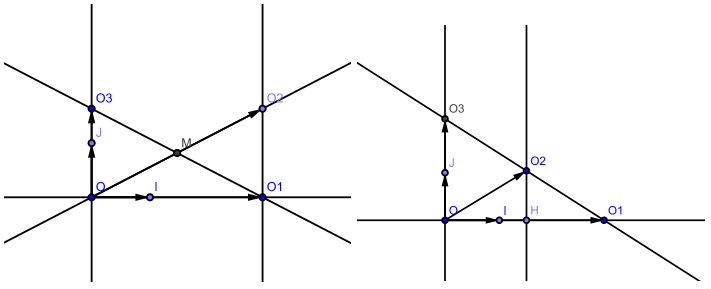
Table 1. The axiom system for the area method

1. $\overline{AB} = 0$ if and only if the points A and B are identical
2. $\mathcal{S}_{ABC} = \mathcal{S}_{CAB}$
3. $\mathcal{S}_{ABC} = -\mathcal{S}_{BAC}$
4. If $\mathcal{S}_{ABC} = 0$ then $\overline{AB} + \overline{BC} = \overline{AC}$ (Chasles' axiom)
5. There are points A, B, C such that $\mathcal{S}_{ABC} \neq 0$ (dimension; not all points are collinear)
6. $\mathcal{S}_{ABC} = \mathcal{S}_{DBC} + \mathcal{S}_{ADC} + \mathcal{S}_{ABD}$ (dimension; all points are in the same plane)
7. For each element r of F, there exists a point P, such that $\mathcal{S}_{ABP} = 0$ and $\overline{AP} = r\overline{AB}$ (construction of a point on the line)
8. If $A \neq B, \mathcal{S}_{ABP} = 0, \overline{AP} = r\overline{AB}, \mathcal{S}_{ABP'} = 0$ and $\overline{AP'} = r\overline{AB}$, then $P = P'$ (uniqueness)
9. If $PQ \parallel CD$ and $\frac{\overline{PQ}}{\overline{CD}} = 1$ then $DQ \parallel PC$ (parallelogram)
10. If $\mathcal{S}_{PAC} \neq 0$ and $\mathcal{S}_{ABC} = 0$ then $\frac{\overline{AB}}{\overline{AC}} = \frac{\mathcal{S}_{PAB}}{\mathcal{S}_{PAC}}$ (proportions)
11. If $C \neq D$ and $AB \perp CD$ and $EF \perp CD$ then $AB \parallel EF$
12. If $A \neq B$ and $AB \perp CD$ and $AB \parallel EF$ then $EF \perp CD$
13. If $FA \perp BC$ and $\mathcal{S}_{FBC} = 0$ then $4\mathcal{S}_{ABC}^2 = \overline{AF}^2\overline{BC}^2$ (area of a triangle)

The crucial thing in the formalization of these two basic notions is constructing three points O, I and J that form a Cartesian coordinate system. With the support of these points, we can define the two primitive notions of the area method in our system.

O, I and J form a Cartesian coordinate system, in other words, they satisfy $\overrightarrow{OI} \cdot \overrightarrow{OJ} = 0$ (or $\overrightarrow{OI} \perp \overrightarrow{OJ}$), $|OI| = 1$ and $|OJ| = 1$. The axiom 13 gives us the existence of three non-collinear points O, O_1 and O_2. The axiom system for mass point (the key here is axiom 1 for the barycenter) allows us to produce a new point C from two given points A, B and a real number k such that $\overrightarrow{AC} = k \times \overrightarrow{AB}$. We explain how to construct O, I, J from O, O_1, O_2 using this rule for producing new points.

We first construct a point O_3 such that $\overrightarrow{OO_1} \perp \overrightarrow{OO_3}$ and $O \neq O_3$. Let's consider two configurations of O, O_1, O_2 in Fig. 3. For the first case $\overrightarrow{OO_1} \perp \overrightarrow{O_1O_2}$, we first construct M such that $\overrightarrow{OM} = \frac{1}{2} \times \overrightarrow{OO_2}$, we then construct O_3 such that $\overrightarrow{O_1O_3} = 2 \times \overrightarrow{O_1M}$. M is the midpoint of OO_2 and O_1O_3, so $\overrightarrow{OO_3} \parallel \overrightarrow{O_1O_2}$. From the hypothesis $\overrightarrow{OO_1} \perp \overrightarrow{O_1O_2}$ we have $\overrightarrow{OO_3} \perp \overrightarrow{OO_1}$. For the second case $\overrightarrow{OO_1} \not\perp \overrightarrow{O_1O_2}$, we construct H as the orthogonal projection of O_2 on OO_1. O_3 is constructed from O_1, O_2 by $\overrightarrow{O_1O_3} = \frac{|O_1O|}{|O_1H|} \times \overrightarrow{O_1O_2}$. Thanks to an extension of the Thales theorem for parallel lines in our library, we can prove $\overrightarrow{OO_3} \parallel \overrightarrow{HO_2}$ from $\frac{|O_1H|}{|O_1O|} = \frac{|O_1O_2|}{|O_1O_3|}$. From the definition of H, we get $\overrightarrow{O_2H} \perp \overrightarrow{OO_1}$, therefore $\overrightarrow{OO_3} \perp \overrightarrow{OO_1}$. $O \neq O_3$ in the two cases is easily verified thanks to distinctions of O, O_1, O_2 in the axiom 13.



(a) $OO_1 \perp O_1O_2$ **(b)** $OO_1 \not\perp O_1O_2$

Fig. 3

Now we construct I from O, O_1 and J from O, O_3 by $\overrightarrow{OI} = \frac{1}{|OO_1|} \times \overrightarrow{OO_1}$ and $\overrightarrow{OJ} = \frac{1}{|OO_3|} \times \overrightarrow{OO_3}$ respectively. We have $O \neq O_1$ (axiom 13), hence $\overrightarrow{OI}$ is a unit representation of $\overrightarrow{OO_1}$, and we have $|OI| = 1$. By the same way, we have $\overrightarrow{OJ}$ is a unit representation of $\overrightarrow{OO_3}$, and we have $|OJ| = 1$. By the construction of O_3) explained above, we have $\overrightarrow{OO_1} \perp \overrightarrow{OO_3}$, so we can prove $\overrightarrow{OI} \perp \overrightarrow{OJ}$ □.

Three points O, I, J offer an orientation for the Euclidean plane. This is used to define sign of the signed distance $\overline{AB}$. The magnitude of $\overline{AB}$ is defined by their Euclidean distance $|AB|$. The definition is as follows:

$$\begin{cases} \text{if } \overrightarrow{AB} \cdot \overrightarrow{OI} > 0 & \overline{AB} = |AB| \\ \text{if } \overrightarrow{AB} \cdot \overrightarrow{OI} < 0 & \overline{AB} = -|AB| \\ \text{if } \overrightarrow{AB} \cdot \overrightarrow{OI} = 0 & \begin{cases} \text{if } \overrightarrow{AB} \cdot \overrightarrow{OJ} > 0 & \overline{AB} = |AB| \\ \text{if } \overrightarrow{AB} \cdot \overrightarrow{OJ} < 0 & \overline{AB} = -|AB| \\ \text{if } \overrightarrow{AB} \cdot \overrightarrow{OJ} = 0 & \overline{AB} = 0 \end{cases} \end{cases}$$

On the other hand, the Cartesian coordinate system with the new points O, I, J enables us to construct the orthogonal vector for the given vector.

$\overrightarrow{v}^{\perp} := (-\overrightarrow{v} \cdot \overrightarrow{OJ}) \times \overrightarrow{OI} + (\overrightarrow{v} \cdot \overrightarrow{OI}) \times \overrightarrow{OJ}$.

We can then formalize the trigonometric functions of two vectors and the signed area of a triangle as follows:

$\cos \overrightarrow{v_1}\overrightarrow{v_2} = \frac{\overrightarrow{v_1} \cdot \overrightarrow{v_2}}{|\overrightarrow{v_1}| \times |\overrightarrow{v_2}|}$ and $\sin \overrightarrow{v_1}\overrightarrow{v_2} = \frac{\overrightarrow{v_1}^{\perp} \cdot \overrightarrow{v_2}}{|\overrightarrow{v_1}^{\perp}| \times |\overrightarrow{v_2}|}$

$\mathcal{S}_{ABC} = \frac{1}{2} \times |AB| \times |AC| \times \sin \overrightarrow{AB}\overrightarrow{AC}$

Two primitive notions of the area method are formalized, many properties are also proved. They are properties of the signed distance $\overline{AB} = k \times \overline{CD} \leftrightarrow \overline{AB} = k \times \overline{CD} \wedge \overrightarrow{AB} \parallel \overrightarrow{CD}$, *col A B C* $\rightarrow \overline{AB} + \overline{BC} = \overline{AC}$; properties of the orthogonal vector $(\overrightarrow{v_1} + \overrightarrow{v_2})^{\perp} = \overrightarrow{v_1}^{\perp} + \overrightarrow{v_2}^{\perp}$, $(k \times \overrightarrow{v})^{\perp} = k \times \overrightarrow{v}^{\perp}$, $|\overrightarrow{v}^{\perp}| = |\overrightarrow{v}|$, $\overrightarrow{v_1} \cdot \overrightarrow{v_2}^{\perp} = -\overrightarrow{v_1}^{\perp} \cdot \overrightarrow{v_2}$; as well as properties of signed area $\mathcal{S}_{ABC} = \frac{1}{2} \times \overrightarrow{AB}^{\perp} \cdot \overrightarrow{AC}$, $\mathcal{S}_{ABC} = \mathcal{S}_{BCA}$, $\mathcal{S}_{ABC} = -\mathcal{S}_{ACB}$... They serve as basic properties to prove the axiom system in Table 1. We do not intend to detail long proofs, we give only the following proof of the property 6: $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} + \mathcal{S}_{DCA} = \mathcal{S}_{ABC}$ presented with

two column method as an example. The full Coq proofs consists of approximately 1500 lines of Coq tactics.

Property	Reasoning
1. $\mathcal{S}_{DAB} = \mathcal{S}_{BDA} = \frac{1}{2} \times \overrightarrow{BD}^{\perp} \cdot \overrightarrow{BA}$	signed area props
2. $\mathcal{S}_{DBC} = \mathcal{S}_{BCD} = \frac{1}{2} \times \overrightarrow{BC}^{\perp} \cdot \overrightarrow{BD}$	signed area props
3. $\frac{1}{2} \times \overrightarrow{BC}^{\perp} \cdot \overrightarrow{BD} = -\frac{1}{2} \times \overrightarrow{BC} \cdot \overrightarrow{BD}^{\perp}$	orthor vect props
4. $-\frac{1}{2} \times \overrightarrow{BC} \cdot \overrightarrow{BD}^{\perp} = \frac{1}{2} \times \overrightarrow{CB} \cdot \overrightarrow{BD}^{\perp}$	orthor vect props
5. $\mathcal{S}_{DBC} = \frac{1}{2} \times \overrightarrow{CB} \cdot \overrightarrow{BD}^{\perp}$	from (2) (3) (4)
6. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times \overrightarrow{BD}^{\perp} \cdot \overrightarrow{BA} + \frac{1}{2} \times \overrightarrow{CB} \cdot \overrightarrow{BD}^{\perp}$	from (1) (5)
6. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times \overrightarrow{BD}^{\perp} \cdot (\overrightarrow{BA} + \overrightarrow{CB})$	distrib prop
7. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times \overrightarrow{BD}^{\perp} \cdot \overrightarrow{CA}$	$(\overrightarrow{BA} + \overrightarrow{CB}) = \overrightarrow{CA}$
8. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times (\overrightarrow{BC} + \overrightarrow{CD})^{\perp} \cdot \overrightarrow{CA}$	$\overrightarrow{BD}^{\perp} = \overrightarrow{BC} + \overrightarrow{CD}$
9. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times (\overrightarrow{BC}^{\perp} \cdot \overrightarrow{CA} + \overrightarrow{CD}^{\perp} \cdot \overrightarrow{CA})$	$\overrightarrow{BD}^{\perp} = \overrightarrow{BC} + \overrightarrow{CD}$
10. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \frac{1}{2} \times (-\overrightarrow{CB}^{\perp} \cdot \overrightarrow{CA} + \overrightarrow{CD}^{\perp} \cdot \overrightarrow{CA})$	$\overrightarrow{BC}^{\perp} = -\overrightarrow{CB}^{\perp}$
11. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = -\mathcal{S}_{CBA} + \mathcal{S}_{CDA}$	signed area props
12. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} = \mathcal{S}_{CAB} - \mathcal{S}_{CAD}$	signed area props
13. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} + \mathcal{S}_{CAD} = \mathcal{S}_{CAB}$	signed area props
14. $\mathcal{S}_{DAB} + \mathcal{S}_{DBC} + \mathcal{S}_{DCA} = \mathcal{S}_{ABC}$	signed area props
Qed.	

4.2 Usability of the Area Method

This method consists in expressing the goal to be proved using three geometric quantities (the signed area of a triangle, the signed distance and the Pythagoras difference), and eliminating the points from the goal in the reverse order of their construction using some elimination lemmas. It deals with problems stated in terms of sequences of specific geometric construction steps. So to make it runnable in our library, we have to convert geometry statements in our library to its statements. Precisely, from sequences of geometric constructions of our library, we have to construct sequences in the area method.

The first step in this process is to normalize the constructions in our library. Compound constructions are unfolded and replaced by sequences of primitive constructions. These sequences are also reduced without loosing their semantics. For example, a line passing through A that is parallel with perpendicular bisector of BC is simplified to a line passing through A that is perpendicular to line BC.

In the second step, we try to extract constructions of the area method from the sequence of constructions we get in the first step. Let's consider constructions of the area method. In fact, each construction aims to create a geometric object with a precise semantics. For example, *on_inter_line_perp Y R U V P Q* is a construction in the area method and defined by

```
Definition on_inter_line_perp (Y R U V P Q : Point):=
  Col Y U V /\ perp Y R P Q /\ ~ perp P Q U V.
```

It means that Y is at the intersection of UV and the perpendicular to PQ going through R. With the semantics of construction, we can construct a sequence of primitive constructions of our library that produces this object. For each constructions, a lemma is introduced to ensure that sequences of primitive constructions give us exactly the object defined by this construction. In our example, the corresponding sequence of the construction *on_inter_line_perp Y R U V P Q* is *Y = intersectionPoint (lineT R (line P Q)) (line U V)*. The lemma that ensures exactness is as follows:

```
Lemma constr_on_inter_line_perp:
  forall Y R U V P Q : Point,
  P <> Q -> U <> V ->
  ~ perpendicular (line P Q) (line U V) ->
  Y = intersectionPoint (lineT R (line P Q)) (line U V) ->
  on_inter_line_perp Y R U V P Q .
```

To complete this step, we write tactics to automatically introduce constructions of the area method when their corresponding sequence of primitive constructions appears in hypotheses. The following tactic is for our example:

```
Ltac convert_to_AMConstructions_12 :=
  repeat match goal with
   | H: ?Y =
  intersectionPoint (lineT ?R (line ?P ?Q)) (line ?U ?V)|- _
  => try (assert (on_inter_line_perp Y R U V P Q) by
    (apply (@constr_on_inter_line_perp Y R U V P Q);
    auto with geo);revert H)
  end;
  intros.
```

5 Conclusion

Our development provides a library in Coq for interactive and automated geometry theorem proving. As far as we know, this is the first system which integrate both *interactive* and *automatic* theorem proving in geometry in a formal setting. We give an axiom system, from which we build up Euclidean plane geometry. We then verify the axiom system of the area method and integrate this method into our library. In this library, geometric objects are formalized in a constructive manner. Proof of properties, lemmas, and theorems are formal and traditional proofs, they are adapted to student knowledge. The combination with the automated deduction method offers flexibility and power in proving geometry theorem. The integration of this library in a dynamic geometry tool offers advantages in education. It allows student to build proofs themselves. In addition, it highlights the structure of the proofs. Student have a logical view of geometry problems and understands what are the hypotheses and the goal. This work could also be used to study the combination of different automatic theorem proving methods in geometry.

In the future, we will study the integration of algebraic approaches to automatic theorem proving in geometry such as Gröbner bases and Wu's method. To be useful in high-school, we will also need to extend our proof system to allow proofs which look less formal. For that purpose we will study how to deal automatically with degenerated cases.

References

1. Berger, M., Pansu, P., Berry, J.P., Saint-Raymond, X.: Problems in Geometry. Springer, Heidelberg (1984)
2. Bertot, Y., Guilhot, F., Pottier, L.: Visualizing Geometrical Statements with GeoView. Electronic Notes in Theoretical Computer Science 103, 49–65 (2003)
3. Chou, S.C., Gao, X.S., Zhang, J.Z.: Machine Proofs in Geometry. World Scientific, Singapore (1994)
4. Coq development team: The Coq Proof Assistant Reference Manual, Version 8.3. TypiCal Project (2010), http://coq.inria.fr
5. Dehlinger, C., Dufourd, J.-F., Schreck, P.: Higher-Order Intuitionistic Formalization and Proofs in Hilbert's Elementary Geometry. In: Richter-Gebert, J., Wang, D. (eds.) ADG 2000. LNCS (LNAI), vol. 2061, pp. 306–324. Springer, Heidelberg (2001)
6. Geogebra development team: Introduction to GeoGebra, http://www.geogebra.org/book/intro-en/
7. Guilhot, F.: Formalisation en Coq et visualisation d'un cours de géométrie pour le lycée. TSI 24, 1113–1138 (2005) (in french)
8. Hilbert, D.: Les fondements de la géométrie. In: Gabay, J. (ed.) Edition critique avec introduction et compléments préparée par Paul Rossier. Dunod, Paris (1971)
9. Janičić, P.: Geometry construction language. Journal of Automated Reasoning 44, 3–24 (2010)
10. Janicic, P., Narboux, J., Quaresma, P.: The Area Method: a Recapitulation. Journal of Automated Reasoning (2010)
11. Meikle, L., Fleuriot, J.: Formalizing Hilbert's Grundlagen in Isabelle/Isar. In: Theorem Proving in Higher Order Logics, pp. 319–334 (2003)
12. Narboux, J.: A Decision Procedure for Geometry in Coq. In: Slind, K., Bunker, A., Gopalakrishnan, G.C. (eds.) TPHOLs 2004. LNCS, vol. 3223, pp. 225–240. Springer, Heidelberg (2004)
13. Narboux, J.: Toward the use of a proof assistant to teach mathematics. In: Proceedings of the 7th International Conference on Technology in Mathematics Teaching, ICTMT 7 (2005)
14. Narboux, J.: A graphical user interface for formal proofs in geometry. J. Autom. Reasoning 39(2), 161–180 (2007)
15. Narboux, J.: Mechanical Theorem Proving in Tarski's Geometry. In: Botana, F., Recio, T. (eds.) ADG 2006. LNCS (LNAI), vol. 4869, pp. 139–156. Springer, Heidelberg (2007)
16. Pham, T.-M., Bertot, Y.: A combination of a dynamic geometry software with a proof assistant for interactive formal proofs. In: UITP Workshops 2010 (2010)
17. Scott, P., Fleuriot, J.: Idle time discovery in geometry theorem proving. In: Proceedings of ADG 2010 (2010)
18. Tarski, A.: What is Elementary Geometry?. In: Henkin, L., Suppes, P., Tarski, A. (eds.) The axiomatic Method, with special reference to Geometry and Physics, pp. 16–29. North-Holland, Amsterdam (1959)

Teaching Geometry with TutorMates

María José González, Julio Rubio, Tomás Recio,
Laureano González-Vega, and Abel Pascual*

Cantabria University, La Rioja University, Addlink Research, Spain
http://www.tutormates.com

Abstract. TutorMates is a new educational software designed to combine symbolic and geometric tools to teach and learn mathematics at Secondary Education level. Its main technical ingredients are: a Computer Algebra system (Maxima) plus a dynamic geometry and graphical tool (GeoGebra), and an intermediary layer that connects the user interface and the computation engine. The role of this intermediary layer is to adapt the computation kernel to the elementary mathematical needs of the expected users. All these components are adapted to the mathematics curricula of 12 to 18 years old students. In this paper, we describe how TutorMates combines and adapts these complex technological tools in order to improve the computer-aided teaching of geometry.

Keywords: Educational software, symbolic and geometric computation, geometry teaching.

1 Introduction

TutorMates (http://www.tutormates.es) is a research and development project, led by Addlink Research [9], in which a multidisciplinary team of researchers from the Universities of Cantabria and La Rioja (Spain) take an active part. Its main goal is to provide a set of mathematic and informatics' tools integrated in a unique learning environment for students and secondary level mathematics teachers.

There is indeed a wide variety of didactic material based on different technologies, which have been developed to support the teaching and learning of mathematics. However, most of them either provide collections of activities that can be solved with the help of symbolic and numeric packages like Mathematica or Maple, or give access to the user to an assortment of commands, as in a kind of extended calculator with which he can accomplish mathematical activities, or develop mathematics lessons that, although supported by interactive applets, are given as a finished product that the student has to follow in a sequential way, from the beginning to the end, and which is adopted by the teacher without directly taking part on it.

Each one of these materials has a different technological foundation which is highly dependent on the nature of the considered mathematical topic: numerical topics need a specific treatment, which is different from the one needed to

* Partially suported by PET2008_0329.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 384–398, 2011.

make graphical representations of functions; the programs that treat geometric topics manage geometric properties invariants from the dragging viewpoint; statistical topics are considered in programs that manage statistic representations (tables, diagrams, etc). Software programs capable to include all this variety are the powerful scientific and numeric computing tools, which are well adapted to research or professional needs. However, when focusing on in mathematics education learning processes for 12 to 18 years old students, the power of these systems can be seen as a disadvantage. Students in these levels usually have difficulties to manage the syntax of the different commands and to interpret the answers given by the system.

However, mathematics education at these levels cannot ignore the development of technologies that are particularly well adapted to the new educational demands. The current mathematics curriculum in the Spanish context proposes to overcome the approach to mathematics as the learning of calculation routines, in order to focus on problem solving processes that require high order skills, such as modelling, interpretation, justification, generalization, etc. These processes require the reorganization of the working time of the students who, instead of learning calculation and graphical routines, have to interpret the results that they can now obtain in a few seconds using a mathematical computing package. But it is necessary to have materials that adapt the full potential of these systems to the cognitive development of the students.

TutorMates has been conceived with this purpose. It is a complete educational software that gathers in the same environment the basic elements of a mathematics lesson -theoretical elements, exercises, problems and evaluation- together with an adaptation of the corresponding numeric, symbolic and geometric commands. All these elements are combined in a flexible structure that allow the material to be easily adapted to different teaching and learning styles.

In this paper, Section 2 is devoted to describe TutorMates' characteristics, highlighting three ideas: the implementation of lessons in TutorMates, the role of the technology in this implementation and the search of the maximum flexibility and adaptability of these lessons to the curricular context. In Section 3, we give details on TutorMates' technical design and architecture. We devote Section 4 to show a particular example of the use of TutorMates to teach a geometry topic of the secondary education level. We finish with a section of conclusions, in which we reflect on the incorporation of automatic theorem proving tools in TutorMates' future development.

2 TutorMates Description

From the variety of didactical materials that have been developed for secondary level mathematics education, only a few of them have really carried out an adjustment of the technology to the curricular context of the students. This adjustment has to be made from different points of view: it is important to adapt the technology to the curricular orientations, taking into account the specificity of every mathematical topic and the competences that are being developed; it

is necessary to adapt the technological potential to the cognitive level of the students; it is also important to observe the students' habits of study as well as the teachers habits of teaching.

TutorMates has been developed taking into account all these factors. Therefore, its potential rests on three bases:

- The design of lessons adapted to every topic of the mathematics curriculum.
- The incorporation, in the above mentioned lessons, of a computer algebra system (Maxima [3]), a dynamic geometry program (GeoGebra [7]) and an environment of statistical computation (JFreeChart library [8]) by means of a mediated access tool that adapts the communication with the system to the cognitive level of the user.
- The flexibility of use, both for the student and the teacher, thanks to internal management tools (Digital Notebook) and external ones (connection to Moodle).

2.1 Mathematics Lessons in TutorMates

TutorMates is organized around mathematical topics that have been selected following the structure of the current curriculum [2]. Every lesson contains theoretical content, examples, exercises, problems and self-evaluation tests, following what we might consider to be an academic standard scheme. Nevertheless, as we will see afterwards, the treatment that is given to these elements by means of a variety of technological resources, does not limit its use to traditional teaching schemes, but opens a wide range of didactic possibilities.

The theoretical content of a lesson in TutorMates provides, in a brief and concise form, the main notions of the lesson, which are usually presented by means of examples. In Figure 1 we can see one of the theoretical screens of TutorMates in the topic *Data Analysis* for 15 years old students. In particular, the theoretical items incorporate random generators of examples, with the purpose that the student looks up as many examples as needed. In addition, TutorMates' structure presents every example in its varied forms of representation. In the left part of the screen shown in Figure 1, we can see a concise explanation of the notion of *bar graph*, followed by an example and a random generator of statistical samples. This generator provides the data in a table form (visualized in Figure 1) and in a bar graph form (hidden in Figure 1) in the right side of the screen.

The mathematical activities included in TutorMates are classified in two levels of difficulty: exercises and problems. Exercises correspond to the basic procedures that the student must learn. Most of the exercises are accompanied by generators of data that allow us to repeat the structure of the exercise with a variety of examples adapted to the students' level. Problems establish higher cognitive demands. They have been selected observing, as main criteria, the usefulness of the topic in practice and the consideration of reasoning processes, search for regularities, modelling, etc. Problems are the basis to develop the mathematical competences that guide the current curricula.

Every topic, depending on its nature (numerical, algebraic, geometric, functional or statistical) relies on a Computation Zone that provides the user with

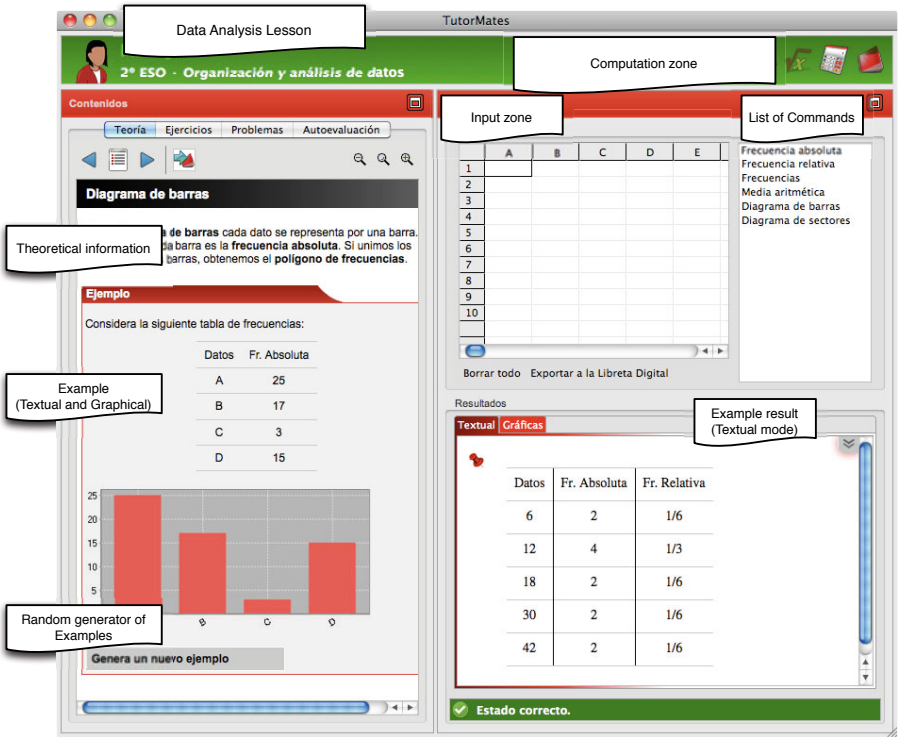


Fig. 1. Example of screen in TutorMates with a random generator of statistical samples

the commands needed to produce interactivity in the resolution of the activities. This zone includes, in particular, the input zone, which provides data to the commands; these data come either from some of the many data generators included in TutorMates or are introduced by the user with the TutorMates Expression Editor. We can see these elements in the upper right side of Figure 1, where the input zone is a table, as corresponds to the emphData Analysis topic. In Sections 2.2 and 3.2 we give more details on this Computation Zone, which has been a basic element in TutorMates' development. Thanks to this zone, TutorMates is not only a set of well published mathematical notes connected between them in intuitive way. In addition, they are conceived as structures that promote the learning based in experimentation.

Finally, every topic incorporates a set of evaluation activities. The system uses these activities to generate random Self-Evaluation Tests. These tests can be automatically marked by the system, which also reports the mistakes and informs the student of the elements to be revised.

2.2 Incorporating Technology in Mathematics Lessons

TutorMates has been designed to take advantage of the educational potential of the scientific software that is being used to teach mathematics in higher

levels. More precisely, TutorMates uses the computer algebra system Maxima, the educational software GeoGebra and the library JFreeChart. By means of these elements:

- We have adapted the Computation Zone to the nature of every mathematical topic, in order to complement the structure of theory, exercises, problems and self-evaluation of every topic.
- We have developed a tool, denominated the Intermediary Layer, with the purpose to adapt to the users' cognitive level the type of examples, data, messages and answers admitted or returned by the system.
- We have designed and incorporated generators of random examples and exercises in every lesson.

For example, in Figure 2 we can see the list of commands corresponding to the lesson *Radicals* for 15 years old students: *Simplify, Compute, Estimate, Are equal?, Factorize*. These commands correspond to the mathematical procedures that our curricula determine. They have been developed using Maxima and both, its input and output, are managed by means of the Intermediary Layer. In Figure 2, we can see that when a 15 years old student introduces a negative root and demands its value using the command *Compute*, he obtains a warning message indicating that the demanded solution does not exist in the set of real numbers.

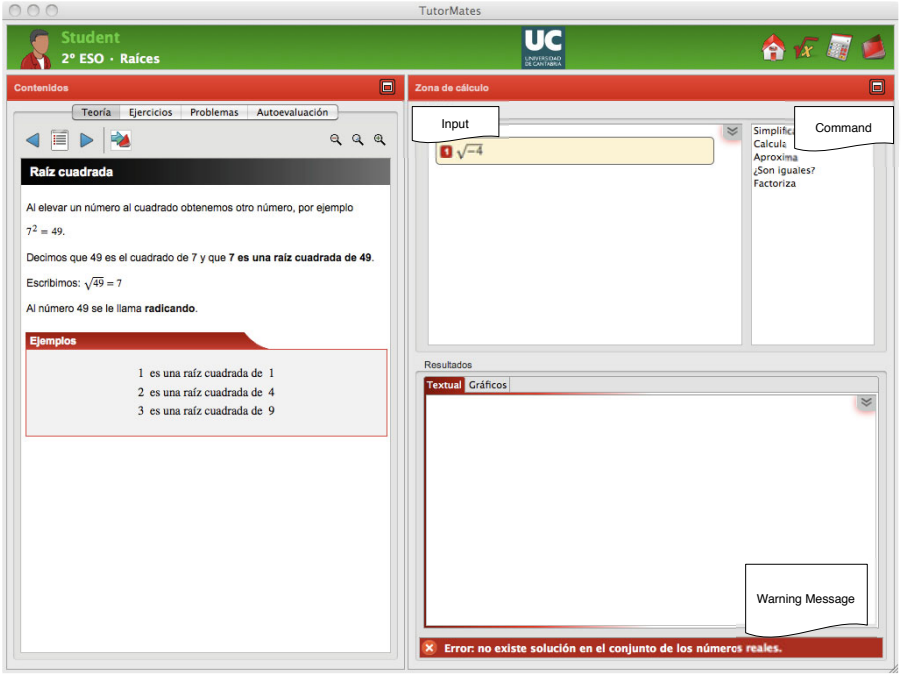


Fig. 2. List of commands in the Computation Zone of a numerical topic and warning message

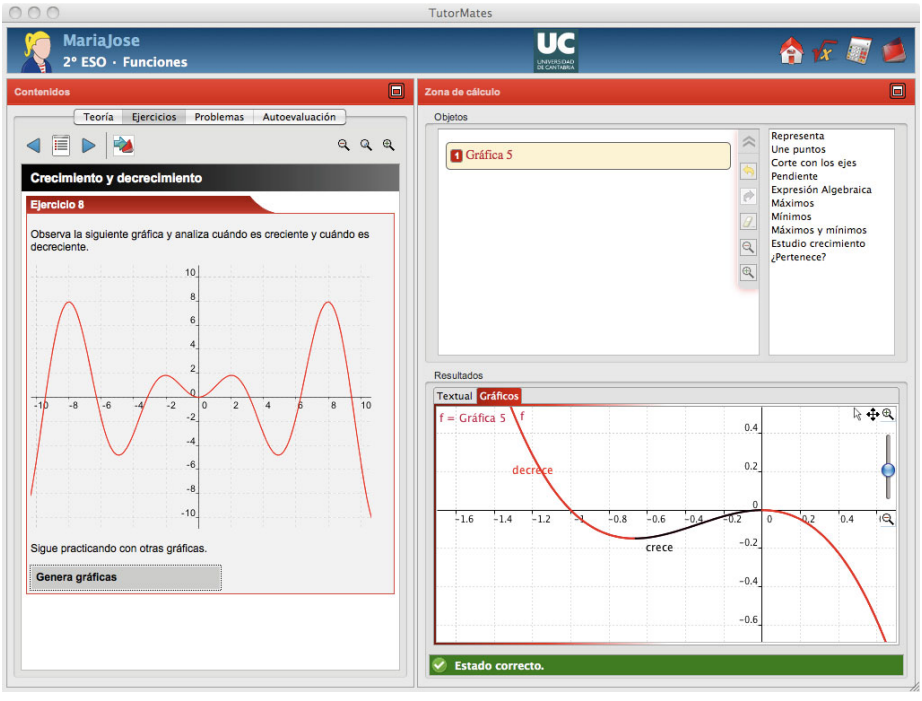


Fig. 3. Function graph produced by a random generator and graphical result of the *Increase-Decrease intervals* command

The Computation Zone of the geometric topics is a graphical environment based on the GeoGebra software. When the considered topic is related to functions, the Computation Zone combines a graphical environment and a list of commands. In these cases, the Intermediary Layer also acts adapting the results to the knowledge of the user. For example, in the Computation Zone of Figure 3 we can see the graph of a function that has been generated by the random generator of graphs of an exercise for 15 years old students. Given that, in this level, students only know the algebraic expressions of linear functions, the system conceals the cubic expression of the function graph, but it allows the student to manipulate this graph by naming it with a literal expression (in this case, 'Gráfica 5'). The result of applying to this expression a command such as *Increase-Decrease intervals*, is of graphical nature (the one that is shown in Figure 3) and of numerical nature (the one that should appear in the 'Textual' tab).

The Computation Zone of Statistics and Probability topics incorporate a table in order to register the input data. These tables are manipulated by means of commands that provide graphical and numerical results that are dynamically linked to the registered input data (Figure 1).

The Computation Zone of every lesson, together with the random generators of data and examples, and all these, integrated in a coherent form in the structure of theoretical content, exercises and problems, succeeded in making of

TutorMates an interactive environment in which the student is involved in new teaching and learning forms.

In the next section, we present some additional characteristics of TutorMates, which are oriented to simplify the users communication with the program and to facilitate the classroom management.

2.3 Classroom Management with TutorMates

TutorMates can be used with two different roles: the *student* and the *teacher* role. Both roles can use all the functionalities of TutorMates in the mathematical lessons. But, when accessing to TutorMates with *teacher* role, the user has, in addition, the possibility of consulting some didactic orientations given to implement every lesson, personalizing the teaching sequences, and managing a Moodle connection that allows the teacher to follow students' progress, evaluate and distribute tasks. Both, in the *student* and the *teacher* roles, one of the most useful managenmet tools is the Digital Notebook of TutorMates. This tool has the same functionality as the paper notebook, but it is enriched with connections to all the lessons elements (Figure 4).

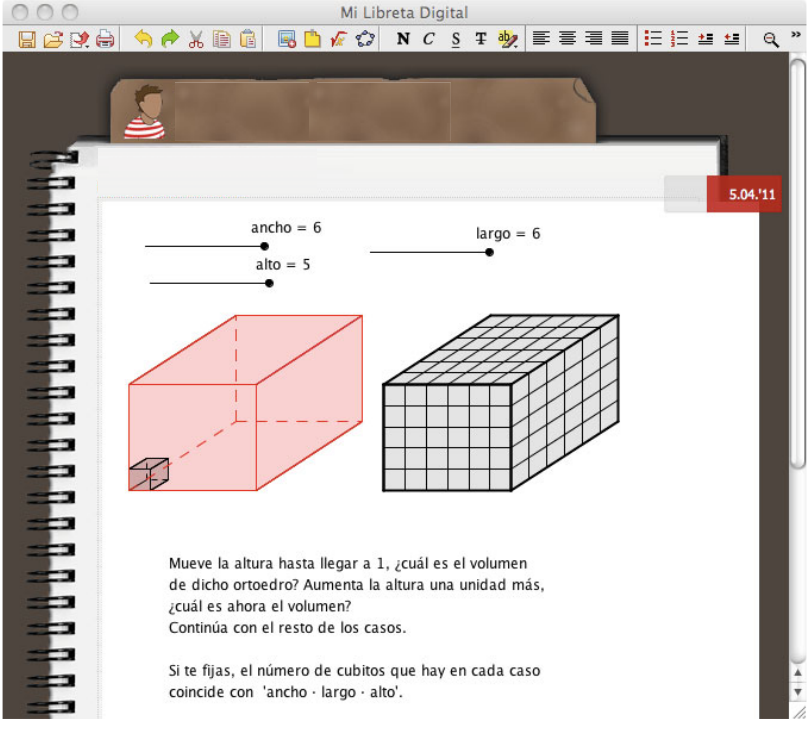


Fig. 4. Digital Notebook with an interactive exercise

Thanks to the Digital Notebook, students can register any personal annotation, bring to the notebook pieces of theoretical information, exercises or problems; can manipulate data or graphic information, send the changes to the Computation Zone, operate there and register again in the notebook the realized actions. The files generated with the Digital Notebook can incorporate specific information prepared by the teacher. The student can send the work made in the Digital Notebook to the teacher to be evaluated, etc. The connectivity of the Digital Notebook with all the elements of the system turns this tool into a key instrument to explore, with the help of technology, problems and open learning situations. In Section 4 we give concrete examples that show the potential of this tool in a Geometry lesson.

3 Technological Issues

TutorMates, as a software system, has been carefully engineered from its beginning, taking into account several technological constraints. For instance, it is a multiplatform application (running under Windows, GNU/Linux and MAC OS X) and it has been designed to have an easy multilingual support. This second constraint is due to the need of being released with the four official Spanish languages: Spanish, Catalan, Basque and Galician.

TutorMates is also multilingual in another sense: it has been programmed in Java, JavaScript and Common Lisp (and use other domain specific languages, as that of Maxima [3]), and relies on several XML tools (as the XUL interfacing language [4]). More concretely, the graphical user interface has been developed in Java, using the power of RCP Eclipse (Rich Client Platform [5]). This has been instrumental to ease the scalability and the components reusing. Java is also the programming language where the TutorMates Expression Editor has been developed. The internal code linking the application with Maxima (which acts as computing engine) is programmed in Common Lisp (let us stress that it is the programming language in which Maxima is written).

The different layers of TutorMates are communicating through a variant of the XML standard MathML [6], enriched to support a treatment of different *profiles* when calling and processing Maxima procedures. Let us explain with some more detail the role of this XML encoding.

3.1 The Intermediary Layer

TutorMates has a *mediated access* (see [1]) to its computational (Maxima) and geometrical (GeoGebra) resources. That is to say that TutorMates does not take profit of the *full* power of these tools. To understand why the capabilities of Maxima must be limited, let us imagine one scenario where a young student introduces an equation in the TutorMates Expression Editor. If Maxima is used without control it could answer some expressions as solutions which could be

beyond the comprehension capacities of our young user (for instance, he/she could be puzzled is some imaginary numbers appear and he/she was not aware yet of the existence of complex numbers). To avoid this kind of *educational inconsistency*, each TutorMates sentence sent to Maxima is accompanied with a *profile* (encoded in our MathML extension). A *profile* consists of several items of information: a *role* (teacher or student), a course, a lesson and other technical information (for instance, whether the result can be reused in other TutorMates frame or not). If TutorMates is used in its *teacher mode*, the system gives full access to Maxima answer but a *warning* is produced if the expression is outside the scope of the active course/lesson (this is a very useful tool when the teacher is using TutorMates to prepare exercises and other material for his/her students). The *lesson* information is also relevant. For instance, in one and the same course, big numbers could be allowed in a lesson on scientific notation, but forbidden in a lesson on formula manipulation.

Similar facilities are provided when accessing to GeoGebra, but this introduces other aspects, which are explained in the following subsection.

3.2 Integrating GeoGebra

TutorMates provides a mediated access to GeoGebra. This implies that the main panel of GeoGebra is presented to TutorMates students with only some options available. While this could be understood as a constraint to the creativity of users, it has the great didactic advantage of guiding the students towards a correct solution (displaying the full power of GeoGebra could loose the student when searching through the different gadgets). This decision can be interpreted in the following way: each geometry exercise is endowed with an implicit proof previously devised by TutorMates designers; then, TutorMates shows only the minimal GeoGebra tools needed to find *that* proof.

Thanks to this mediator, another benefit is obtained. GeoGebra is based on numerical computing, and it is well known that the inherent inaccuracies of this kind of computation raise problems to afford theorem proving. In our educational setting, a student could be puzzled if GeoGebra manipulations give him/her some unexpected result. Our intermediary layer allows us to present (and maintain) a link between textual (symbolic) results (obtained by means of Maxima) and GeoGebra graphical ones. Since TutorMates generators always provide results *with respect to a profile* (recall: a role, a course, a lesson), GeoGebra graphics are, in some sense, "tamed" and the difficulties related to numerical algorithms can be smoothed. To give an example, if the exercise is to find the cuts of a line with the axes, and a student should only know integer numbers, TutorMates will provide straight lines equations only with convenient integer numbers, and the solutions will be shown both graphically and algebraically. Of course, if some manipulation of the graphic is undertaken by the student, then floating-point real numbers could appear, but, at least, the textual representation could help the student to understand better the whole process.

The importance of this symbolic-numerical link to the teaching and learning of Geometry theorem proving needs a more thorough study, and could enhance TutorMates in further releases.

3.3 How Does It Work?

In this subsection, we explain, with the help of an example, how the whole application works. We focus on the interaction between the computer algebra system Maxima and the dynamic geometric application GeoGebra.

Let us consider the situation shown in Figure 3. In this particular example, the user has generated a partially-random function called 'Gráfica 5' in the screenshot. As we have explained before, this name substitutes the algebraic expression of the function, which cannot be managed by 15 years old students. They are now starting to give sense to the notions of local maximum or minimum, increase and decrease intervals, etc. Let us supose that, after generating the function, the user clicks the *Increase-Decrease intervals* command. The steps followed by the software are the following:

1. TutorMates takes the selected MathML objects ('Gráfica 5' in the example) from the input.
2. The application, using the selected command (*Increase-Decrease intervals*) and objects, makes a request to the Intermediate Layer.
3. The Intermediate Layer checks whether the input is appropriate for the level selected and if so, translates the request to Maximas syntax.
4. After the evaluation in Maxima, the result is translated back to MathML syntax and the level is checked again. Notice that the evaluation implies not only getting a solution, but also indirect calculations used, for example, to improve the centering of the graphical representation.
5. If all the tests are passed, the Intermediate Layer returns the result to TutorMates.
6. TutorMates receives the message which also contains information specifying if the result should be represented as a text, graphically or both.
7. If it is textual, the system enhances the textual results with a MathML format.
8. If it is graphical, the system sends the MathML to a module which connects with GeoGebra.
   (a) This module finds the correct XML interface to be used for the command selected before.
   (b) The MathML is translated into GeoGebras syntax.
   (c) Other extra information is also translated, like how to center the figure or how to color the shape of the function.
   (d) The figure is adapted and drawn in GeoGebra.

4 Geometry Teaching with TutorMates

The Digital Notebook, the link of the system with Moodle and the structure of the lessons give big flexibility to TutorMates from the didactic point of view: it is possible to use TutorMates to implement either an expositive teaching style or it is possible to implement a more innovative style in which students experiment and deduce properties by themselves before arriving to a theoretical formalization. Any of these options can take profit of the technology treatment incorporated in TutorMates.

As an example, we show in this section some details concerning a concrete implementation to teach with TutorMates a specific geometry topic: *Similarity* for 15 years old students. We will adopt an interactive approach, in which the student must deduce properties while manipulating dynamic scenes in GeoGebra. This process combines numerical and geometric elements to introduce the students in geometric reasoning processes. As it is not possible in these lines, to show the complete sequence, we will detail two key moments in the process. More precisely, we will see the activities proposed for the student to develop:

- Theoretical knowledge on the similarity notion.
- Procedural knowledge concerning how perimeter and area change by the effect of similarity.

Firstly, students are guided to analyze the regularities that arise when working with different similar figures. They start by identifying what are the geometric elements involved in the notion of similarity. With this purpose, they solve the activity shown in Figure 5, where they are asked to determine what are the similar figures and they must give arguments to explain their answer.

This activity is based on the second theoretical item of the lesson, but the teacher has selected and placed it at the beginning of the students' Digital Notebook for this lesson.

Next, the teacher considers that students have to analyze what happens with the lengths, the angles and the position of similar figures. Instead of demanding students to read the corresponding theoretical information, teacher can propose to solve in the Digital Notebook the activity shown in Figure 6, which has been taken from the Exercise 5 of Tutormates *Similarity* lesson.

Following by, students have to formalize the idea of proportionality constant. They can explore this idea in a GeoGebra scene prepared for this purpose in the first theoretical item of the lesson (Figure 7).

After this process, the teacher can propose to the students a series of GeoGebra activities oriented to apply the previous ideas. These activities can vary from simple mathematic situations to more elaborated contexts in which, for example, students should deal with scales in real situations that can be chosen from the variety of possibilities given in the exercises and problems of the lesson.

Once students have solved the previous activities, the teacher can introduce them into the exploration of new properties on the perimeter and the area of

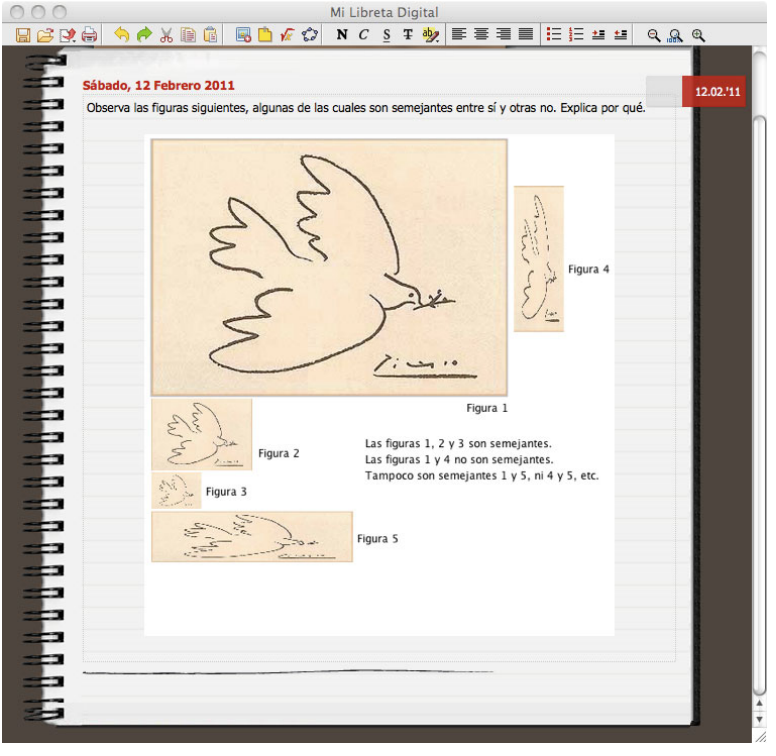


Fig. 5. Exploring activity in the Digital Notebook

similar figures. In Figure 8, we can see a TutorMates exercise to be solved in the Digital Notebook; we remark that students can explore here the properties of the given figures -compute lengths, perimeter or area, draw new lines or segments, etc.- in order to find the answer to the posed questions.

Finally, the teacher can reinforce or formalize students' observations concerning perimeter and area by proposing them to solve the problems of the lesson that specifically treat this property.

With this example, we have tried to show a way of implementing TutorMates's lessons that does not follow the sequential presentation given in the software. The teacher, with the support of the Digital Notebook, can fix the concrete lesson plan that he prefers; the student has an adaptation of Geogebra to every activity he is solving. Considering that TutorMates has been designed to support student centered methodologies, the teacher has the responsability to decide the degree up to which he wants to involve the student in questioning, problem-solving and investigation activities.

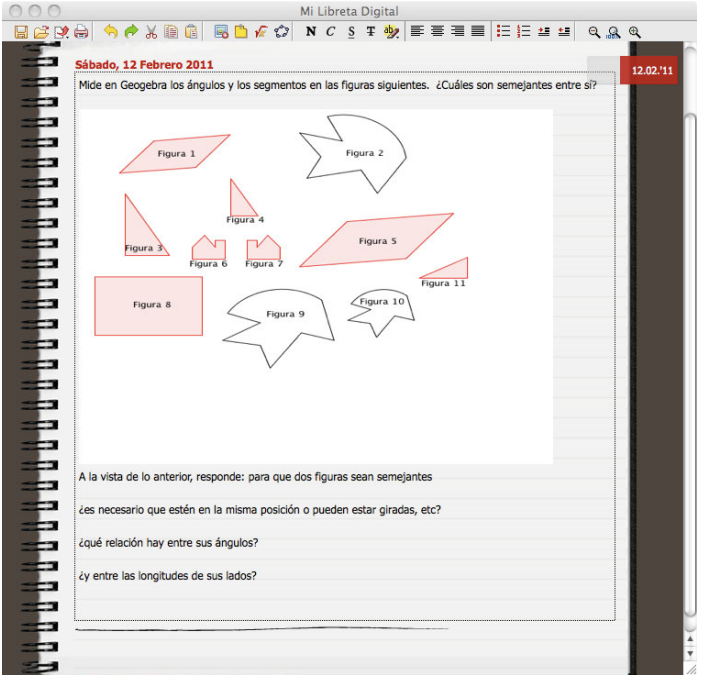


Fig. 6. Activity to be solved in the Digital Notebook

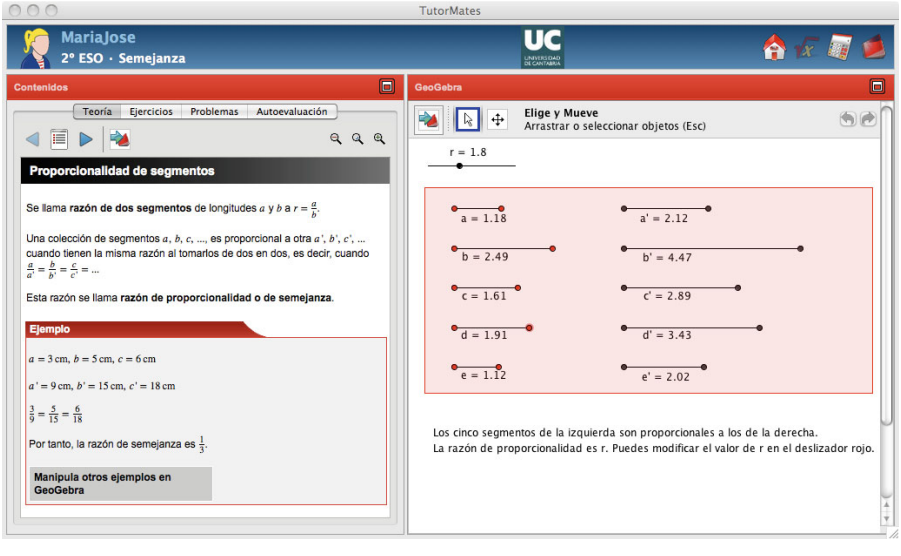


Fig. 7. GeoGebra scene proposed to deal with the proportionality constant

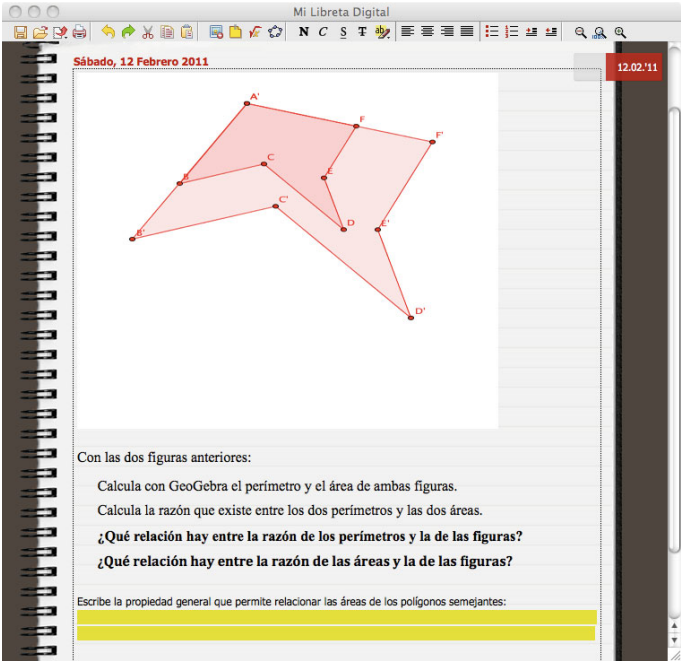


Fig. 8. Exploration activity proposed to deal with perimeter and area of similar figures

5 Conclusions

Undoubtedly, scientific computing software packages have transformed both the research and the professional domains in the last years. Moreover they have also influenced the mathematics education programs at university levels. Their use at the secondary education level is possible only if the powerful computational methods are adapted to the cognitive level of the students, if they are implemented in well-structured interactive processes and if they fit in curricular contexts with attractive and easy to use interfaces. The TutorMates Project team has dealt with these questions and has developed a low cost software, since it is based on open source code, which is proposed as a helpful material for the teaching and learning of mathematics at secondary education levels.

TutorMates' design of mathematics lessons allows this material to be adapted to different educational styles. The consideration of technology in these lessons does not necessarily produce a complete break with respect to the traditional teaching styles of teachers, nor the study habits of the students. TutorMates flexibility admits several interpretations: from its use as a calculation tool, to other approaches in which the student explores or deduces properties.

Finally, we remark that the integration in TutorMates of a symbolic and numerical computation system as well as a dynamic geometry software, will facilitate in the future the incorporation of new advances in this field (for instance

the incorporation of the automatic discovery and proof tools in geometric contexts) which are actually being developed by different research teams and whose development requires, essentially, the cooperation between these two types of technological tools.

References

1. Heras, J., Pascual, V., Rubio, J.: Mediated Access to Symbolic Computation Systems. In: Autexier, S., Campbell, J., Rubio, J., Sorge, V., Suzuki, M., Wiedijk, F. (eds.) AISC 2008, Calculemus 2008, and MKM 2008. LNCS (LNAI), vol. 5144, pp. 446–461. Springer, Heidelberg (2008)
2. Ministerio de Educación y Ciencia: Real Decreto 1631/2006, por el que se establecen las Enseñanzas Mínimas correspondientes a la Educación Secundaria Obligatoria. In: Boletín Oficial del Estado, vol. 5, pp. 677–773 (2007)
3. The Maxima Computer Algebra System, http://maxima.sourceforge.net
4. XULRunner package, https://developer.mozilla.org/en/xulrunner
5. Rich Client Platform, http://wiki.eclipse.org/index.php/Rich_Client_Platform
6. Mathematical Markup Language MathML, http://www.w3.org/Math/
7. GeoGebra mathematics software, http://www.geogebra.org
8. JFree Java chart library, http://www.jfree.org/jfreechart/
9. Addlink Research, http://www.addlink.es

On Equivalence of Conditions for a Quadrilateral to Be Cyclic

Pavel Pech

Faculty of Education, University of South Bohemia,
Jeronýmova 10, 371 15 České Budějovice, Czech Republic
pech@pf.jcu.cz

Abstract. In the paper we will prove a theorem that puts together three conditions — Ptolemy, Cubic and Quartic — for a convex quadrilateral to be cyclic. Further Ptolemy inequality is proved. Some related formulas from geometry of polygons are derived as well. These computations were done by the theory of automated geometry theorem proving using Gröbner bases approach. Dynamic geometry system GeoGebra was applied to verify Ptolemy conditions. These conditions were subsequently proved by Wu–Ritt method using characteristic sets. The novelty of the paper is the method of proving geometric inequalities. Also some relations among Ptolemy, Cubic and Quartic conditions seem to be new.

Keywords: cyclic quadrilaterals, Ptolemy inequality, automated geometry theorem proving.

1 Introduction

In the paper we will be concerned with cyclic quadrilaterals, i.e., those whose vertices lie on a circle.

Consider a plane quadrilateral $ABCD$ with lengths of sides a, b, c, d and diagonals e, f, where $a = |AB|$, $b = |BC|$, $c = |CD|$, $d = |DA|$, $e = |AC|$, $f = |BD|$.

We formulate the following three cyclicity conditions which are necessary for a *convex* quadrilateral to be cyclic:

Ptolemy (quadratic) condition:

$$P := ac + bd - ef = 0, \tag{1}$$

Cubic condition:

$$S := e(ab + cd) - f(ad + bc) = 0, \tag{2}$$

Quartic condition:

$$K := e^2(ab + cd) - (a^2 + b^2)cd - (c^2 + d^2)ab. \tag{3}$$

The best known from the conditions above is the Ptolemy condition $P = 0$ which ensures that a quadrilateral is cyclic and convex. Conversely, if a quadrilateral is convex and cyclic then the Ptolemy condition $P = 0$ follows. Moreover for an arbitrary quadrilateral the inequality $P \geq 0$ holds.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 399–411, 2011.

If a quadrilateral is cyclic and convex then the quartic condition $K = 0$ holds. The converse statement is not true. Quartic condition follows from the subtended arc criterion and the cosine rule.

The cubic condition $S = 0$ is less known [10], [8]. For a cyclic and convex quadrilateral $S = 0$ holds true whereas the converse statement does not hold. The cubic condition can be used for instance to simplify expressions containing Cayley-Menger determinants [10], [6].

All three conditions (1), (2), (3) are closely connected. We will show that $P = 0$ is equivalent to $(K = 0 \wedge S = 0)$. Some related formulas are derived as well.

In proofs we applied the theory of automated geometry theorem proving based on Gröbner bases computation [4], [6], [9] and Wu–Ritt method with characteristic sets [2], [11], [13], [12]. Also dynamic geometry system GeoGebra is applied to demonstrate and verify statements. Although the topic is from elementary geometry, it is not easy to prove all statements in automated way. The novelty of the paper is the method of proving geometric inequalities. Some relations among cyclicity conditions (1), (2), (3) seem to be new as well.

Throughout the paper we will use dynamic geometry system GeoGebra and computer algebra systems Maple, CoCoA[1] and Epsilon.[2]

2 Preliminaries

It is obvious that a quadrilateral $ABCD$ in a plane is convex if the line AC separates the vertices B, D and the line BD separates the vertices A, C. Let us express the convexity of a quadrilateral in an algebraic form.

In a rectangular system of coordinates $A = [0, 0]$, $B = [a, 0]$, $C = [u, v]$, $D = [w, z]$, (Fig. 1) expressing diagonals AC and BD, we get:

Theorem 1. *A quadrilateral $ABCD$ is convex iff*

$$N \geq 0 \wedge M \geq 0, \tag{4}$$

where

$$N := \begin{vmatrix} 0, & 0, & 1 \\ u, & v, & 1 \\ w, & z, & 1 \end{vmatrix} \cdot \begin{vmatrix} u, & v, & 1 \\ 0, & 0, & 1 \\ a, & 0, & 1 \end{vmatrix} = (uz - vw)av \tag{5}$$

and

$$M := \begin{vmatrix} a, & 0, & 1 \\ w, & z, & 1 \\ 0, & 0, & 1 \end{vmatrix} \cdot \begin{vmatrix} w, & z, & 1 \\ a, & 0, & 1 \\ u, & v, & 1 \end{vmatrix} = az(av + uz - vw - az). \tag{6}$$

Notice that by (5) and (6), N, M are equal to four times the product of the *oriented* areas of triangles ACD, CAB and BDA, DBC respectively.

[1] Program CoCoA is freely distributed at `http://cocoa.dima.unige.it`

[2] Program Epsilon is freely distributed at `http://www-calfor.lip6.fr/~wang/epsilon/` Program is working under Maple.

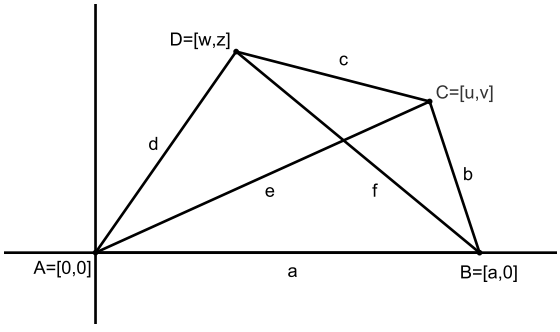


Fig. 1. Convex quadrilateral $ABCD$

The next theorem [5] is the criterion for four points with given coordinates to lie on a circle:

Theorem 2. *Points* $A = [x_1, y_1]$, $B = [x_2, y_2]$, $C = [x_3, y_3]$, $D = [x_4, y_4]$ *are cocyclic iff*

$$C := \begin{vmatrix} x_1^2 + y_1^2, x_1, y_1, 1 \\ x_2^2 + y_2^2, x_2, y_2, 1 \\ x_3^2 + y_3^2, x_3, y_3, 1 \\ x_4^2 + y_4^2, x_4, y_4, 1 \end{vmatrix} = 0.$$

For the choice $A = [0, 0]$, $B = [a, 0]$, $C = [u, v]$, $D = [w, z]$ we get

$$C := a(-avw + vw^2 + auz - u^2 z - v^2 z + vz^2) = 0. \tag{7}$$

Euler's four points relation [1] which gives a mutual dependence of all distances between four vertices of a quadrilateral in a plane is as follows:

Theorem 3. *If* a, b, c, d, e, f *are six mutual distances between four vertices* A, B, C, D *of a plane quadrilateral then*

$$U := \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & a^2 & e^2 & d^2 \\ 1 & a^2 & 0 & b^2 & f^2 \\ 1 & e^2 & b^2 & 0 & c^2 \\ 1 & d^2 & f^2 & c^2 & 0 \end{vmatrix} = 0. \tag{8}$$

The determinant U above is the well known Cayley–Menger determinant for the volume of a tetrahedron.

Furthermore we will need the following theorem [3] to determine definiteness of polynomials by factorization:

Theorem 4. *Let* f, g, h *be polynomials in real variables* $x = (x_1, x_2, \ldots, x_n)$ *such that* $f = gh$, *where* g *and* h *have no common factors. Then*

$$(\forall x \in R^n : f \geq 0) \Leftrightarrow [(\forall x \in R^n : g \geq 0 \ \wedge \ h \geq 0) \vee (\forall x \in R^n : g \leq 0 \ \wedge \ h \leq 0)].$$

In the next theorem we present three relations which will be useful further. Formulas (10) and (11) seem to be new.

Theorem 5. *Let $ABCD$ be a quadrilateral with side lengths a, b, c, d and diagonals e, f. Then the following relations hold:*

$$S^2 = PV \tag{9}$$

$$eS = K + (bc + ad)P, \tag{10}$$

$$K^2 = PQR, \tag{11}$$

where P, S, K are as above and

$V := ac(-a^2 - c^2 + b^2 + d^2 + e^2 + f^2) + bd(a^2 + c^2 - b^2 - d^2 + e^2 + f^2) - ef(a^2 + c^2 + b^2 + d^2 - e^2 - f^2)$,

$Q := ac + bd + ef$,

$R := (bc + ad)^2 - e^2(a^2 + b^2 + c^2 + d^2 - e^2 - f^2)$.

Proof. It suffices to verify whether the polynomials corresponding to the relations (9), (10), (11) belong to the ideal $I = (h_1, h_2, \ldots, h_5)$ which describes the quadrilateral $ABCD$, see the next section for $h_1, h_2, \ldots, h_5$ (we omit it).

Remark 1. The formula (9) which is due to V. P. Varin [10] can be written (without reduction to canonical form modulo ideal I) in the form

$$S^2 \;=\; PV - (1/2)U, \tag{12}$$

where U is the Cayley–Menger determinant from (8). This is easy to verify by direct computation. As in a plane quadrilateral $U = 0$ by (8) then from (12) the relation $S^2 = PV$ follows.

Remark 2. Let us note that from (9) and (11) and the Ptolemy inequality $P \geq 0$ inequalities $V \geq 0$ and $R \geq 0$ follow.

In the Theorems 7 and 8 we will use *saturation*. Let us remind it [12]:

Definition 1. *Let I be an ideal and F a polynomial in $K[x]$. The saturation of I with respect to F is the set*

$$I : F^{\infty} = \{P \in K[x] : F^q P \in I \text{ for some integer } q > 0\}.$$

3 Some Equivalences among Cyclicity Conditions

We start with the best known condition for a quadrilateral to be cyclic — the Ptolemy condition (1). Unlike the classical proof we will prove it using automated tools based on Gröbner bases computation.

Consider an arbitrary quadrilateral $ABCD$ with side lengths a, b, c, d and diagonals e, f and denote $|AB| = a$, $|BC| = b$, $|CD| = c$, $|DA| = d$, $|AC| = e$, $|BD| = f$.

Choose a rectangular coordinate system such that $A = [0, 0]$, $B = [a, 0]$, $C = [u, v]$, $D = [w, z]$, (Fig. 2). We will prove:

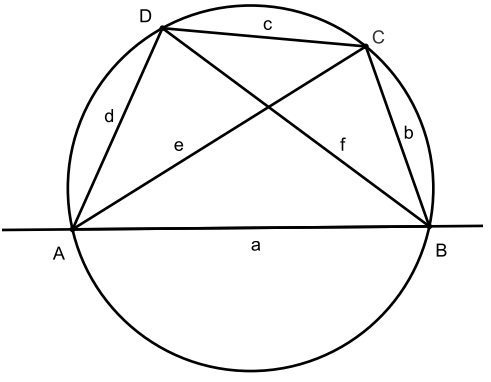


Fig. 2. Convex cyclic quadrilateral $ABCD$

Theorem 6. *A quadrilateral $ABCD$ with side lengths a, b, c, d and diagonals e, f is convex and cyclic iff*

$$P := ac + bd - ef = 0. \tag{13}$$

Proof. Let us describe the quadrilateral $ABCD$ in an algebraic form:

$b = |BC| \Rightarrow h_1 := (u - a)^2 + v^2 - b^2 = 0,$
$c = |CD| \Rightarrow h_2 := (w - u)^2 + (z - v)^2 - c^2 = 0,$
$d = |DA| \Rightarrow h_3 := w^2 + z^2 - d^2 = 0,$
$e = |AC| \Rightarrow h_4 := u^2 + v^2 - e^2 = 0,$
$f = |BD| \Rightarrow h_5 := (w - a)^2 + z^2 - f^2 = 0.$

Further add the cocyclic condition (7)

$C := a(-avw + vw^2 + auz - u^2z - v^2z + vz^2) = 0,$

and polynomials (5),(6) for investigating convexity of a quadrilateral

$N := (uz - vw)av,$

$M := az(av + uz - vw - az).$

First suppose that $P = 0$. Consider the ideal $I = (h_1, h_2, \ldots, h_5)$. We will show that $C \in \sqrt{I \cup \{P\}}$ which implies that $ABCD$ is cyclic. To do this we will use Gröbner bases approach and the program CoCoA. We enter

```
Use R::=Q[abcdefuvwzt];
J:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,
u^2+v^2-e^2,(w-a)^2+z^2-f^2,P,Ct-1);
NF(1,J);
```

and obtain the result `NF(1,J)=0`. Thus the cyclicity of $ABCD$ is confirmed.[3]

Further we will show that $ABCD$ is convex, i.e., that $N \geq 0$ and $M \geq 0$ by (4). To carry out this we will use the method of reduction of polynomials to canonical form modulo an ideal [3]. Consider the ideal $K = I \cup \{P\} \cup \{N - t\}$, where t is a slack variable. Then the elimination of variables e, f, u, v, w, z in K

```
Use R::=Q[abcdefuvwzt];
K:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,
u^2+v^2-e^2,(w-a)^2+z^2-f^2,P,N-t);
Elim(e..z,K);
```

gives in 13s the relation

$$t = \frac{abcd(-a+b+c+d)(a-b+c+d)(a+b-c+d)(a+b+c-d)}{4(ab+cd)^2}. \quad (14)$$

From (14) non-negativity of $N = t$ follows, since in a quadrilateral inequalities $-a+b+c+d \geq 0$, $a-b+c+d \geq 0, \ldots$, which are analogous to triangle inequalities, hold.

Similarly, for $M = s$ we get

$$s = \frac{abcd(-a+b+c+d)(a-b+c+d)(a+b-c+d)(a+b+c-d)}{4(bc+ad)^2} \quad (15)$$

which implies that $M \geq 0$.

Note that the values of t from (14) and s from (15) equal four times the product of areas of triangles ABC, CDA and BCD, DAB in the case that $ABCD$ is convex and cyclic [6].

Now we will prove the converse implication. Suppose that a quadrilateral $ABCD$ is cyclic and convex. We are to prove that $P = 0$.

If $ABCD$ is cyclic and convex then $C = 0$, $N - t = 0$, $M - s = 0$, where t and s are the values from (14) and (15). The conditions $N - t = 0$ and $M - s = 0$ may be written as $T1 = 0$ and $T2 = 0$, where

$$T1 := 4N(ab+cd)^2 - abcd(-a+b+c+d)(a-b+c+d)(a+b-c+d)(a+b+c-d),$$

$$T2 := 4M(bc+ad)^2 - abcd(-a+b+c+d)(a-b+c+d)(a+b-c+d)(a+b+c-d).$$

We will investigate whether P belongs to the radical of the ideal $I \cup \{C, T1, T2\}$. After searching for additional (non-degeneracy) conditions (we omit this)

[3] Using the weaker criterion whether $C \in I \cup \{P\}$ we find out that $NF(C, I \cup \{P\}) \neq 0$. Hence C does not belong to $I \cup \{P\}$ but C^2 does, as we can easily verify. This is another example of the ideal $I \cup \{P\}$ which is not a radical ideal [2], [7].

```
Use R::=Q[abcdefuvwzsrt];
L:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,
u^2+v^2-e^2,(w-a)^2+z^2-f^2,C,T1,T2,(a+c)(a-c)(b-d)(b+d)s-1,
(e(ab+cd)+f(bc+ad))r-1,Pt-1);
NF(1,L);
```

we get in 12s the result `NF(1,L)=0`. It remains to show that the theorem holds true also for the values $a = c$ and $b = d$ that were excluded. As $ABCD$ is cyclic then it is a rectangle with $e - f = 0$. This implies that $P := a^2 + b^2 - e^2$ equals zero due to the Pythagorean and Thales theorems. The theorem is proved.

Next we will show how the three conditions (1), (2) and (3) are connected. The following theorem seems to be new.

Theorem 7. *In a quadrilateral $ABCD$ with side lengths a, b, c, d and diagonals e, f*

$$(K = 0 \ \wedge \ S = 0) \ \Leftrightarrow P = 0. \tag{16}$$

Proof. Using the same notation as in the previous theorem we first prove the necessity of $P = 0$. Consider the ideal $I = (h_1, h_2, \ldots, h_5)$. We are to show that P belongs to the radical of the ideal $L = I \cup \{K, S\}$. Evaluation of the normal form of 1 with respect to the ideal J gives

```
Use R::=Q[abcdefuvwzt];
J:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,
u^2+v^2-e^2,(w-a)^2+z^2-f^2,K,S,Pt-1);
NF(1,J);
```

the result `NF(1,J)=1`. Searching for additional conditions yields

```
Use R::=Q[abcdefuvwz];
L:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,
u^2+v^2-e^2,(w-a)^2+z^2-f^2,K,S);
Elim(u..z,Saturation(L,Ideal(P)));
```

the polynomial $bc + ad$. Then the normal form of 1 with respect to the ideal $J \cup \{(bc + ad)s - 1\}$ is equal to zero. This completes the proof of the necessity part of the theorem.

To prove that the condition $P = 0$ is sufficient for holding $K = 0 \wedge S = 0$, we take the ideal $I \cup \{P\}$ and investigate whether $K, S \in \sqrt{I \cup \{P\}}$. In both cases we get, without the use of additional conditions, a positive answer, that is $P = 0$ implies $K = 0 \wedge S = 0$. The theorem is proved.

Putting the Theorems 6 and 7 together we get:

Consequence. *The following conditions are equivalent:*

1. *$ABCD$ is cyclic and convex,*
2. *$P = 0$,*
3. *$K = 0 \wedge S = 0$.*

Remark 3. From the relation $eS = K + (bc + ad)P$ in (10) immediately $(K = 0 \wedge S = 0) \Rightarrow P = 0$ follows.

Similarly, from $S^2 = PV$ in (9) we get $(P = 0 \Rightarrow S = 0)$.

And finally from $K^2 = PQR$ in (11) the implication $(P = 0 \Rightarrow K = 0)$ follows.

4 Ptolemy and Related Inequalities

In this section we will prove the well-known Ptolemy inequality in an automated way. The method, which seems to be new, is based on saturation of an ideal with respect to a given polynomial.

The theorem on the Ptolemy inequality is as follows:

Theorem 8. *In a quadrilateral $ABCD$ with side lengths a, b, c, d and diagonals e, f the inequality*

$$P := ac + bd - ef \ \geq 0 \tag{17}$$

holds. The equality is attained iff $ABCD$ is a convex cyclic quadrilateral.

Proof. In the proof of the inequality (17) we will use the method which is based on saturation of an ideal with respect to a given polynomial. We can call it the auxiliary polynomial method as well.

Using the same notation as above we eliminate dependent variables b, c, d, e, f in the ideal $I = (h_1, h_2, \ldots, h_5) \cup \{P\}$

```
Use R::=Q[abcdefuvwz];
I:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,u^2+v^2-e^2,
(w-a)^2+z^2-f^2,P);
Elim(b..f,I);
```

to obtain a polynomial

$$M := a(-avw + vw^2 + auz - u^2z - v^2z + vz^2)^2. \tag{18}$$

Note that $M \geq 0$ (we suppose that $a > 0$) and realize that $M = 0$ is the condition (7) for the points A, B, C, D to be cocyclic.

Now we will use saturation of the ideal $J = I \cup \{M\}$ with respect to the polynomial P. We get

```
Use R::=Q[abcdefuvwz];
J:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,u^2+v^2-e^2,
(w-a)^2+z^2-f^2,M);
Elim(u..z,Saturation(J,Ideal(P)));
```

a few polynomials from which we take a polynomial

$N := (-ac + bd + ef)(ac - bd + ef)(ac + bd + ef)$.

The normal form `NF(NP,J)` of the product of polynomials NP with respect to the ideal J equals zero which implies that NP belongs to the ideal J. Using general representation of the polynomial NP with respect to the ideal J we get

```
Use R::=Q[abcdefuvwz];
J:=Ideal((u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,u^2+v^2-
e^2,(w-a)^2+z^2-f^2,M);
GenRepr(NP,J);
```

$$NP = c_1h_1 + c_2h_2 + c_3h_3 + c_4h_4 + c_5h_5 + 4aM,$$

where $c_1, c_2, \ldots, c_5$ are some polynomials. As $h_1 = 0$, $h_2 = 0$, $h_3 = 0$, $h_4 = 0$, $h_5 = 0$ then

$$NP = 4aM. \tag{19}$$

From (18) and (19) we conclude that

$$NP \geq 0. \tag{20}$$

As $ac + bd + ef > 0$ instead of (20) we can investigate

$$(ac + bd - ef)(ac - bd + ef)(-ac + bd + ef) \geq 0. \tag{21}$$

By the Theorem 4 each polynomial $ac + bd - ef$, $ac - bd + ef$, $-ac + bd + ef$ from (21) does not change the sign. If all polynomials in (21) are non-negative then we are ready. Now suppose that e.g. $ac + bd - ef \geq 0$ and $ac - bd + ef \leq 0$, $-ac + bd + ef \leq 0$. Then $(ac - bd + ef) + (-ac + bd + ef) = 2ef \leq 0$ which is a contradiction. Thus we get three inequalities

$$ac + bd - ef \geq 0,\ ac - bd + ef \geq 0,\ -ac + bd + ef \geq 0. \tag{22}$$

The case of equality in $ac + bd - ef \geq 0$ was investigated in previous parts and we omit it. The theorem is proved.

All equality cases from (22) will be studied in the next section.

5 Verification and Proofs of Ptolemy Conditions

The first part of this section is important especially from a pedagogical point of view. We will deal with a verification — a mighty tool of dynamic geometry systems. Verification in DGS can be considered as a weaker form of the proof which can be used in teaching mathematics at lower classes of schools.

We will verify the cases of equality of the Ptolemy and related inequalities with dynamic geometry software GeoGebra. Then we will decompose a corresponding polynomial set into irreducible triangular sets and evaluate pseudo-remainders of individual Ptolemy conditions with respect to these components.

Let us look at the following statement:

Denote $P = ac + bd - ef$, $M = ac - bd + ef$, $N = -ac + bd + ef$. *Then*

$$ABCD \text{ is cyclic } \Rightarrow\ (P = 0\ \vee\ M = 0\ \vee\ N = 0).$$

Conditions $P = 0$, $M = 0$, $N = 0$ are called Ptolemy conditions.

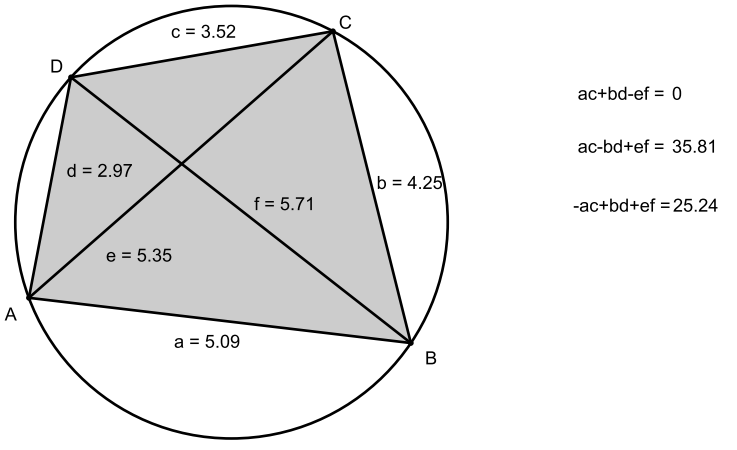


Fig. 3. Verification of Ptolemy conditions - convex case

First we will verify the statement in GeoGebra. Suppose that the vertices A, B, C are fixed and the vertex D moves on the circumcircle of ABC, Fig 3. Points A, B, C divide the circle into three circular arcs. In GeoGebra we calculate values of polynomials P, M, N for actual lengths of sides a, b, c, d and diagonals e, f that they attain when the vertex D moves along the circumcircle of ABC. We get three different results in accordance with the position of D. If the vertex D lies on the circular arc between the points A, C then $P = 0$, whereas $M \neq 0$, $N \neq 0$. In this case the quadrilateral $ABCD$ is convex, see Fig. 3. If D lies between B, C then $M = 0$ whereas $P \neq 0$, $N \neq 0$. Finally if D lies between the points A, B then $N = 0$ whereas $P \neq 0$, $M \neq 0$, Fig 4.

In the second part of this section we will prove what we have found by verification. Namely we show that the case $P = 0$ occurs if the point D of a quadrilateral $ABCD$ is on the circular arc between the vertices A, C etc. To do this we use the program Epsilon using Wu–Ritt approach with characteristic sets [13], [12].

Suppose that a cyclic quadrilateral $ABCD$ is described by the polynomials $h_1, h_2, \ldots, h_5, C$, where we use the same notation as before. Now we will decompose this set of polynomials into irreducible triangular sets with the ordering $a \prec b \prec c \ldots \prec z$. Using the package TriSys we enter

```
with(trisys); C:=a*(-a*v*w+v*w^2+a*u*z-u^2*z-v^2*z+v*z^2);
X:=[a,b,c,d,e,f,u,v,w,z];
its({(u-a)^2+v^2-b^2,(w-u)^2+(z-v)^2-c^2,w^2+z^2-d^2,u^2+v^2-e^2,
(w-a)^2+z^2-f^2,C},X);
```

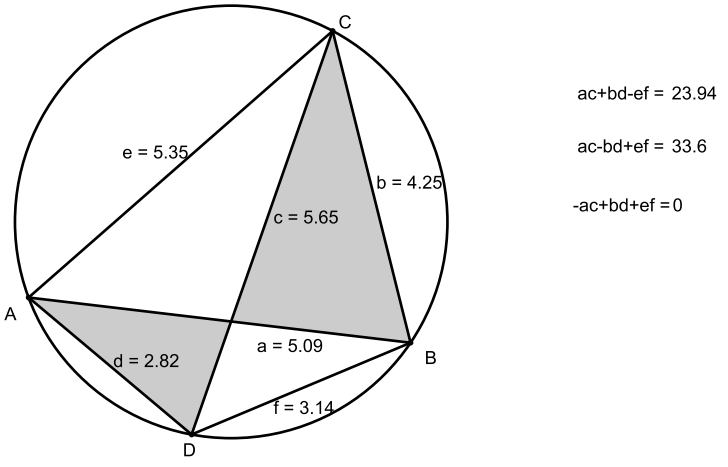


Fig. 4. Verification of Ptolemy conditions - non-convex case

and in 8.2s obtain *four* non-degenerate irreducible triangular sets V_1, V_2, V_3, V_4. Let us denote

$f_{11} := e^2(ab - cd) + (a^2 + b^2)cd - (c^2 + d^2)ab,$

$f_{12} := e^2(ab + cd) - (a^2 + b^2)cd - (c^2 + d^2)ab,$

$f_{21} := e(ab - cd) - f(bc - ad),$

$f_{22} := e(ab - cd) + f(bc - ad),$

$f_{23} := e(ab + cd) - f(bc + ad),$

$f_{24} := e(ab + cd) + f(bc + ad),$

$f_3 = e^2 - 2ua + a^2 - b^2,$

$f_{41} := b^2e^2a^4 - 4a^3bdcv^2 - 2a^2e^2b^4 - 2a^2b^2e^2c^2 - 2a^2e^2d^2b^2 + 4a^2v^2b^2d^2 + 4a^2v^2b^2c^2 + 4d^2c^2a^2v^2 + 8ab^3e^2cd - 4ab^3dcv^2 - 4abd^3cv^2 - 4abc^3dv^2 + e^2b^6 - 2b^4e^2d^2 - 2b^4e^2c^2 - 2b^2e^2c^2d^2 + b^2e^2d^4 + 4b^2d^2c^2v^2 + b^2e^2c^4,$

$f_{42} := b^2e^2a^4 + 4a^3bdcv^2 + 4a^2v^2b^2c^2 + 4d^2c^2a^2v^2 - 2a^2e^2d^2b^2 + 4a^2v^2b^2d^2 - 2a^2e^2b^4 - 2a^2b^2e^2c^2 - 8ab^3e^2cd + 4ab^3dcv^2 + 4abd^3cv^2 + 4abc^3dv^2 - 2b^4e^2d^2 - 2b^4e^2c^2 - 2b^2e^2c^2d^2 + b^2e^2d^4 + 4b^2d^2c^2v^2 + b^2e^2c^4 + e^2b^6,$

$f_5 := -2wu + u^2 - 2zv + v^2 - c^2 + 2wa - a^2 + f^2,$

$f_6 := -d^2 + 2wa - a^2 + f^2.$

Then the irreducible triangular sets are as follows:

$V_1 = \{f_{11}, f_{21}, f_3, f_{41}, f_5, f_6\},$

$V_2 = \{f_{11}, f_{22}, f_3, f_{41}, f_5, f_6\},$

$V_3 = \{f_{12}, f_{23}, f_3, f_{42}, f_5, f_6\},$

$V_4 = \{f_{12}, f_{24}, f_3, f_{42}, f_5, f_6\}.$

The pseudo-remainder of P equals zero only on the component V_3, hence $prem(P, V_3) = 0$. Notice that V_3 contains the polynomials K, S which means that $P = 0$ occurs if the vertex D lies on the circular arc between the vertices A, C, i.e., if the quadrilateral $ABCD$ is convex.

Similarly we get $prem(M, V_2) = 0$, $prem(N, V_1) = 0$ for non-convex cases and finally $prem(ac + bd + ef, V_4) = 0$.

Concluding Remarks

1. In the previous automated proof of the geometric inequality (17) in the Theorem 8 human factor is used at least in two steps. First when we choose by saturation an appropriate polynomial N from a few polynomials. Second when we need to prove (17) from the inequality (21), where we used the proof by contradiction. It is a question how to reduce such human intervention.

2. From the topic just presented one problem remains open. How to prove in automated way that if a quadrilateral $ABCD$ is convex and the condition $S = 0$ holds, then $ABCD$ is cyclic? Classical solution is known [8], computer solution is very long (24 pages) [10]. It seems that Rabinowitsch/Seindenberg device of converting polynomial inequalities to equations could help [3], [14].

Acknowledgments. The author wish to thank the referees for their valuable and helpful suggestions. The research is partially supported by the University of South Bohemia grant GAJU 089/2010/S.

References

1. Berger, M.: Geometry I. Springer, Heidelberg (1987)
2. Chou, S.C.: Mechanical Geometry Theorem Proving. D. Reidel Publishing Company, Dordrecht (1987)
3. Chou, S.C., Gao, X.S., Arnon, D.S.: On the mechanical proof of geometry theorems involving inequalities. Advances in Computing Research 6, 139–181 (1992)
4. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties and Algorithms. Springer, Berlin (1997)
5. Kowalewski, G.: Einführung in die Determinantentheorie. Veit & Comp. Leipzig (1909)
6. Pech, P.: Selected topics in geometry with classical vs. computer proving. World Scientific Publishing, New Jersey (2007)
7. Pech, P.: On the Need of Radical Ideals in Automatic Proving: A Theorem about Regular Polygons. In: Botana, F., Recio, T. (eds.) ADG 2006. LNCS (LNAI), vol. 4869, pp. 157–170. Springer, Heidelberg (2007)
8. Rashid, M.A., Ajibade, A.O.: Two conditions for a quadrilateral to be cyclic expressed in terms of the lengths of its sides. Int. J. Math. Educ. Sci. Techn. 34, 739–742 (2003)
9. Recio, T., Vélez, M.P.: Automatic Discovery of Theorems in Elementary Geometry. J. Automat. Reason. 12, 1–22 (1998)

10. Sadov, S.: Sadov's Cubic Analog of Ptolemy's Theorem (2004), http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/sadov.html
11. Wang, D.: Gröbner Bases Applied to Geometric Theorem Proving and Discovering. In: Buchberger, B., Winkler, F. (eds.) Gröbner Bases and Applications. Lecture Notes of Computer Algebra, pp. 281–301. Cambridge Univ. Press, Cambridge (1998)
12. Wang, D.: Elimination Methods. Springer, Wien New York (2001)
13. Wang, D.: Elimination Practice. Software Tools and Applications. Imperial College Press, London (2004)
14. Wu, W.-t.: Mathematics Mechanization. Science Press, Beijing; Kluwer Academic Publishers, Dordrecht (2000)

An Algorithm for Prediction of Overhead Messages in Client-Server Based Wireless Networks

Azeem Irshad, Muddesar Iqbal, Amjad Ali, and Muhammad Shafiq

University of Gujrat, Pakistan
{irshadazeem,m.iqbal,amjad.ali,shafiq}@uog.edu.pk

Abstract. Discrete math has countless applications in computer science. This paper also identifies another application in computer networks by developing the possible sequences of interaction among nodes from different platforms. The mobile nodes scattered across the network need to communicate and exchange messages either directly or through servers. The more dynamic the mobile nodes, the more diffusion or high turnover of nodes leaving and entering different servers or platforms and hence, higher the overhead. This paper proposed an algorithm that indicates and predicts the overhead in different scenarios of grouping of nodes under different servers. On the basis of this prediction and information about the geographical distribution, the nodes can foresee and arrange for resources or the requisites for establishing communication sessions among nodes of different servers.

Keywords: Discrete math, mobile nodes, Geographical distribution, Grouping, Servers.

1 Introduction

This paper identifies an application of discrete math [12], [13], [14] in computer networks. This paper studies the impact of nodes dynamically joining or leaving the servers. This dynamism leads to switching among different combination of sequences or strings [1]. Every sequence has its own overhead. The switching might lead to an increased or decreased overhead sequence. We have proposed an algorithm that generates all sorts of sequences that might help the stakeholders to foresee the resources required. The numbers in a sequence lead to the output of interaction messages that will help the end users i.e., nodes to prepare for the exchange of maximum number of messages to cover the whole network.

In networks there might be many servers operating and comprising some nodes under each of them. The nodes under a single server can be termed as a single group. For the proper functioning of the network the nodes need to be prepared for communication in terms of resources. Hence, the node might also want to know about the maximum number of interaction messages [5] for communication among the nodes of diverse groups. As the nodes might need to communicate and elicit information from the nodes of other groups of different servers. We assume that there

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 412–423, 2011.

is no need for communication among the nodes of a single server. As they already know about their own servers but they might be unknown about the nodes of other servers. For the nodes of a group, the server broadcasts its own information to all nodes frequently. Thus no interaction messages might be needed to place internally, however, they remain ignorant of other servers. Hence, the external messages are crucial and will be exchanged in this scenario among the nodes to reveal their identities and particulars about the origin of servers. These external messages are said to be interaction messages in the context of this paper.

The nodes are generally mobile in the state of the art networks, the random entry and exit of nodes from groups leaves the strength of nodes in a group quite dynamic. If the node comes to know about the current or recent distribution of nodes among groups, it anticipates whether its possible interaction messages are going to increase or decrease with the time in the near future, and if so, then by how much amount. If we assume that for a specific short time period the geographic division can be predictable, the nodes can foresee overhead for such time period and can arrange the resources accordingly, like battery power or alternative energy resources etc., to maintain the communicating sessions.

A group looses its identity if all of its nodes move to other groups; however, its server or docking station might remain there to serve. Any node in the future might be present in its domain and can join it anytime. Here we have shown the hierarchy of increasing or decreasing number of messages. Different sequence patterns might also lead to similar number of interaction messages. The sum of each pattern or sequence remains constant, as the number of nodes is assumed constant. The number of nodes entering the network is taken as a variable and the calculation overhead will increase with the increasing number of nodes and vice-versa. However, for a specific time period the strength of nodes will be assumed as constant. In that time period the nodes would be managing their resources according to a single calculated output of a sequence. The requirements for a node might be changed in another time period due to the change in the sequence. The provision of updated calculations to each node becomes a challenge with growing mobility on the part of nodes.

In this paper different organized sequences are used from which the users come to know about how many messages would be required for the whole network. Different groups having varying strength of nodes represent the values of a sequence. If they know about the geographical division of nodes among different groups, they can calculate the overall number of interactions required to cover the whole network. It may also be calculated by the server and broadcasted in the network. So the burden of calculation can be put on the server and the nodes as well. It would be more optimal if the algorithms are applied on the server and the generated output is provided to the nodes. Some algorithms have been presented for generating sequences and analysis in the proposed model.

The paper has been organized into three sections. The section 2 describes about the previous related work. The section 3 describes the proposed model. The section 4 concludes the findings.

2 Related Work

The overhead analysis in the context of geographical distribution of nodes has not been found in the literature, as far as I researched. The literature review has been stated with reference to networks that can modeled in the frame of network servers. Such type of networks can be MANETs, Mesh based networks, persuasive networks, WIFI, WIMAX particularly based on hierarchical architecture.

The MANETs [3] are mobile ad hoc networks and function in a peer-to-peer and decentralized manner. There might be several Group Representatives (GR) [2] under which the nodes may operate. For security reasons and optimal mobility management, the ad hoc networks cease to function purely on peer-2-peer and there must be some authority, be it local or external. There might be hierarchy of levels for GRs and Fixed Gateways in a MANET and can be called as servers. The nodes needs to authenticate themselves, belonging to different servers or establish communicating sessions, directly or indirectly through GRs. Being on different platforms, a relatively higher number of messages would be required to authenticate or establish communicating sessions. The MANETs lack the analysis of overhead for establishing the sessions among the nodes belonging to various groups.

In WIFI 802.11 [6], different Access Points act as the key authority for handling the traffic among different nodes. These APs might be operating at different hierarchical levels. The first level APs could be having fixed infrastructure as a backbone. The second level APs could be wireless-based and mobile. The third level hierarchy will be having end user nodes. These APs might be called as servers in the context of our model. The WIFI networks are deficient in the analysis of overhead for the nodes belonging to various groups.

Wireless Mesh Networks [4] consists of many stationary wireless routers that will act as multihop wireless backbone. The mobile clients route their messages through these wireless routers. Like MANETs these networks can also be randomly deployed. There might be many nodes operating under a single wireless router. The mobile clients can use any available wireless router to send their messages to IP backbone. WMN can support a number of online applications like gaming, conferencing, multimedia streaming and broadcasting. There might also be different servers in WMN acting as different groups for various nodes. Like the earlier two networks WMN networks also lack of overhead analysis with respect to inter-group node communication.

The WIMAX networks [7], [10] provide the last mile access with the hotspot size of 48km. The nodes are scattered in the form of different groups under various servers in these networks. These networks also lack the overhead analysis when the nodes from different servers interact and try to establish communicating sessions.

The Bluetooth [9] and Pervasive networks [11], though operate in the limited area, can also fall short of overhead analysis procedures for client-server applications.

3 Proposed Model

In proposed model we assume that the nodes need to contact every other node belonging to a different server. The strength of nodes can be assumed as constant or

variable, it depends. We assume it here as variable. We also assume that the nodes place the messages to each node in other server's domain to cover the whole network. The nodes do not contact nodes belonging to the same server using external messages. They use internal messages with light overhead for this and hence negligible.

We have presented two algorithms for generating and analyzing sequences in this model. The first algorithm named as **Sequencing_Algorithm_for_grouping_nodes** is for the calculation of every possible sequence and on the basis of those sequences, the resulting interaction messages are generated through another algorithm named as **Algorithm_MNPS**. The latter algorithm takes every sequence as input and provides with the output of interaction messages. The administrator or master server can not only calculate all possible sequences to consider all options for its ease of operation but also calculate number of messages through **Algorithm_MNPS** for that sequence calculated in the previous algorithm.

The patterns or sequences are organized in the order of increasing number of messages. An end user node might wish to know about the current sequence pattern, as there are several patterns, in order to analyze number of interactive messages, if the end node is computing those messages. However, in most cases the server is likely to compute them for less overhead. A resourceful server might predict the number of nodes and calculate the number of sequence and messages accordingly. Due to the random entry and exit of the nodes in network, the total number of nodes remains uncontrolled, which requires computing an absolutely different range of sequences and messages? This task lies with the server rather than end nodes. The hierarchical of order of sequences can be helpful for the administrator to control the number of overhead messages for the whole network. The overhead messages tend to rise if the nodes are increasing in the network and scattered as well among different groups. On the other hand, though it might also not serve the purpose, if all nodes get clustered in a single group. The mobility for the nodes is vital and in many sort of networks the nodes are highly dynamic and cannot be restricted to a single platform or server. The abnormal mobility factor also leads to dispersion of nodes and too much clustering factor restricts the functioning of network, so there must be a tradeoff.

Different servers, geographically distributed serve as docking stations or platforms. The mobile nodes might join them while finding themselves in the domain of respective servers. A single master server might maintain the position of nodes and carry out the accounting of overheads. It might also furnish the information to other servers and the nodes come to know about the geographical distribution and physical topology of nodes this way.

As far as algorithm has been concerned, we assume that no group members can interact themselves. However the members can only interact across the groups so the nodes can only interact if divided among various groups. Here, for groups having smaller number of nodes or members and hence large number of groups, the interactive messages for the network grows. The number of interactive messages across various groups tends to be smaller with the increasing number of nodes in a group and hence small number of groups. The size of a group increases at the cost of

reduction in the size of other group by one node/member/unit. The number of interaction messages for a network would be the most if each group comprises only a single member, while the least if there is a single group comprising all of the nodes. For the later scenario, the number of interactive messages can as less as zero.

Table 1. All possible Sequences for total 10 nodes

S/#	Sequences	Total Interaction messages
1.	10	0
2.	9 1	9
3.	8 2	16
4.	8 1 1	17
5.	7 3	21
6.	7 2 1	23
7.	7 1 1 1	24
8.	6 4	24
9.	5 5	25
10.	6 3 1	27
11.	6 2 2	28
12.	6 2 1 1	29
13.	5 4 1	29
14.	6 1 1 1 1	30
15.	5 3 2	31
16.	5 3 1 1	32
17.	4 4 2	32
18.	4 4 1 1	33
19	4 3 3	33
20.	5 2 2 1	33
21.	5 2 1 1 1	34
22.	5 1 1 1 1 1	35
23.	4 3 2 1	35
24.	4 2 2 2	36
25.	4 3 1 1 1	36
26.	3 3 3 1	36
27.	4 2 2 1 1	37
28.	3 3 2 2	37
29.	4 2 1 1 1 1	38
30.	3 3 2 1 1	38
31.	3 3 1 1 1 1	39
32.	4 1 1 1 1 1 1	39
33.	3 2 2 2 1	39
34.	3 2 2 1 1 1	40
35.	2 2 2 2 2	40
36.	3 2 1 1 1 1 1	41
37.	2 2 2 2 1 1	41
38.	3 1 1 1 1 1 1 1	42
39.	2 2 2 1 1 1 1	42
40.	2 2 1 1 1 1 1 1	43
41.	2 1 1 1 1 1 1 1 1	44
42.	1 1 1 1 1 1 1 1 1 1	45

The table 1 shows all sequences for the strength of nodes as 10 in the network. There are 42 numbers of sequences possible with this strength and hence 42 interaction messages. A sequence is composed of a number of groups. The number of values represents the number of groups. These values represent the number of nodes in each group. This range of sequences has been generated by applying algorithm as mentioned below. This is rather a description based approach of defining algorithm pseudocode. As our pseudocode is not the code based algorithm so we have not assumed any data structure like arrays or linked list and this algorithm just helps us display those sequences.

The figure 1 shows the algorithm for sequencing of grouping nodes. This algorithm is self-descriptive form of pseudocode and is composed of 6 steps. The figure 1 shows the how the next sequence will be generated. We can generate the sequences by this algorithm and display them. However, the output of those sequences might not be in order. So we can also sort the sequences on the basis of output in the end so it depends upon the user's requirement. It is already sorted to a large extent. Sorting is only required at few places.

The communication can be maintained among different servers and they may exchange the status of distribution of nodes after a specific time period. At a time only one of the sequences would be enabled. The algorithm gives the complete and every possible range of the sequences within the given number of nodes. The final output messages can be computed at each server. The servers will broadcast the beacon signals and the nodes, if there, will respond to the beacon messages. This will enable the server to compute the number of nodes in its domain. This number will be reciprocated with other servers and all the servers can generate the overall picture of geographical distribution of nodes which will be communicated to nodes after all. There are possibilities that a node is not part of any of the group and is lying at a position with no signals or without any contact with the server. So the overall number of nodes in a network might change any time and we cannot fix a number. For a specific time period, the strength the strength of nodes will be assumed as constant. There should be a reasonable time period. As the greater time period will cease the server tables to be updated and show the right picture of nodes distribution in network. Smaller time period creates a lot of overhead, so it should be a trade off.

The hierarchy of interaction messages shows that the number of these messages is somehow predictable. So the servers can predict that the total messages, will remain around at such and such figure, in case there occurs some change of a single node movement. If there is more dynamism then prediction might be difficult and in some cases might also prove wrong. On the basis of this prediction, the nodes can be notified to arrange or equip their resources in this proportion. There might be some other types of resources besides battery power. We, though, focus on battery power.

Sequencing_Algorithm_for_grouping_nodes {

Step 1. Start checking the sequence shown at Eq (1) from right to left, if there is any value greater than 1 at any position, take it and note down its position.

$$X1, X2, X3, X4,\ldots\ldots\ldots Xn \qquad \text{----- Eq (1)}$$

In the sequence shown at Eq (1), say that position is X3.

Step 2. Take the sum of all of the values from X3 up to Xn (inclusive of X3). Like

$$Xs = X3 + X4 + X5 + \ldots\ldots\ldots + Xn \qquad \text{------Eq (2)}$$

The sum of those values in that sequence is Xs as shown in Eq (2). Suppose the value at X3 is P. Now, how you would make the next sequence.

Step 3. In the next sequence write the values of X1 and X2 as it is in the previous sequence.

Step 4. Subtract a 1 from X3 of previous sequence and write the result Q, i.e., (Q = P – 1), in the position of X3 in the new sequence and subtract it from Xs.

$$Xs = Xs - Q$$

Step 5. Keep repeating this process, by writing Q repeatedly in the next positions and subtracting Q from Xs, till the Xs is greater than Q. If the resultant Xs is smaller than Q then write the remainder value in Xs at the respective position of the new sequence.

Step 6. In the end the Xs comes to null value and the sequence becomes complete. The value at the position Xn might be the same Q or some other value less than Q as remainder (Xs – Q).

}

Consider the sequence No. 34 of table 1.

34. 3 2 2 1 1 1
X1 X2 X3 X4 X5 X6
P = 2 at position X3
While, Xn = X6
Xs = 2 + 1 + 1 + 1
= 5
Q = 2 – 1
Q = 1

Fig. 1. Showing generation of next sequence

The table 2 shows all possible sequences with 6 nodes in the first group for all sequences, assuming 10 nodes in the network. The table 2 is a sub-part of table 1. The algorithm **Sequencing_Algorithm_for_grouping_nodes** generates the sequences in this manner. We can get all these sequences in increasing order by applying sorting algorithm on the output of **Algorithm_MNPS.** The table 1 has been displayed in the same manner.

Table 2. All Possible Sequences for total 10 nodes with 6 nodes in the first group

S/#	Sequences	Total Interaction messages
1.	6 4	24
2.	6 3 1	27
3.	6 2 2	28
4.	6 2 1 1	29
5.	6 1 1 1 1	30

In the same manner, the table 3 shows the generation of all possible sequences for total 10 nodes with 5 nodes in the first group. This is also the sub-part of table 1. The figure 2 displays how the next sequence has been generated. A node switches its place from one group to another which impacts the distribution and hence the sequence gets disturbed. The algorithm would be run after a specific time period to generate the new sequence on the basis of this movement and finally the total number of output interaction messages will be calculated. The nodes after getting this calculated output may predict their future needs in terms of required resources.

Table 3. All possible Sequences for total 10 nodes with 5 nodes in the first group

S/#	Sequences	Total Interaction messages
1.	5 5	25
2.	5 4 1	29
3.	5 3 2	31
4.	5 3 1 1	32
5.	5 2 2 1	33
6.	5 2 1 1 1	34
7.	5 1 1 1 1 1	35

The table 4 shows all possible sequences for a total of 6 nodes to clear the conception about sequences. The figure 2 is also elaborated by taking table 4 as a reference.

Table 4. All Possible Sequences for total 6 nodes

S/#	Sequences	Total Interaction messages
1.	6	0
2.	5 1	5
3.	4 2	8
4.	4 1 1	9
5.	3 3	9
6.	3 2 1	11
7.	2 2 2	12
8.	3 1 1 1	12
9.	2 2 1 1	13
10.	2 1 1 1 1	14
11.	1 1 1 1 1 1	15

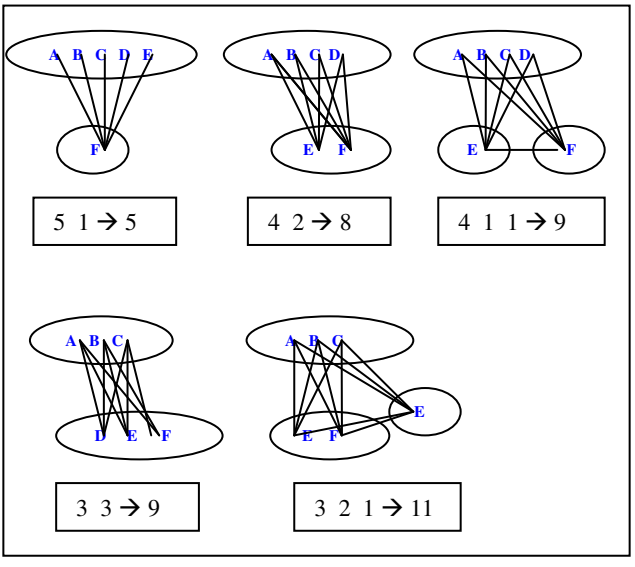


Fig. 2. Showing generation of next sequence

The above algorithm generates the sequences while to calculate the maximum interaction messages or to calculate the cost for establishment of maximum number of sessions among the nodes, another algorithm will be required. The maximum number of interaction messages resulting from the output of sequences, are calculated by an algorithm known as Algorithm_MNPS given as under:

INPUT: An Array A storing N node items.

OUTPUT: Maximum Number of Possible communicating Sessions S among nodes belonging to different servers

```
Algorithm_MNPS (A, N)   {
m = 0, M = 0
For i ← 0 to N - 1    {
                do   M ← M + ( m * A [i] )
                     m ← ( m + A [i] )
    }
```

In the above algorithm, an array 'A' shows the number of nodes under each server group. Let's assume, the nodes in a network belong to three servers. The first server group contains 4; second server contains 5 and the third server as 3 numbers of nodes under them. Then the input to algorithm contains N=3 for three server groups and an array 'A' initialized with 3 values of 4, 5 and 3. The total number of interaction messages 'M' among the nodes of three server groups is 47 as produced by the Algorithm_MNPS at the end of 3rd iteration. We ignore the value of 'm' here. We

take the sequence no. 34 to calculate the maximum possible number of sessions as shown in figure 3.

34. 3 2 2 1 1 1

(3 * 2) +
{(3 + 2) * 2} +
{(3 + 2 + 2) * 1} +
{(3 + 2 + 2 + 1) * 1} +
{(3 + 2 + 2 + 1 + 1) * 1}
= 6+10+7+8+9
= 40

Fig. 3. A sequence generating output of 40

The fig. 3 shows the conversion of sequence into the output of total interaction messages and it is for that sequence only. The figure 4 shows the degree of dispersion for 10 nodes. For 10 nodes, there are 45 interaction messages in the worst case. As the number of groups for a given number of nodes increases, the size of bars for range of messages shrink and the bars tend to position them in the upper right side of the graph relative to the lower left side bar. The figure 4 has been drawn on the basis of output messages as shown in table 1. The range of the first bar is from 9 to 25 for the network with 2 groups only. So the messages for such a group will be required as minimum as 9 and for the maximum as 25.

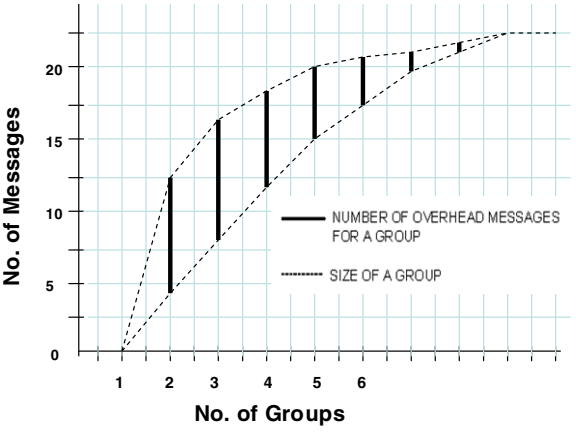


Fig. 4. Graph for Degree of dispersion for 10 nodes

For clarity, the figure 5 has been drawn to exhibit the degree of dispersion for 6 nodes. For 6 nodes, there are 15 interaction messages in the worst case as shown in table 4. The figure 5 is drawn similarly on the basis of the messages shown in table 4.

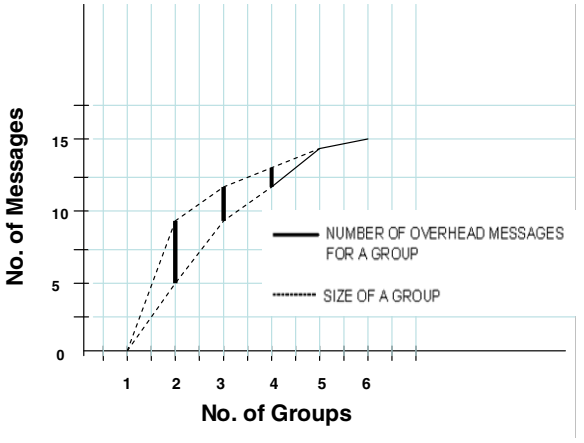


Fig. 5. Graph for Degree of dispersion for 6 nodes

The degree of dispersion for the nodes distribution in the network can be observed from these graphs that tend to rise with the increasing number of groups. The number of groups can be increased from new entry of nodes in the network, or it can also be increased for the constant number of nodes in the network but with high mobility or dynamism. So, the degree of dispersion can be observed in both cases of constant or unfixed number of nodes in the network.

This proposed study can assist the users in better analyzing the overheads. Further we will find the applications of this algorithm in future.

4 Conclusion

This paper signals an application of discrete math towards computer networks. The maximum possible sequences have been generated with the help of algorithm. We can manipulate these sequences to get to the number of total output messages, which can be achieved through another algorithm. In the light of the number of output messages either generated by the nodes or by the servers, the nodes can plan, efficiently utilize their existing resources and foresee the requisite resources that might be needed in the near future.

References

[1] Irshad, A., Shafiq, M., Khurram, S., Gilani, S.M., Irshad, E., Asif, M.S.: An algorithm for strings generation for Regular Expression in a controlled manner. In: IEEE ICIIT, UCP Lhr, ISBN: 978-0-387-25282-7

[2] Irshad, E., Noshairwan, W., Irshad, A., Usman, M., Gilani, M.: Group Mobility in Mobile Ad hoc Networks. In: WWW/Internet, October 13-15, pp. 381–385. IADIS, Germany (2008), ISBN: 978-972-8924-68-3

[3] Buttyan, L., Hubaux, J.-P.: Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks, Mobile Network Applications. Special issue on Mobile Ad Hoc Networks, vol. 8(5), pp. 579–592. Kluwer Academic Publishers, Dordrecht (2003)
[4] Akyildiz, I.F., Wang, X., Wang, W.: Wireless Mesh Networks: A Survey. Computer Networks 47(4), 445–487 (2005)
[5] Irshad, A., Noshairwan, W., Shafiq, M., Khurram, S., Usman, M., Irshad, E.: Security Enhancement in MANET Authentication by checking the CRL Status of Servers. In: DEC 2008, SERSC-IJAST, Korea, vol. 1, pp. 91–98. Springer, Heidelberg (2008) ISSN: 2005-4238
[6] The Wi-Fi Alliance. Website, http://www.wi-fi.org/
[7] TheWiMAX Forum, http://www.wimaxforum.org/home
[8] IEEE 802.11 Standard Group Web Site, http://www.ieee802.org/11/
[9] IEEE 802.15 Standard Group Web Site, http://www.ieee802.org/15/
[10] IEEE 802.16 Standard Group Web Site, http://www.ieee802.org/16/
[11] Stajano, F.: Security for Ubiquitous Computing. John Wiley & Sons, Chichester (2002), ISBN 0470844930
[12] http://courses.csail.mit.edu/6.042/spring10/mcs.pdf
[13] Rosen, K.H.: Discrete Mathematics and Its Applications, 6th edn. McGraw-Hill, New York, ISBN 0-07-288008-2, http://highered.mcgraw-hill.com/sites/0072880082/information_center_view0/
[14] http://www.cs.princeton.edu/courses/archive/fall06/cos341/handouts/mathcs.pdf

TCP Hybla+: Making TCP More Robust against Packet Loss in Satellite Networks

ManKyu Park[1], MinSu Shin[1], DeockGil Oh[1],
ByungChul Kim[2], and JaeYong Lee[2]

[1] Satellite Broadcasting & Telecommuncations Convergence Research Team, ETRI
138 Gajeongno, Yuseong-gu, Daejeon, 305-700, Korea
{neomkpark,msshin,dgoh}@etri.re.kr
[2] Department of Information Communications Engineering, Chungnam National University
79 Daehangno, Yuseong-gu, Daejeon, 305-764, Korea
{byckim,jyl}@cnu.ac.kr

Abstract. Conventional TCP congestion control is based on the fundamental premise that packet loss is an indicator for network congestion only. But actually there are many packet transmission failures in wireless environment because of fading or interference. But this cause of packet loss coming from wireless link has been looked upon as the network congestion. The result of such failure is that TCP will "blindly" halves its sending rate after receiving three duplicated ACKs regardless of the loss occurrence reason. So, TCP will suffer sending rate degradation severely. In this paper, we propose TCP Hybla+ which modifies TCP Hybla to overcome a TCP performance degradation problem by adapting the timestamp option based available bandwidth estimation scheme with one-way TCP data path for asymmetric satellite network. We use the ns-2 network simulator to verify the performance enhancement for the proposed TCP Hybla+. Test results show that the proposed scheme is more suitable TCP than the original TCP Hybla, because the proposed scheme can sustain high sending rate based on the measured available bandwidth for long delay and heavy loss probability environments in satellite networks.

Keywords: TCP Congestion Control, Packet loss, TCP Hybla, Bandwidth Estimation, and Satellite Network.

1 Introduction

Today, the Internet has become the *de facto* global networking infrastructure. IP-based services have increasingly been used for delivering multimedia applications, including various voice, video, and data applications. Such applications generate traffic with different characteristics and consequently require various levels of services. Service differentiation and end-to-end quality of service (QoS) provisioning in IP networks have thus become major preoccupations [1].

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 424–435, 2011.

It is well known that, Satellite networks play a significant role in developing a hybrid (satellite/terrestrial) network infrastructure that can support those emerging applications because of its number of advantages, including global coverage, bandwidth flexibility, reliability, QoS control, and multicast capability. However, satellite network characteristics, for example, long propagation delay, channel impairment, and bandwidth asymmetry, pose new threats to TCP/IP performance [2].

To solve the inherent problems for satellite networks, solutions that cope with these impairments usually rely upon either the adoption of enhanced versions of transport protocols, or the insertion of intermediate agents, like PEPs (Performance Enhancing Proxies). Although PEP allows the end systems to run unmodified and can overcome some problems with TCP window sizes on the end systems being set too low for satellite communications, the end-to-end argument is one of the architectural principles of the Internet [3]. This principle requires that a certain end-to-end function can only be performed by the end systems, not by the network. The application of such a principle has been crucial for the success of the Internet. As a consequence of this principle, most of the potential negative implications associated with the use of PEPs are related to the breaking of the end-to-end of connections, with security problems which use IPsec will be the most critical [2].

Also, conventional TCP protocols have lower throughput performance in satellite networks mainly due to the effects of long propagation delays and high link error rates. That is reason why the TCP protocol was basically designed to recover only from congestion situations, losses of TCP segments due to errors in the transmission link are erroneously ascribed to congestion problems. In recent years, to overcome these problems, several TCP variants have been proposed to improve TCP performance for satellite networks.

TCP-Peach [4] is an innovative congestion control scheme using a novel but simple measurement method based on low-priority data segments which are dummy segments. It introduces the Sudden Start and Fast Recovery scheme instead of the Slow Start and Fast Recovery scheme in conventional TCP. And TCP-Peach is followed through by the proposal for TCP-Peach+ [5], which also exploits the benefits of low-priority data segments, NIL segments, and use the Jump Start and Quick Recovery to occupy network resource quickly. However, since both the schemes put duplicates of already transmitted regular data segments in the low-priority data segments, the associated overhead affects the net realizable performance of the network.

In case of TCP-Cherry, it probes an uncertain satellite link using low-priority data segments, unlike dummy segments in TCP-Peach or NIL segments in TCP-Peach+, and low-priority data segments in TCP-Cherry, named supplement segments, carry data that have not been transmitted yet. TCP-Cherry proposes Fast Forward Start and First Aid Recovery to maintain the high link utilization. They are counterparts of Jump Start and Quick Recovery in TCP-Peach+, respectively [6].

However, those proposed TCP variants are very complex to adopt a real system, because their TCP should be modified TCP sender and TCP receiver all together. But TCP mechanism is one of the most important parts in the Kernel OS, so modification

or installation of the new TCP receiver in an operation system for users is difficult except OS provider like a Microsoft or Linux group.

In this paper, we focus on the mitigation of TCP performance degradation at only TCP sender using a dynamic congestion control algorithm which occurs at the three duplicated ACK received due to satellite wireless link errors. We propose TCP Hybla+ which modified TCP Hybla to overcome a TCP performance degradation problem adapting the TABE (Timestamp-based Available Bandwidth Estimation) scheme for one-way TCP data path in asymmetric satellite networks.

The remainder of the paper is organized as follows: we start with related work for understanding proposed system and mechanisms which are TCP Hybla, TCP New Jersey and ABE (Available Bandwidth Estimation) Scheme in section 2. After that we will introduce our proposed mechanisms about TCP Hybla+ which providing performance enhancement in a large BDP (Bandwidth Delay Product) and high packet loss rate network in section 3. Next, we present our simulation results in section 4, and section 5 concludes the paper.

2 Related Work

In this section, we briefly overview the characteristics of satellite network and its penalty for TCP performance, then we look over the related TCP variants which are improved TCP versions suitable for satellite network such as TCP Hybla, TCP Jersey, and TCP New Jersey for large BDP and leaky networks, and we explain the available bandwidth estimation schemes.

2.1 Satellite Network Characteristics and TCP

The characteristics of GEO satellite network that mainly affect performance of the TCP are the large delay and the available bandwidth variations. In case of GEO satellites system, one-way propagation delay is commonly 250ms and it may be more when forward error correction (FEC) is used. So, TCP round-trip time (RTT) is more than 500ms, which is the time it takes for a TCP sender to determine whether a packet has been successfully received at the receiver. This long propagation delay affects the TCP performance degradation severely because the TCP congestion window (CWND) size is determined by the successful acknowledgement reception per RTT. Thus, the longer the RTT, the narrower the CWND growth and the slower TCP response is experienced.

Actually, many TCP applications are likely to be “short lived TCP [18]” services which are based on the transfer of small files. Thus, it can happen that the entire transfer occurs within the Slow Start phase. It means that a TCP connection is not able to utilize the full bandwidth of the network. Moreover, the classical DVB-S or S2 satellite channels are characterized by higher Packet Error Rate (PER) values. Due to high PER, a large amount of packets are dropped during transmission. TCP demands reduction in its window size to half, even through these packet drops are not caused by congestion.

2.2 TCP Hybla

TCP Hybla[7] is a relatively new approach that has been proposed specifically for high RTT connection like a satellite link. The basic idea of TCP Hybla is to obtain the same instantaneous transmission rate, (*B(t)= W(t)/RTT*), for connections with short or long RTTs using a comparatively fast reference TCP connection (e.g. wired connection which RTT takes 25ms).

Commonly, when a TCP is established a connection between TCP sender and receiver, the TCP sender probes for available bandwidth by gradually increasing the congestion window (*CWND*) value *W*. In the Slow-Start phase, the congestion window is increased by maximum segment size (*MSS*) bytes per non-duplicate ACK received, when a congestion window reaches to the Slow-Start Threshold (*SSThresh*), the TCP sender switches to the Congestion-Avoidance (*CA*) phase, during which the congestion window is increased by *MSS/W* bytes per new ACK received [8]. In a real channel, this rise continues until either the size of receiver buffer (*AWND, Advertised Window*) is reached, or a segment is lost. In this case, action is taken following a recovery procedure that depends on the specific TCP version adopted [9].

By expressing the value *W* of the congestion window in *MSS* units, the standard *CWND* update rules previously described are given by Eq. (1)

$$W_i = \begin{cases} W_{i-1} + 1 & SS \\ W_{i-1} + 1/W_{i-1} & CA \end{cases} \qquad W(t) = \begin{cases} 2^{\frac{t}{RTT}} & 0 \le t < t_\gamma \quad SS \\ \dfrac{t - t_\gamma}{RTT} + \gamma & t \ge t_\gamma \quad CA \end{cases} \tag{1}$$

Where the index *i* denotes the reception of the *i* th ACK. Finally, note that the actual transmission of TCP segments occurs in relation to a window size given by *MIN (CWND, AWND)*, where the advertised window corresponds to the receiver buffer size [10].

TCP Hybla introduce the normalized round trip time, *ρ (RTT/RTTref)*, as the ratio between the actual RTT and the round trip time of the reference connection to which TCP Hybla aims to equalize the performance. In the reference it is shown that the same segment transmission rate of the reference connection can be achieved by longer RTT connections by replacing the standard congestion control rules with the following Eq. (2),

$$W_i' = \begin{cases} W_{i-1}' + 2^\rho - 1 & SS \\ W_{i-1}' + \rho^2 / W_{i-1}' & CA \end{cases} \qquad W'(t) = \begin{cases} \rho 2^{\rho \frac{t}{RTT}} & 0 \le t < t_{\gamma,0} \quad SS \\ \rho \left[\rho \dfrac{t - t_{\gamma,0}}{RTT} \right] + \gamma & t \ge t_{\gamma,0} \quad CA \end{cases} \tag{2}$$

2.3 TCP New Jersey

TCP New Jersey [11] modifies the original TCP Jersey [12] and improves inaccuracy in bandwidth estimation caused by the ACK burstiness for a reverse path.

The original TCP Jersey employs an available bandwidth estimator (*ABE*) using packet fair algorithm [13] at the sender side to probe the end-to-end path bandwidth for the connection and the congestion warning mechanism (CW) such as Explicit Congestion Notification (ECN) scheme to help the sender to effectively differentiate the cause of packet loss at an intermediate router.

This simple yet effective rate estimator with Eq. (3) provides a good guideline for the sending rate when packet drop mostly occurs at the forward link and the reverse link has relatively moderate error rates for ACKs to get throughput. The result from the estimator reflects the desirable sending rate for the forward channel.

$$R_n = \frac{RTT \times R_{n-1} + L_n}{(t_n - t_{n-1}) + RTT} \tag{3}$$

In Eq.(3), where R_n is the estimated bandwidth when n th ACK arrives at time t_n, t_{n-1} is the previous ACK arrival time, L_n is the size of data that nth ACK, and RTT is the TCP's estimation of the end-to-end round trip time delay at time t_n.

In practice, however, the reverse link conditions are not always as good as expected. For instance, when the network transmits bi-directional traffic of a similar amount, or when the reverse link is a wireless channel, which suffers from non-negligible transmission errors, the reliability and accuracy of the result from the ABE module is compromised.

TCP New Jersey for the above weakness use the timestamp based available bandwidth estimation scheme. The TCP timestamp option proposed in RFC 1323[14] provides a readily available solution to overcome asymmetric link problems. In TCP New Jersey, instead of using the ACK arrival time, the packet arrival time, which is stamped by the receiver and echoed back by the ACK according to RFC 1323, is used in the bandwidth estimation. Therefore, in Eq. (3), t_n becomes the arrival time of the nth packet at the receiver. This is equivalent to having the estimation done at the receiver but without the need for the receiver to explicitly feed back the estimation. The estimated available bandwidth by TABE is thus less affected by the reverse link conditions and immune to the ACK loss on the reverse link [11].

2.4 Available Bandwidth Estimation Schemes

A fundamental design philosophy of the Internet TCP congestion control algorithm is that it must be performed end-to-end. But a network is considered as a "black box". A TCP source cannot receive any explicit congestion feedback from the network. Therefore the source, to determine the rate at which it can transmit, must probe the path by progressively increasing the input load until implicit feedback, such as timeouts or duplicate acknowledgements, signals that the network capacity has been reached [15].

TCP-Reno, and its variants detect the available bandwidth of the bottleneck link by continuously increasing the window size until the network is congested and then decrease the window size multiplicatively, e.g., the AIMD algorithm. However, the decrement of window size is rather heuristic and coarse, i.e., halving the current window size. This is because these TCP schemes lack the ability to quantitatively

estimate the available bandwidth before the congestion happens. These TCP scheme's congestion control mechanisms are reactive rather than proactive and preventive [12].

In order to measure the bandwidth used by a TCP flow for calculating and updating the *ssthresh*, we use the measurement method proposed in [11].

When TCP ACK packets arrive at TCP sender as shown in Fig. 1, let E_BW_{k-1} be the estimated bandwidth for the *(k-1)*th ACK. Then, the newly measured bandwidth E_BW_{new} and its moving average E_BW_k become as follows Eq. (4), (5).

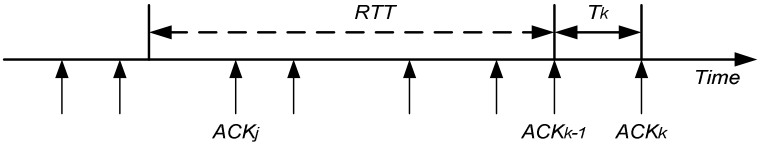


Fig. 1. Bandwidth measurement scheme

$$E\_BW_{new} = \frac{E\_BW_{k-1} \cdot RTT + PacketSize_k}{RTT + T_k} \tag{4}$$

$$E\_BW_k = (1-\beta) \cdot E\_BW_{k-1} + \beta \cdot E\_BW_{new} \tag{5}$$

In the above equation, T_k indicates the time interval between the *(k-1)*th and *k*th ACKs, and $PacketSize_k$ represents the total packet length newly acknowledged by the *k*th ACK. β is a parameter used for moving average calculation ($0 < \beta < 1$) Then, the ssthresh is calculated as follow,

$$ssthreh = \frac{E\_BW_k \cdot RTT}{Max.SegmentSize} \tag{6}$$

To verify the effectiveness of our used ABE and provide a comparative result with the TCP-Westwood's bandwidth estimator, we conducted the following simulation using the ns-2[16] simulator. The simulation topology shows in Fig. 2 which consists of 6 nodes and the bottleneck link between r1, r2. The bottleneck link is configured as a 5Mbps error free duplex link with one-way link delay of 40ms.

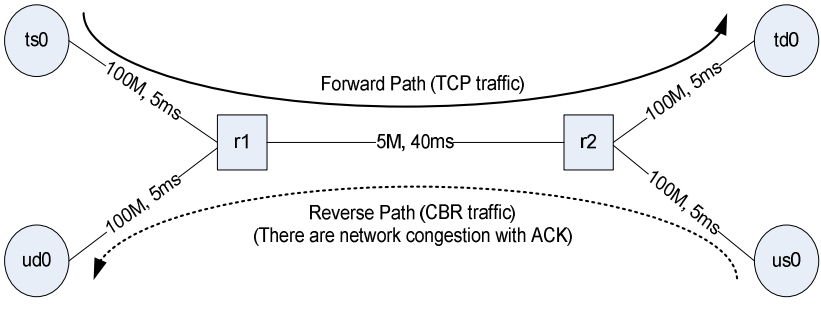


Fig. 2. Simulation Topology for a test of the effectiveness of bandwidth estimators

Simulation results are shown in Fig. 3. Each plotted point on the curve represents the goodput with increasing background traffics. In case of TCP Westwood and TCP Jersey, a bandwidth estimator is used to measure the available bandwidth, and then the estimator is used to control the packet sending rate at TCP source.

Although those TCP assume that there are same network conditions between TCP data path and TCP ACK path, DVB-S2/RCS network is asymmetric network. When the offered load for background traffic is increased, bandwidth estimation values are affected from congestion situation in the reverse link. In TCP Vegas, bandwidth estimation scheme base on RTT variation controls the packet transmission rate. As mentioned with other TCP variants, TCP Vegas is also affected from the reverse congestion state. This is the reason that TCP variants did not originally clearly distinguish between the forward and reverse congestion situation.

However, TCP Newreno and TCP New Jersey can sustain the packet sending rate regardless increasing background traffics. TCP Newreno did not calculate CWND based on available bandwidth estimation scheme, So TCP Newreno did not reduce CWND unnecessarily. For TCP New Jersey, CWND value is computed using the Timestamp based available bandwidth estimator, and this scheme is able to estimate the available bandwidth for forward link only.

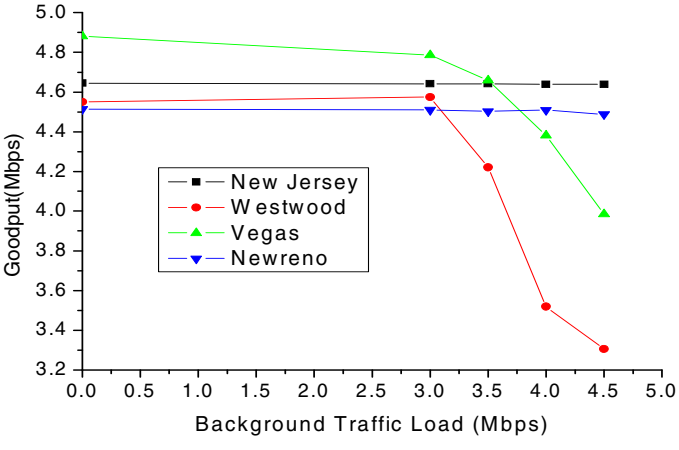


Fig. 3. Comparison of Goodput according to various TCP versions

3 TCP Hybla+: Our Proposal

TCP Hybla [7] has been proposed specifically for high RTT connection like a satellite link. Although we find TCP Hybla which was modeled from TCP reno with RTT 25ms to be quite effective even in case of high RTT, we observe that there is still room for proposing a more robust solution to yield better TCP performance over a satellite link.

To introduce our proposed TCP Hybla+, we first analyze features of the original TCP Hybla to find out key factors which affect TCP goodput for TCP Hybla.

First of all, Although TCP Hybla was designed to overcome a long propagation delay in satellite radio links; TCP Hybla had too aggressive CWND increment during Slow-Start period. For satellite communication environment, ρ value (RTT/RTT_{ref}) in TCP Hybla becomes 19.2 (0.48/0.025 = 19.2), and W_2 was calculated with 602,248 ($=1+2^{19.2}-1$) when TCP sender got the first ACK packet in Slow-Start period. To avoid setting this unrealistic value to an initial CWND, TCP Hybla in the real state uses Hoe's modification method [17]. But when TCP Hybla adapts Hoe's modification method to set an initial CWND value, it will be same to an *ssthresh* (Slow-Start threshold) value. This situation means that TCP Slow-Start does not work for original exponential increment but it will be action as the jump to *ssthresh* at Slow-Start begin, it can be called "jump start [4]". This TCP action may look good because TCP can achieve the available bandwidth in very short time. However if there are multiple connections staring up at approximately the same time, or other large volume traffic joins in when a connection is in Slow-Start, the Hoe's modification will set the initial *ssthresh* too high, resulting in multiple loss and coarse timeout.

Second, TCP Hybla was suitable for a satellite network but it lacks which is useful special congestion control scheme when packet loss occurs. As we know that the satellite network has a high packet loss probability and TCP Hybla does not possess the capacity to distinguish and isolate congestion loss from wireless loss. As a consequence, TCP Hybla reacts to wireless loss with a drastic reduction of congestion window.

For these problems, we fist adjust the RTT_{ref} value to 70ms from 25ms. In [18], author introduced the median RTT for traffic to be approximately 70ms from set of experiments. When the proposed TCP used 70ms for the reference RTT, ρ value (RTT/RTT_{ref}) becomes 6.8 (0.48/0.07 = 6.8), and value of W_2 is 111.4 ($=1+2^{6.8}-1$), therefore we can expect the Slow-Start period to probe the gradually increment for available network capacity when TCP Hybla+ is used. We adapted the congestion control mechanism using the timestamp based available bandwidth estimator (TABE) [10] for considering the bandwidth asymmetric of a satellite network. The measured available bandwidth is used to calculate the CWND value after detecting packet loss with 3-duplicated ACKs.

4 Evaluation

In this section, we evaluate and compare the proposed TCP Hybla+ with the original TCP Hybla through simulations. We consider the DVB-S2/RCS [19] GEO based test scenario. The topology used for simulation is shown in Fig. 4 which is simulated using network simulator ns-2[16].

In this topology, there are two earth stations, RCST, HUB, between N hosts with a TCP sender and N hosts with a TCP receiver. Each sender or receiver host is connected to the each station. We consider GEO system model, it is defined by DVB-RCS standard [19], the link HUB-to-RCST is called forward link and the link RCST-to-HUB is called return link. We also assume that the forward link (TCP sender to receiver for TCP data segments) capacity is equivalent to 10Mbps and the return link (TCP receiver to TCP sender for TCP ACK segments) is 5Mbps. All the results

shown in this section have been obtained by considering the system behavior for connection duration of 500ms.

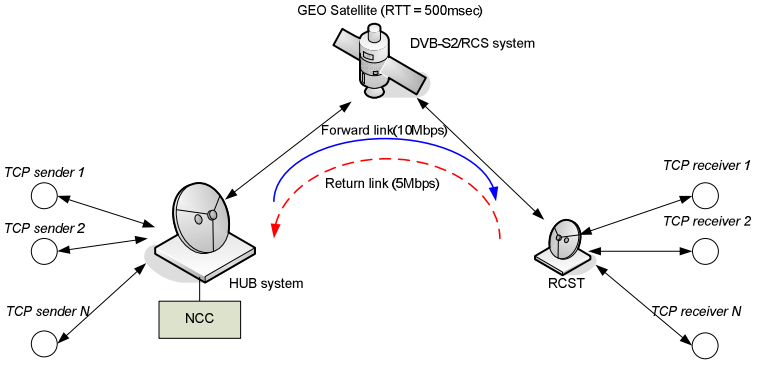


Fig. 4. Simulation scenario for the DVB-S2/RCS system

4.1 Simulation Results

Let us refer to the results of the comparison of TCP congestion windows between the conventional TCP Hybla and the proposed TCP Hybla+ which uses the time stamp based available bandwidth estimator (TABE) and dynamic congestion control scheme. The proposed TCP implemented base on the ns-2 source code from [20] for TCP Hybla.

The goodput results of the performance evaluation are that the conventional TCP congestion window is a multiple decrease because there are a lot of packet loss and out of order packet sequences. Those problems lead the congestion window size to sharply drop, and the TCP then enters the slow-start phase again.

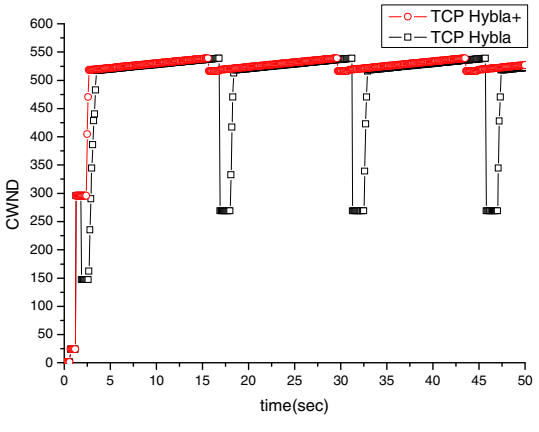


Fig. 5. Comparison of TCP CWND values (PER = 0%)

In Fig. 5, we can see that congestion window dynamics of TCP Hybla+ at 0% PER (Packet Error Rate) sustains the high packet sending rate using adaptive decrement based on the TABE scheme, however the conventional TCP Hybla halves its sending rate involved in static factor *(CWND = CWND/2)* whenever meet three duplicated ACKs.

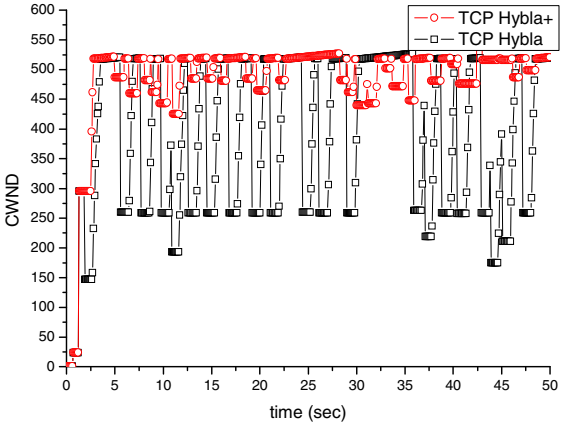


Fig. 6. Comparison of TCP CWND values (PER= 0.1%)

In Fig. 6, we can see that the goodput performance for TCP Hybla+ is better than TCP Hybla when wireless link PER = 0.1%. The proposed TCP does not use the blindly static factor, but it can calculate a properly CWND value based on TABE. However, the conventional TCP Hybla cannot distinguish between the packet loss due to wireless link error or wireless link congestion.

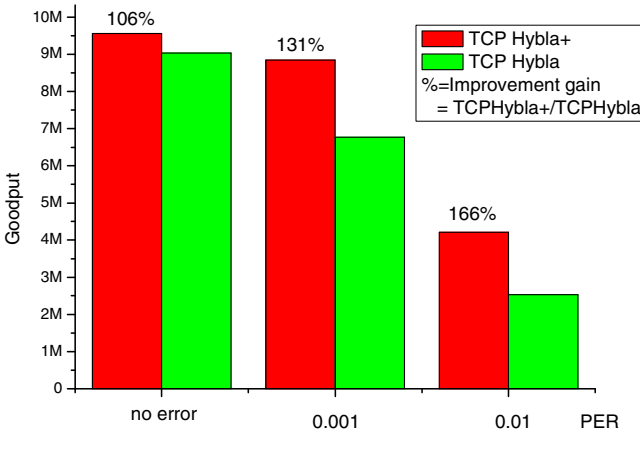


Fig. 7. Comparison of TCP Goodput with various PER

In this experiment, we analyze the goodput performance results for various PER between TCP Hybla+ and TCP Hybla in Fig. 7. Performance of TCP Hybla+ at no link error is very close to the TCP Hybla. Even when the segment loss probability is 10^{-3} or 10^{-2}, i.e., at high link error rates, goodput of TCP Hybla+ can attain about 31% and 66% higher goodput (Improvement gain=*TCPHybla+/TCPHybla*) then TCP Hybla respectively.

Consequentially, the proposed TCP Hybla+ works efficiently at high transmission rate as compared to the original TCP Hybla.

5 Conclusion

In this paper, we proposed a new TCP scheme for dynamic congestion control, called TCP Hybla+, to improve the TCP performance in the Large BDP and leaky satellite networks.

TCP Hybla+ adopts the proper ρ value (RTT/RTT_{ref}) using the new reference RTT (=70ms). It needs to prevent buffer overflow from the packet overshoot for Slow-Start duration and we use the dynamic congestion control algorithm based on the available bandwidth estimator which is a modified TABE algorithm with timestamp option enabled for asymmetric network, DVB-S2/RCS system, to optimize the congestion window value when sender received three duplicated ACKs.

Our simulation shows that under no error to 10^{-2} PER, TCP Hybla+ outperforms the original TCP Hybla in goodput from 6% to 66% respectively. These simulations show results that proposed TCP is more suitable TCP than original TCP Hybla, because the proposed scheme is able to distinguish congestion losses from wireless error. Hence it can sustain high sending rate under the long delay and high loss probability in satellite network.

Acknowledgments. This work was supported by the IT R&D program of KCC [2009-S-039-02, Development of Satellite Return Link Access Core Technology for High Efficient Transmission].

References

[1] Pace, P., Aloi, G.: Effective Admission Policy for Multimedia Traffic Connections over Satellite DVB-RCS Network. ETRI Journal 28(05), 593–606 (2006)
[2] Hassan, M., Jain, R.: High Performance TCP/IP Networking: Concepts, Issues and Solutions, pp. 206–237. Pearson Prentice Hall, USA (2004)
[3] Carpenter, B. (ed.): Architectural principles of the internet. RFC 1958, IETF (June 1996)
[4] Akyildiz, I.F., Morabito, G., Palazzo, S.: TCP peach: A new congestion control scheme for satellite IP networks. In: IEEE/ACM (2001)
[5] Akyildiz, I.F., Zhang, X., Fang, J.: TCP Peach+: enhancement of TCP Peach for satellite IP networks. IEEE Communications Letters 6(7) (July 2007)
[6] Utsumi, S., Zabir, S.M.S., Shiratori, N.: TCP-Cherry : A New Scheme of TCP Congestion Control for Satellite IP Networks. Computer Communications 32, I.12 (July 2009)

[7] Caini, C., Firrincieli, R.: TCP Hybla: a TCP Enhancement for Heterogeneous Networks. International Journal of Satellite Communications and Networking 22(5), 547–566 (2004)
[8] Allman, M., Stevens, W.: TCP congestion control. RFC 2581, IETF (April 1999)
[9] Floyd, S., Henderson T.: The New Reno modification to TCP's fast recovery algorithm. RFC 2582, IETF (April 1999)
[10] Transmission Control Protocol, Darpa Internet Program. RFC793 (September 1981)
[11] Xu, K., Tian, Y., Ansari, N.: Improving TCP performance in integrated wireless communications networks. Computer Networks 47(2), 219–237 (2005)
[12] Xu, K., Tian, Y., Ansari, N.: TCP-Jersey for Wireless IP Communications. IEEE Journal on Selected Areas in Communications 22(4), 747–756 (2004)
[13] Prasad, R.S., Murray, M., Dovrolis, C., Claffy, K.: Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. In: IEEE Network (November/December 2003)
[14] Jacobson, V., Braden, R., Borman, D.: TCP extensions for high performance. Request for Comments 1323 (May 1992)
[15] Casetti, C., Gerla, M., Mascolo, S., Sanadidi, M.Y., Wang, R.: TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links. In: Proceedings of Mobicom 2001, Rome, Italy (July 2001)
[16] ns-2, http://www.isi.edu/nsnam/ns/
[17] Hoe, J.C.: Improving the Start-up Behavior of A Congestion Control Scheme for TCP. In: Proc. ACM SIGCOMM 1996, pp. 270–280 (1996)
[18] Chu, J., Dukkipathi, N., Cheng, Y.: Increasing TCP's Initial Window (2010), draft-hkchu-tcpm-initcwnd-00.txt
[19] ETSI, Digital Video Broadcasting(DVB); Interaction channel for satellite distribution systems. EN 301 790, V1.5.1 (2009)
[20] TCP Hybla Homepage, http://hybla.deis.unibo.it/

A Secure Privacy Preserved Data Aggregation Scheme in Non Hierarchical Networks

Arijit Ukil and Jaydip Sen

Innovation Lab, Tata Consultancy services
BIPL, Sector-5, SaltLake, Kolkata – 700091, India
{arijit.ukil,jaydip.sen}@tcs.com

Abstract. Privacy is an important issue in today's context of extreme penetration of Internet and mobile technologies. Recently, large-scale data collection and integration efforts increased privacy concerns. It becomes very challenging to ensure the protection of user's private data. In this paper, we address this issue from a practical deployment point of view. We have considered a non hierarchical or flat wireless sensor network, where there are several sensor/sink nodes reporting to a query server and the server aggregates the data sent by the sink nodes, while preserving the privacy of the data. The philosophy of this work is that the server nodes cannot be able to find out the content of the data of the sink nodes. It can only perform the aggregation operation and execute further processing on the aggregated data only. The scheme presented in this paper has the advantages of low computation time even in the presence of large number of sink nodes. The proposed scheme unlike many other published schemes is highly scalable [3, 10]. We demonstrate through simulation results the efficacy of our scheme in terms of privacy disclosure probability and computational efficiency.

Keywords: security, privacy-preservation, data aggregation, wireless sensor networks, in-network processing, set top box.

1 Introduction

The immense growth and advancement of the information age, data collection and data analysis have exploded both in size and complexity. This in turn has impacted on the privacy preservation of the data of individual user or the network itself. Privacy in our context can be defined as the control over access to information about oneself. Privacy is also the limited access to a person or a process and to all the features related to it. Privacy preservation is important from both individual as well as organizational perspectives. For example, customers might send to a remote database queries that contain private information. Two competing commercial organizations might jointly invest in a project that must satisfy both organizations' private and valuable constraints, and so on. In his seminal paper [1], Yao has introduced the millionaire problem, which can be summarized as follows: Let *A* and *B* are two millionaires who want to find out who is richer without revealing the precise amount of their wealth. This problem is analogous to a more general problem where there are

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 436–451, 2011.

two numbers a and b and the goal is to whether the statement $a \geq b$ is true without revealing the actual values of a and b. Another important issue is data aggregation. In sensor networks, data aggregation is useful, particularly when in-network processing [2] is required to minimize the traffic volume. There are numerous use cases where aggregated data result is important and the individual data values are to be kept private. Like the case of rating of television viewership, where the aggregated sum viewership result of a particular program is the requirement for the surveying authority. But the advertisers or other parties may be interested on the viewership details of the individual for their business interest. If these parties can access the micro details of individual viewership pattern, the privacy of the individual viewers is severely violated.

In this work, we develop our scheme for data-oriented privacy preservation with total preservation of user data. In this paper, we propose a scheme which provides privacy-preserving secure data aggregation. We assume a non hierarchical model of the network like that of finding viewership rating of a TV program. We consider a subset of a wireless sensor network where the data aggregation is performed in single-hop topology with no peer-to-peer connection between the nodes. In the process we consider the algorithm proposed in [3] for privacy preservation and improve its performance in terms of reduced computation time and lesser probability of private data disclosure with the help of an appropriate key management scheme. When we consider the case of other hierarchical wireless sensor networks, our proposal is a sub-set of the solution for the multi-hop case with only linear increase of the complexity with the increment of number of hops. Our proposed scheme has the following advantages over [3] under the assumed scenario:

1. Less computational overhead, hard real-time applications can be executed.
2. Peer-to-peer connection between nodes is not required. Simple client-server model is sufficient.

The paper is organized as follows. In section 2, we discuss the available existing works related to our proposed scheme. In section 3, we describe the motivation of this work. In section 4, we present the system model. In section 4, we describe our scheme, which consists of three parts: the cluster formation algorithm, the key management and privacy is preservation. In section 5, we show through computational complexity analysis that our scheme is much more practical in the assumed scenario. Even in the scenario, where the assumed scenario is a sub-set, our scheme performs better in terms of computational complexity than the one described in [3]. In section 6, we present the simulation results and analysis. In section 7, we conclude the paper.

2 Related Work

Privacy preservation has been studied for a long time in the data mining discipline. Substantial effort has been spent for developing privacy preserving techniques with data mining algorithms in order to prevent the disclosure of sensitive information during the knowledge discovery process. The objective of privacy preserving data mining is to meet the required privacy requirements and to provide data mining

outcome [4]. Privacy-preserving data mining finds numerous applications in surveillance which are naturally supposed to be "privacy-violating" applications. The key is to design methods, which continue to be effective, without compromising security [5]. Privacy preserving schemes and protocols are designed to preserve privacy even in the presence of adversarial participants that attempt to gather information about the inputs of their peers and mostly with malicious intention [6]. There are three major classes of privacy preservation schemes are applied. In the first category of techniques, user data is authenticated while marinating the anonymity of the user [7]. This technique exploits the concept of group signature-based authentication. The second types of mechanisms are based on data perturbation approach, where certain distribution is added to the private data. In the third category, randomized data is used to mask the private values. However, data perturbation techniques have the drawback that they do not yield accurate aggregation results. It is noted by Kargupta et al. [8] that random matrices have predictable structures in the spectral domain. In additive perturbation, randomized noise is added to the data values. The overall data distributions can be recovered from the randomized values. Another approach is multiplicative perturbation, where the random projection or random rotation techniques are used in order to perturb the values. In tune of their argument [8], we apply the second technique of masking the private data by some random numbers to form additive perturbation. Along with privacy preservation, another important thing needs to be considered is the data aggregation requirement. Data aggregation is an efficient mechanism in query processing in which data are processed and aggregated within the network [9]. In-network processing is forwarding the raw data from the sender/sink nodes in a processed form, by reducing redundancy or by extracting information out of the data. In [10, 11], the authors reviewed privacy-preserving techniques for protecting two types of private information: data-oriented and context-oriented. The proposed scheme is for data-oriented privacy preservation.

3 Motivation

As noted in the previous section, privacy preserving data aggregation has many practical use cases and deployable scenarios. Number of research proposals and algorithms exist [10]. In [12], the problem of privacy preservation in a peer-to-peer network application is addressed. He et.al. [3] propose schemes to achieve data aggregation while preserving privacy. The scheme they proposed, CPDA (Cluster-based Private Data Aggregation) performs privacy-preserving data aggregation in low communication overhead with high computational overhead in a self-organized multi-hop wireless sensor networks. CPDA in spite of its elegance and efficiency suffers from two critical limitations:

1. Computation of the privacy preservation algorithm increases with the increase in number of sink nodes.
2. In most of the practical scenarios, the sink nodes cannot communicate directly with each other in a peer-to-peer mode. In such cases, usefulness of CPDA is doubtful.

If we consider the case of privacy preserved viewership computation of Television Rating Points (TRP) computation, we observe that these two limitations forbid CPDA to be useful. TRP is the criterion that indicates the popularity of a channel or program and this data is very useful for the advertisers [13], which has good amount of business impact. TRP is generally calculated by frequency monitoring of the audience by preparing an aggregated data on the viewership to find the statistics of different channels at different location and different time. For the TRP calculation [13] perspective, individual viewership statistics is not required, the aggregated viewership value of a particular location of particular channel is sufficient. There is a scope that individual viewership in the most granular for is recorded and utilized for commercial purpose. But the viewers might not be willing to share the information on their profile of channel viewing. So, the aggregation of individual viewer's data at the service provider end needs to be done in a privacy protected manner.

The objective is that each node should be assured that its private data is only known to itself. Furthermore, the data aggregation scheme should not only be able to preserve privacy, but also be able to handle to some extent the typical security attacks. Furthermore, wireless links may be eavesdropped by attackers to reveal private data. Our proposed private data aggregation scheme should be robust to such attacks.

In view of that, we present our scheme which is called SPPDA (Simplified Privacy Preserving Data Aggregation). This scheme assumes single-hop or non hierarchical network topology without peer-to-peer communication between client/sink nodes, like that of the cases for TRP generation. In SPPDA, the privacy preserving algorithm is based on [2], but the complexity becomes very less. In this scheme, sink nodes do not require peer-to-peer communication. Our scheme consists of three parts:

1. Cluster formation
2. Key management
3. Privacy preserving data aggregation through data value distortion

In rest of the paper, we present this scheme and show how it is superior to CPDA in the assumed scenario. First, we present the system model assumed for the scheme. Then we describe the key management scheme. After that the privacy preservation scheme will be discussed, in which we present the scheme for data value distortion based on [3] and subsequent data aggregation. We have made few substantial contributions in this work. We propose one self-adaptive cluster formation to facilitate privacy preservation and one key management scheme for the assumed scenario. We modify the scheme as in [3], to suit it for practical purposes. The proposed scheme involves less overhead than in [3]. In fact, the computational complexity in [3] increases with the addition of sink nodes and cluster, where as the complexity of the proposed scheme is independent of the number of sink nodes.

4 System Model

In this paper, we assume a single-hop wireless sensor network (WSN) with no peer-to peer communication between the client/sink nodes. This network can be considered as a sub-set of a bigger multi-hop WSN, but for the sake of simplicity single-hop WSN model is presumed in the rest of the discussion. We evaluate the performance of

SPPDA with respect to a single-hop WSN architecture. There are three types of nodes in the network: base station (BS), aggregator, and sink node. The BS is the node who answers the queries. Hence it is the node where aggregation result is destined. The aggrgator or server node has the responsibility of data aggregation and further processing of the aggregated data and then to send the result to the BS. This server node has connection with N number of sink nodes, which are connected with the server node through wireless links. These sink nodes collect the data on its own or as per the instruction by the server node. It is assumed that the sink nodes do not have peer-to-peer connectivity. If one sink node likes to establish a connection with other sink nodes, it has to do it through the server node only. This in fact, eliminates the scalability infeasibility problem of the previous proposals [3]. This scenario is shown below. A data aggregator function is defined as:

$$Y(t) = f(d_1(t), d_2(t), \ldots, d_N(t))$$

Without loss of generality, we consider sum function, though typical aggregation functions like MEAN, MIN, MAX can also be included.

$$Y(t) = \sum_{n=1}^{N} d_n(t)$$

In order to communicate with other sink nodes, the sender/sink node requests to the server node. The server node performs forwarder function. In fact, to implement the privacy preserving policy, one sink node needs to communicate with other one. This is done securely by pair-wise key establishment technique. As it is noted that in many practical cases, this pair-wise key establishment and direct communication between the sink nodes are not possible. So, we propose one efficient key management scheme, which takes care of the typical security issues.

Other than that, it is assumed that the communication between server and sink are accurate without loss of data. It is also assumed that sink nodes are always in the communication range with the server. In the event the failure of the sink node with the server node, the server node notifies the event to the neighbors of the failed node.

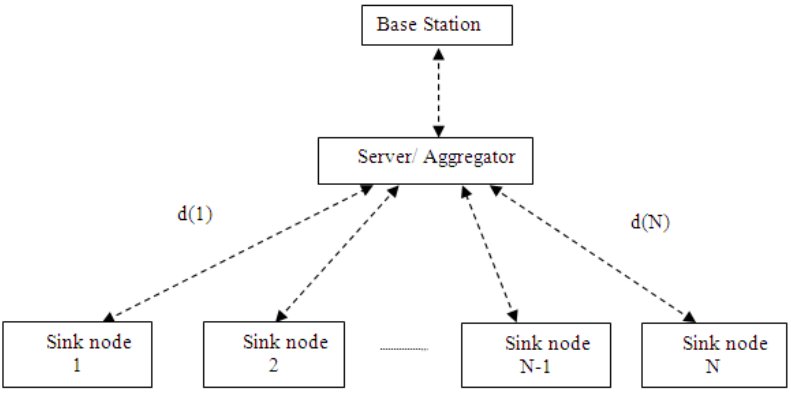


Fig. 1. SPPDA syster

4.1 Cluster Formation

In order to provide scalability of the system, secure communication between the nodes and minimum probability of privacy disclosure, an efficient cluster formation is required. This cluster formation helps in providing security of sink to sink communication, which eventually leads to successful implementation of the privacy preservation algorithm without direct sink to sink communication. These sink nodes collect data and send it to the server. The cluster formation algorithm is as follows:

1. The server broadcasts *HELLO* message. Up on receiving the response from the sink nodes, the server gets the idea of the active nodes.

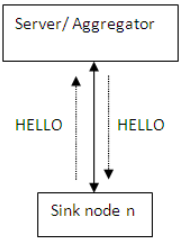


Fig. 2. Cluster formation step 1

2. Each cluster consists of four sink nodes reports to the server node. So, there are *N/4* number of cluster, where *N* = number of active nodes. If *N/4* is not an integer, few clusters with five nodes (total number of clusters having five sink nodes <=*3*) is formed. This process of cluster formation is initiated by the server/aggregator and is done with the help of the client nodes.

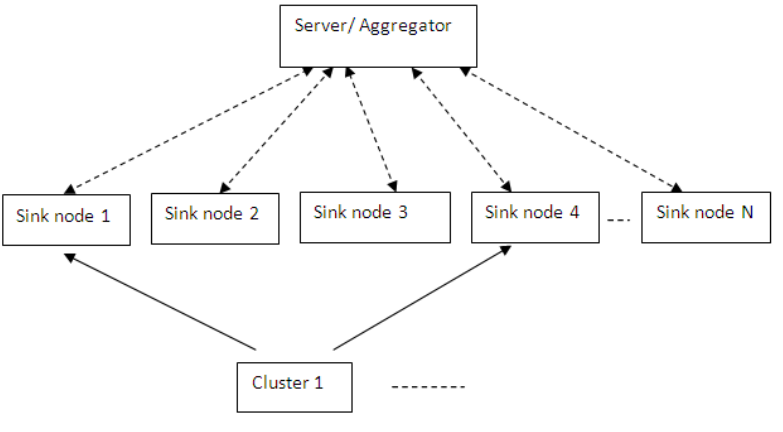


Fig. 3. Cluster formation step 2

3. This cluster formation information is shared among the sink nodes by the server. The other nodes of the cluster the sink node belongs are its neighbors. The server selects the sink node pair within the cluster and informs that to the sink nodes. For example, let us consider the cluster formation in Fig. 3. The server selects sink node *1* and *2* as a pair and sink node *3* and *4* as a pair and these nodes form cluster 1. These node pairs are denoted as: pair *11* and pair *21*. This is shown in Fig. 4. These pairs are termed as friend pair.

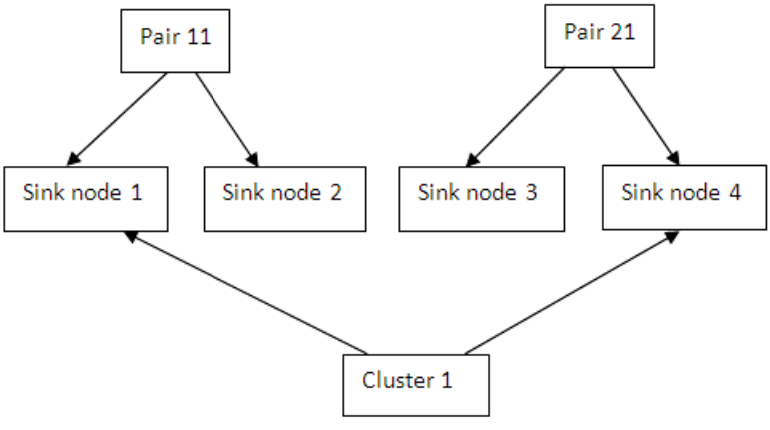


Fig. 4. Friend pair formation

4. The server keeps track with the status of the node by periodical *HELLO* message. In case of failure of any node to respond and if that repeats, the node is declared as passive or dead node. This requires server to regroup the friend pair formation, this may include 3 sink nodes within one pair. Server informs the corresponding sink nodes accordingly. The objective of the server is not to disturb the cluster after it is formed. In the scenario when one node enters the network, the server will try to accommodate it inside one of the clusters if that cluster has less than four members. Otherwise, that node is temporarily accommodated in one of the clusters. When four of such sink nodes available which have temporary membership of some clusters, these four forms a new cluster.

4.2 Key Management

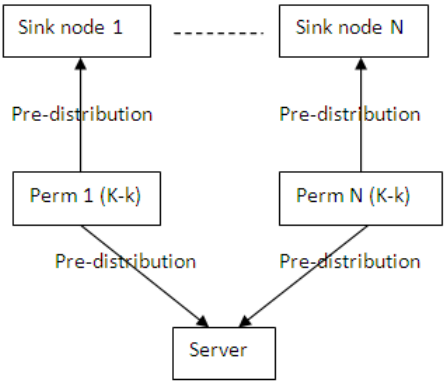
To provide support for ensuring privacy and integrity of messages sent from sink nodes to their corresponding aggregator or server, robust key exchange and management scheme is required [14]. In [15], Eschenauer and Gligor proposed one random key pre-distribution for secure key distribution in sensor networks. In this work, the philosophy is to present a key management scheme designed to satisfy operational and security requirements of a non hierarchical, single-hop sensor

network by selectively distributing and removing keys from sensor nodes (including sink nodes and server) as well as re-keying nodes without substantial computations or bandwidth usage. The objectives of our key management scheme are as follows:

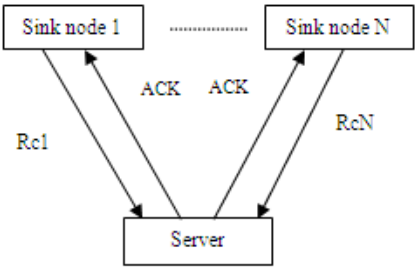
1. The scheme must establish a key between all the sensor nodes that must exchange data securely.
2. Node addition or deletion should be supported.
3. It should work in undefined deployment environment. Unauthorized nodes should not be allowed to communicate with any other nodes in the network.

In order to accomplish these objectives, we first form cluster of the sink nodes. Let, there be N number of sink nodes and each cluster consists of n number of sink nodes. So, there will be N/n number of clusters. The key management process starts by key pre-distribution stage. In the pre-distribution phase, a large key-pool of K keys and corresponding identities are generated. These K number of keys are divided into two banks. One bank consists of k number of keys, which is used for sink node's communication with other sink nodes via the aggregator and $(K-k)$ keys for sink nodes communication with the server/aggregator. So, the key management scheme consists of two parts:

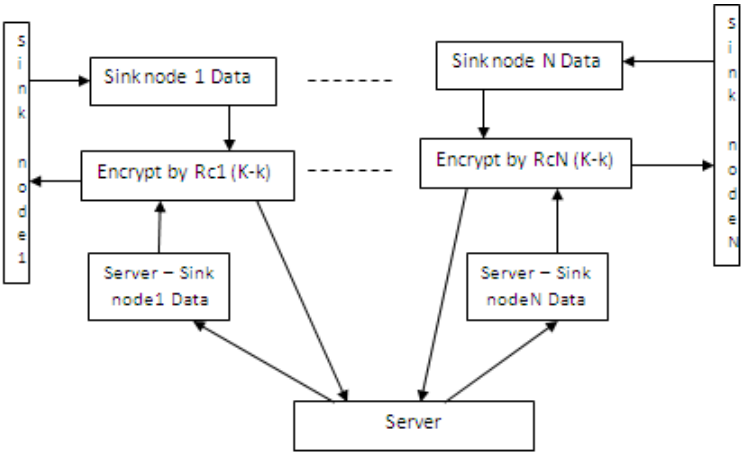
1. Sink node to server key establishment: It is described that each sink node has $K-k$ number of keys shared with the server. As, all the sink nodes possess the same keys, it is totally unsecure when a sink node communicates with the server node with the shared key. Any malicious sink node can decipher the sink nodes' communication with the server and can launch attack very easily. In order to avoid this, in the pre-distribution phase, the sink-server key bank is randomly permuted and reordered for each sink-server pair. This ordering of the key bank is stored in the server for each sink. This is shown in Fig. 5(A). When a new sink node is added, same procedure of key bank randomization is followed and the key order is stored offline for the new sink. Now, the sink node communicates with the server through one of its shared keys. To accomplish this action, the sink node first generates a random number between 1 and $(K-k)$. This random number (Rc) is sent to the server in plain text. The server understands that the sink node will encrypt the next message by the Rc^{th} number key of the key bank. The server responds by sending ACK. This is shown in Fig. 5(B). After receiving ACK, the sink node encrypts the data with the particular key and sends to the server. Every time the sink node likes to communicate with the server, it does the same steps. The random number generation process at each session hinders the probable guessing attack. It can also be observed that the random number (Rc) is sent to the sink in plain text. But this does not arise any vulnerability issue, as getting the random number by a malicious node does not harm because of the pre-randomization of the key bank order. The Rc^{th} number key is different in different sink nodes. The mapping is stored in the server offline in pre-distribution stage.



(A)



(B)



(C)

Fig. 5. Server to sink node key establishment and secure communication

2. Sink node to sink node key establishment: It is assumed that sink to sink direct communication does not exist and this has to happen through the server/aggregator. It is also to be noted that the server should not decipher the communication between the sinks; otherwise the privacy preservation algorithm becomes trivial for server to break. In order to achieve that, the *k* number of keys are pre-stored in the sink nodes at the time of deployment, which the server is unaware of (server is only aware of *(K-k)* keys required for server-sink communication). It is also a requirement that other sink nodes should not decipher the message sink 1 sends to sink *2*. As the k keys are same for all the sink nodes, it becomes easy for another sink node to decrypt the plain text, i.e. sink *3* can decrypt what sink node *1* and sink node *2* are communicating. To avoid this situation, sink node 1 and sink node *2* separately permute the key bank order of the k number of keys dedicated for sink-sink communication and reorder that randomly. After that, they pass the permute function to each other through the server using their pair-wise key with the server. After successful delivery of permute functions, one of the sink nodes (sink node *1*, for example) sends another random number between *1* and *k* to the other sink node (sink node *2*), which indicates the particular key of the permuted key bank. This pair-wise key between sink nodes will be used for the subsequent communication until the data aggregation is complete. For next round of data aggregation process, same key establishment procedure will be followed.

4.3 Privacy Preservation

After establishing the secure communication, we describe the privacy preservation algorithm. This privacy-preservation data aggregation policy is based on the additive property of the polynomial [3]. Good amount of research works have been done to find privacy preserving data aggregation; like using modular arithmetic through secure multi party computation [16]. The objective of these algorithms is that the server or the aggregator can not make out the individual content of the data sent be the sink node. But the published methods require peer-to-peer communication between client/sink nodes, which is not available in most practical scenarios like TRP computaion among STBs.

In this work, we modify the CPDA scheme under practical constraints. Our proposed SPPDA scheme modifies, simplifies and transforms the CPDA algorithm. More specifically, with intelligent cluster formation algorithm, efficient key mangemnt and proper privacy preserving algorithm, our proposed SPPDA scheme solves the practical constraints.

In the system model described, the friend pairs‘ data are aggregated together. After receiving the aggregated data of all the friend pair the server sends that to the base station. It is shown in the Fig. 1. In order to illustrtae this, we assume server/aggregator as node *‘A‘* and two sink nodes of the friend pair is *‘S1‘* and *‘S2‘*.This algorithm consists of two parts:

1. Value distortion: Let the data values in the sink node *S1* and *S2* be *x* and *y* and *z* be the dummy variable at the aggregator node *‘A‘*. In the first step, the server/aggregator sends three seeds a,b and c to the friend pairs. Based on that *A* computes

$$\begin{aligned} \alpha_{S1}^{A} &= z + R_1^A b + R_2^A b^2 \\ \alpha_{S2}^{A} &= z + R_1^A c + R_2^A c^2 \\ \alpha_{A}^{A} &= z + R_1^A a + R_2^A a^2 \end{aligned} \tag{1}$$

where R_1^A and R_2^A are two random numbers generated by *A*.
Similarly, *S1* computes

$$\begin{aligned} \alpha_{S1}^{S1} &= x + R_1^{S1} b + R_2^{S1} b^2 \\ \alpha_{A}^{S1} &= x + R_1^{S1} a + R_2^{S1} a^2 \\ \alpha_{S2}^{S1} &= x + R_1^{S1} c + R_2^{S1} c^2 \end{aligned} \tag{2}$$

Similarly *S2* computes

$$\begin{aligned} \alpha_{A}^{S2} &= y + R_1^{S2} a + R_2^{S2} a^2 \\ \alpha_{S1}^{S2} &= y + R_1^{S2} b + R_2^{S2} b^2 \\ \alpha_{S2}^{S2} &= y + R_1^{S2} c + R_2^{S2} c^2 \end{aligned} \tag{3}$$

where R_1^{S1} and R_2^{S1} are two random numbers generated by sink node *S1*, R_1^{S2} and R_2^{S2} are other two random numbers generated by sink node *S2*. After that, the calculated, α_{S1}^{A} and α_{S2}^{A} are sent to sink node *S1* and sink node *S2* by A, securely as described earlier. Similarly, $\alpha_{A}^{S1}$and α_{S2}^{S1} are sent to sink node *S2* and A by sink node *S1* and α_{A}^{S2} and α_{A}^{S2} and α_{S1}^{S2} are sent to *A* and sink node *S1* by sink node *S2*.

2. Value aggregation: After the private data values (*x* and *y*) are distorted, all the nodes aggregates the values available to them and generates aggregated result. Sink node calculates Ψ_{S1} , sink node *S2* calculates Ψ_{S2} and A calculates Ψ_A from (1- 3).

$$\begin{aligned} \Psi_A &= \alpha_A^A + \alpha_A^{S1} + \alpha_A^{S2} = (x+y+z) + R_1 a + R_2 a^2 \\ \Psi_{S1} &= \alpha_{S1}^A + \alpha_{S1}^{S1} + \alpha_{S1}^{S2} = (x+y+z) + R_1 b + R_2 b^2 \\ \Psi_{S2} &= \alpha_{S2}^A + \alpha_{S2}^{S1} + \alpha_{S2}^{S2} = (x+y+z) + R_1 c + R_2 c^2 \end{aligned}$$

where, $R_1 = R_1^A + R_1^{S1} + R_1^{S2}$ and $R_2 = \mathrm{R}_2^{\mathrm{A}} + \mathrm{R}_2^{\mathrm{S1}} + \mathrm{R}_2^{\mathrm{S2}}$. These aggregated results from sink node *S1* and sink node *S2* are securely sent to the aggregator A. Now, the aggregator has the simple task to solve the above equation for $(x+y+z)$ with the knowledge of the values of *a,b,c* and Ψ_A , Ψ_{S1} and Ψ_{S2} . After solving for $D = x+y+z$, node A internally knows its own data z, so it can find out the result $(x+y)$.

5 Complexity Analysis

In this section, we compare the complexity of the proposed SPPDA scheme with CPDA scheme proposed in [3].

5.1 Cluster Formation

In CPDA [3] cluster formation stage has a complexity of *O(NcP)* i.e. pseudo polynomial in *P*, where *Nc* stands for the probability of a node independently becoming a cluster leader. This is done each time cluster formation is required in order to counter the dynamics of the sensor networks.

In SPPDA, this has fixed value, as $P = 2$ and $Nc = 1$. It can be noted that if P is high, the complexity of the CPDA scheme becomes high, as it is linearly proportional to P.

5.2 Broadcasting Seeds

Broadcasting of seeds in [3], within a cluster takes $O(k)$ messages, where k is the cluster size; this is done in each cluster. So overall complexity becomes *(number of clusters)* x $O(k)$.

In SPPDA, $O(k)$ is fixed, as $O(k)$ takes fixed three messages only and instead of broadcast, multicast is required by aggregator/server to the friend pair. So, this complexity is also fixed for SPPDA.

5.3 Encrypt and Send Computed Values

In CPDA, every node within a cluster sends a message to every other node within the cluster and this happens in all the clusters. So, message complexity is *O (cluster size* x *pseudo polynomial order)*, i.e. $O(kP)$.

In SPPDA, k and P are both fixed with values 3 and 2 respectively.

5.4 Aggregate Information

In CPDA the message complexity is simply the size of cluster in each cluster. So, overall complexity is the order equal to number of nodes in a cluster *(P)* x *number of clusters (K),* which is $O(PK)$.

In our proposed scheme, $P=2$, this complexity is linear to the number of clusters, i.e. $O(K)$.

6 Simulation Result

In this section, we show a comparative study between our proposed SPPDA scheme and the CPDA scheme in [3].

The objective of our work is to find a simpler, efficient privacy preserved data aggregation scheme, which has scalability and can be highly effective in some practical scenario as discussed in Section 3.

We evaluate the performance of SPPDA for multiple runs and compare the mean results with CPDA in an Intel Core 2 duo PC with CPU speed 3 GHz and RAM of 2 GB. The comparison is shown in Table 1.

Table 1. Mean computational time in millisecond for different number of nodes for SPPDA and CPDA

SPPDA	CPDA	Number of nodes
40	220	3
50	250	4
51	310	5

It is observed from Table 1 that when the number of sink nodes increases, the computational time in SPPDA remains almost constant, where as in CPDA the computational time increases more rapidly. This is shown in Fig. 6, which also reveals that in SPPDA the computational time is substantially low than that of CPDA. Computational time difference between these two schemes becomes more visible when more number of nodes is considered. This is because of the fact that in SPPDA, for most of the cases (not considering the failure of a certain sink node, where three sink nodes may need to form a cluster), there will be fixed two number of sink nodes involved in the privacy preservation computation, the computational time becomes fixed. This is indeed a necessary requirement when the overall system is real-time in nature. In CPDA algorithm, the computational time increases almost linearly with the addition of number of nodes. SPPDA on the other hand, incurs very less computational load and does not have much affect with the addition of number of clients/ sink nodes. As in CPDA, with number of nodes increases, the computational time increases, we limit the number of client nodes to five. It is also impractical in CPDA to have large number of sink nodes in a single cluster. It is to be noted that we have compared only the algorithm performance.

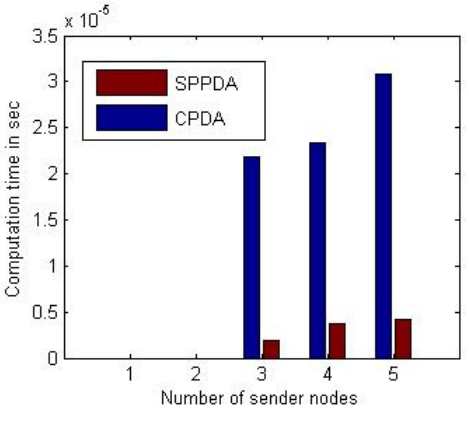


Fig. 6. Computation time requirement

In CPDA scheme, there exists certain probability where private data may be disclosed. This can only happen when the sink nodes exchange messages within the cluster. This can be estimated as

$$P(b) = \sum_{m=pc}^{Dmax} P(k=m)(1-(1-b^{m-1})^m)$$

Where *Dmax* = maximum cluster size, *pc* = minimum cluster size *(= 2*, two sink nodes), *k* = cluster size, *b* = probability that link level privacy is broken, *P(k=m)* = probability that a cluster size is m. In the case of SPPDA, *pc = Dmax = k = 2, P(k=m) =1*. So, we have plotted *P(b)* for CPDA and SPPDA in Fig. 7. It is observed

that the probability of privacy compromised in CPDA has much steeper slope than that of SPPDA.

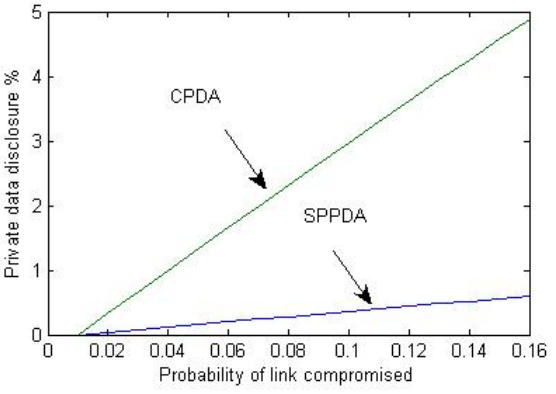


Fig. 7. Probability of private data disclosure

In CPDA, a requirement is that a pair of sink nodes possessing same pair of keys, where the keys are taken randomly from a large pool of key, should be high. Otherwise, the scheme cannot work. But, this requirement helps other nodes to capture at least some of the communication, if it has common pair of keys. Consider the scenario in Fig. 8. There are three nodes, node *1*, node *2* and node *3*. The key between node 1 and aggregator is M_1that between node *2* and aggregator is M_2. same way the key between node *3* and aggregator, node *4* and aggregator are M_3 and M_4 respectively. According to CPDA:

$$\begin{aligned} P(M_1 = M_{2|3|4}) &= \varepsilon_1 \, (0 \le \varepsilon_1 \le 1) \\ P(M_2 = M_{1|3|4}) &= \varepsilon_2 \, (0 \le \varepsilon_2 \le 1) \\ P(M_3 = M_{1|2|4}) &= \varepsilon_3 \, (0 \le \varepsilon_3 \le 1) \\ P(M_4 = M_{1|2|3}) &= \varepsilon_4 \, (0 \le \varepsilon_4 \le 1) \end{aligned} \quad (4)$$

Where *P(X)* means probability that X is true. From (4), it is understood that certain probability (may be very small) exists that same keys may be shared between the nodes in CPDA. This violates the strict security requirement in certain use cases. However, in SPPDA, key distribution is totally deterministic and avoids this kind of security vulnerability. But this requirement calls for more number of keys in the key pool.

In fact, this probability of key shared between the nodes increases with number of sink nodes increase when total number of keys in the key pool is constant. This is shown in Fig. 9. In SPPDA, there is no requirement like that, the key management policy is such that total randomization is done for key distribution and the nodes form friend pair is controlled by the trusted aggregator. The number of nodes forms a friend pair is determined by the server, unless one of the sink nodes becomes passive or

dead. Elimination of this kind of situation that a pair of nodes finds common key is not deterministic makes SPPDA more secure and robust against different attacks.

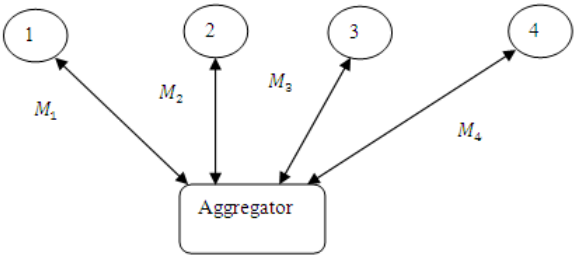


Fig. 8. Pair-wise key distribution

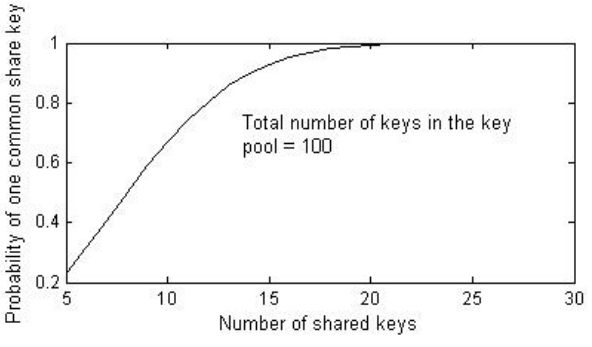


Fig. 9. Probability that a pair of nodes possess one common key

7 Conclusion

In this work, a scheme is presented to perform secure privacy preservation data aggregation. The scheme is developed from the requirement of a practical scenario of non hierarchical, single-hop sensor network without the requirement of communication between client/sink sensor nodes. We compared the performance of the scheme with the one described in [3]. Through analysis and simulation results, it is shown that the scheme is more efficient and scalable. It is found that in most of the practical scenarios, direct sink-to-sink communication in peer-to-peer mode does not exist. This restriction is unsuitable for most of the current privacy preserving data aggregation scheme [3]. SPPDA has intelligently taken care of this constraint by proper key management and self-organized adaptive cluster formation. The advantage of less computational time results in smooth running of real-time applications. However, SPPDA is developed for single-hop structured sensor networks. Our future work includes generalizing the scheme in multi-hop sensor networks without compromising any of the advantages of the SPPDA scheme.

References

1. Yao, A.: Protocols for secure computations. In: Proceedings of the 23rd Annual Symposium on Foundations of Computer Science, pp. 160–164 (1982)
2. Sorniotti, A., Gomez, L., Wrona, K., Odorico, L.: Secure and Trusted in-network Data Processing in Wireless Sensor Networks: a Survey. J. Inf. Ass. Sec. 2, 189–199 (2007)
3. He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.T.: PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks. In: IEEE INFOCOM, pp. 2045–2053. IEEE Press, New York (2007)
4. Oliveira, S.R.M., Zaïane, O.R.: Achieving privacy preservation when sharing data for clustering. In: Jonker, W., Petković, M. (eds.) SDM 2004. LNCS, vol. 3178, pp. 67–82. Springer, Heidelberg (2004)
5. Sweeney, L.: Privacy Technologies for Homeland Security. Testimony before the Privacy and Integrity Advisory Committee of the Department of Homeland Security, Boston, MA, June 15 (2005)
6. Ukil, A.: Security and Privacy in Wireless Sensor Networks. In: Smart Wireless Sensor Networks, Intechweb, Croatia, pp. 395–418 (2010)
7. Sen, J.: An Efficient and User Privacy-Preserving Routing Protocol for Wireless Mesh Networks. International Journal Scalable Computing: Practice and Experience 11(4), 345–358 (2010)
8. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: Random-data perturbation techniques and privacy-preserving data mining. Knowledge and Information Systems 7, 387–414 (2005)
9. Chan, H., Perrig, A., Przydatek, B., Song, D.: SIA: Secure Information Aggregation in Sensor Networks. J. Com. Sec. 5(1), 69–102 (2007)
10. Li, N., Zhang, N., Das, S., Thuraisingham, B.: Privacy-preserving in wireless sensor networks: A state-of-the-art survey. In: Ad Hoc Networks, vol. 7, pp. 1501–1514. Elsevier, Amsterdam (2009)
11. Conti, M., Zhang, L., Roy, S., Di Pietro, R., Jajodia, S., Mancini, L.V.: Privacy-preserving robust data aggregation in wireless sensor networks. In: Secur. Commun. Netw., vol. 2, pp. 195–213 (2009)
12. Huang, Q., Wang, H.J., Borisov, N.: Privacy-preserving friends troubleshooting network. In: Symposium on Network and Distributed Systems Security, pp. 184–194 (2005)
13. Television Audience Measurement, `http://www.tamindia.com`
14. Sen, J.: A Distributed Trust and Reputation Framework for Mobile Ad Hoc Networks. In: Meghanathan, N., Boumerdassi, S., Chaki, N., Nagamalai, D. (eds.) CNSA 2010. Communications in Computer and Information Science, vol. 89, pp. 538–547. Springer, Heidelberg (2010)
15. Eschenauer, L., Gligor, V.D.: A key-management scheme for distributed sensor networks. In: 9th ACM Conf. Computer and Communication Security, pp. 41–47. ACM, New York (2002)
16. Ukil, A., Sen, J.: Secure multiparty privacy preserving data aggregation by modular arithmetic. In: International Conference on Parallel Distributed and Grid Computing (PDGC), pp. 344–349. IEEE Press, Los Alamitos (2010)

An OWL-Based Context Model for U-Agricultural Environments

Yongyun Cho[1], Sangjoon Park[2], Jongchan Lee[2], and Jongbae Moon[3,⋆]

[1] Information and Communication Engineering, Sunchon National University, 413 Jungangno, Suncheon, Jeonnam 540-742, Korea
yycho@sunchon.ac.kr
[2] Dept. of Computer Information Engineering, Kunsan National University, 1170 Daehangno, Gunsan, Jeonbuk, 573-701, Korea
lubimia@kunsan.ac.kr
[3] Supercomputing Center, Korea Institute of Science and Technology Information, 52-11 Eoeun-dong, Yuseong-gu, Daejeon 305-806, Korea
comdoct@shinbiro.com

Abstract. Services in ubiquitous agricultural environments could be more elaborate and autonomous by intercommunicating with various sensors and systems, based on the service architecture of the device-to-device or device-to-service. To do that, the situation information that sensors gain from agricultural environments should be recomposed into contexts, which can be understood or recognized by services and devices. In this paper, we propose an OWL-based context model for agricultural environments, which can be used as service execution conditions for various context-aware service applications based on ubiquitous sensor networks in u-agriculture. The proposed context model is based on OWL ontology, which can be easily and efficiently customized in the various fields of agricultural service domains. With the proposed context model, developers can easily implement context-aware agricultural services through idiomatic communication with device-to-device, sensor-to-sensor, or sensor-to-service in agricultural environments. Especially, we hope that the proposed context model can be greatly helpful in the developments of smart agricultural services in the cultivation environments equipped with various sensors such as greenhouse, glasshouse, and vertical farm and so on.

Keywords: context-aware, OWL, ontology, context model, agriculture.

1 Introduction

Conventionally, services in ubiquitous computing environments aim for autonomous processes in which humans do not intervene. Context-aware services in ubiquitous computing environments based on RFID/USN sensors need an appropriate context model and an ontology for a high-level context reasoning

⋆ Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 452–461, 2011.

[1]. A context model and a ontology have to represent possible contexts occurred in specific service domains such as an u-city, a smart home, an u-port and so on. with low-level contexts, profiles, or data from sensors. Recently, work conditions in agricultural environments have moved to be more autonomous and smart. Especially, large-scaled farming systems equipped with various agricultural sensors and telecommunication networks, such as greenhouse, glasshouse, and vertical farm, can automatically control and monitor the growth and development processes of crops in order to gain high productivity and safe growing of crops [2]. Commonly, the situation conditions in agricultural environments are realtime, irregular, and independent each other. So, to get a context-aware and smart agricultural service, we needs a context model to represent the situation conditions with unified data sets and an ontology to define the knowledge shared for agricultural environments.

This paper introduces an OWL-based context model for agricultural environments. The suggested model represents the situation conditions in agricultural environments with RDF-based contexts, and defines the meanings or the relations of the contexts through an OWL-based agricultural ontology. The suggested context model has structural architecture incrementally to gain higher-level contexts through the rule-based composition of low-level contexts, referencing the OWL-based agricultural ontology. By virtue of the structural architecture, because the suggested model's context architecture can be easily modified, the suggested model has high scalability, when the types of contexts and the scopes of knowledge for real agricultural environments are changed. Therefore, the suggested model can be used as a potential and reasonable context model to develop various context-aware service applications in agricultural environments.

2 Related Work

TempCRM [3] proposes a reasoning model for temporal situation conditions in such critical places as smart home and smart office based on RDF(resource description framework) and OWL(Web ontology language). TempCRM defines a OWL-based ontology for reasoning of various temporal contexts which can be occurred in a smart home domain. Through the scenario test with the temporal contexts, TempCRM proves that it can be considerable for using the temporal contexts around smart home environments as execution conditions of appropriate services. CONON [4] is another research for OWL-based ontology to describe contexts in pervasive computing environments. CONON introduces a context model to reason a higher-level composite context from low-level individual entities for situation conditions. To do that, it suggests an OWL-based ontology reasoning rules. CONON is implemented by a prototype consisting of a description logic based ontology reasoner and a first-order logic based situation reasoner according to the rules. TempCRM and CONON may be meaningful researches in the point that they consider ontology-based context model to use situation conditions as important data for the service automation in ubiquitous computing and pervasive computing environments.

Recently, there are many research for context-aware workflow services using context for situation information, which can be obtained from both users and their surroundings [5,6,7]. Those research also consider a reasonable context model to be used as the transition conditions of services. FollowMe [8] uses a workflow model to provide context-aware workflow services. FollowMe introduces a scenario-based workflow model to handle user's demands of services and situation conditions. It introduces a system architecture to have a separate reasoning engine to get rule-based contexts from the situation information and uses CPDL (Compact Process Definition Language) and XPDL (XML Process Definition Language) to control a service flow. As a context-aware workflow language, CAWL [9] is another research to use context-aware workflow technologies in such smart spaces as smart home, smart office, and so on. CAWL can define contexts for real situation conditions directly into workflow a service scenario through RDF-based context triplet to a workflow model to provide context-aware workflow services. Figure 1 shows an architecture of CAWL.

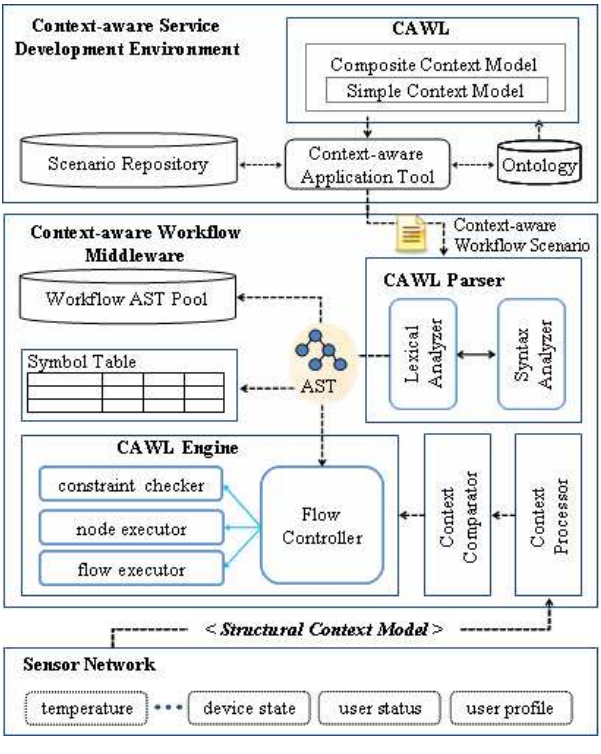


Fig. 1. An architecture of CAWL

3 OWL-Based Context Model for Agriculture

3.1 An Architectural Layer of the Suggested Context Model

Situation information occurred in ubiquitous computing environments is composed of the information such as profile, location, time, and so on.

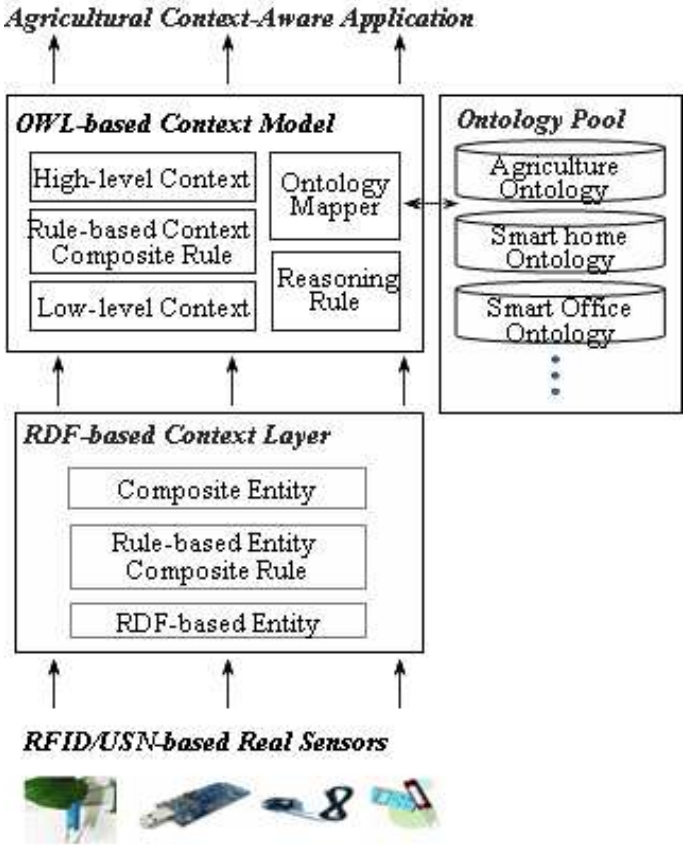


Fig. 2. A conceptual architecture of agricultural services using OWL-based context model

Figure 2 illustrates a brief conceptual architectural layer for supporting context-aware services based on a OWL-based context model in agricultural environments. As shown in Figure 2, the suggested context model consists of two context model layer. The RDF-based Context Layer gathers raw data from various real sensors based on RFID/USN in agricultural environments, and converts them into an entity consisting of a set of the data type and value. Then, the RDF-based Context Layer composites several entities according to a rule-based composite conditions, which consists of *and*, *or*, *not*, and *exclusive or*.

After the entity composition, it produces a RDF-based context in the triplet form of subject-predicate-value. However, the triplet context does not include more meaningful data such as the shared knowledge about the terms of the RDF-based contexts the relation information among the entities. In Figure 2, the OWL-based Context Model Layer adds the useful additional information into the RDF-based contexts according to the shared domain knowledge defined in domain-specific ontology. In this time, several low-level contexts can be composed to a higher-level context by a rule-based context composition and reasoning rule using a domain-specific ontology. As a result, in the OWL-based Context Layer produce a higher-level context including the relation information about class-hierarchies, subclasses, individuals, properties, reverse objects of a low-level context can be produced.

3.2 Context Composition Using the OWL-Based Context Model

The suggested context model includes a context composition process to translate a real sensed data into RDF-based context, and incrementally into OWL-based context in order to offer a high-level context. Figure 3 shows a triplet structure of a RDF-based context to represent real data in the RDF-based context model of the suggested context model.

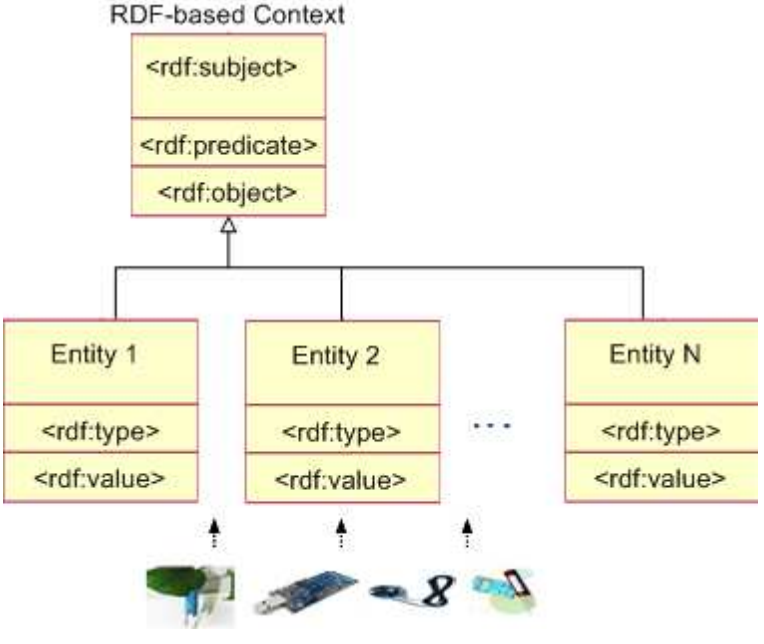


Fig. 3. A triplet structure of the suggested RDF-based context model

As shown in Figure 3, a situation data from real sensors are translated into a entity set of <rdf:type> and <rdf:type>. Then, several entity sets are composed into a RDF-based context which is a triplet of <rdf:subject>, <rdf:predicate>, and <rdf:object>. Figure 4 shows an example of a RDF-based context expression for a possible sensed data in u-agricultural environment.

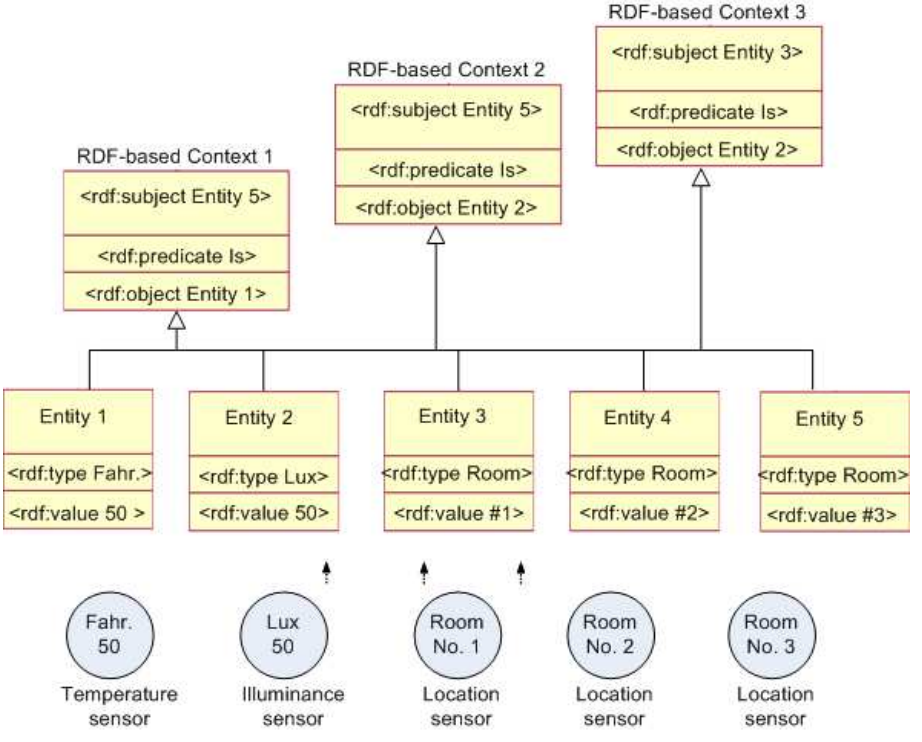


Fig. 4. An example of the context expression using the suggested RDF-based context model

As shown in Figure 4, several entities are used for the composition into appropriate RDF-based contexts. In Figure 4, RDF-based context 1 illustrates that it composes Entity 1 and Entity 2 to indicate that the temperature of the room no. 5 is Fahr. 50. To do that, it can be constructed with a triplet set of <rdf:subject Entity 5>, <rdf:predicate Is>, and <rdf:object Entity 1>.

In the same way, Entity 2 is produced as the triplet set of <rdf:subject Entity 5>, <rdf:predicate Is>, and <rdf:object Entity 2> to represent that the illumination of the room no. 5 is 50 Lux. And Entity 3 is produced as the triplet set of <rdf:subject Entity 3>, <rdf:predicate Is>, and <rdf:object Entity 2> to represent that the illumination of the room no. 3 is 50 Lux. Those RDF-based contexts can be composed into a high-level context through OWL-based

ontology context model. Figure 5 shows the suggested OWL-based context model for several RDF-based context occurred in u-agricultural environment.

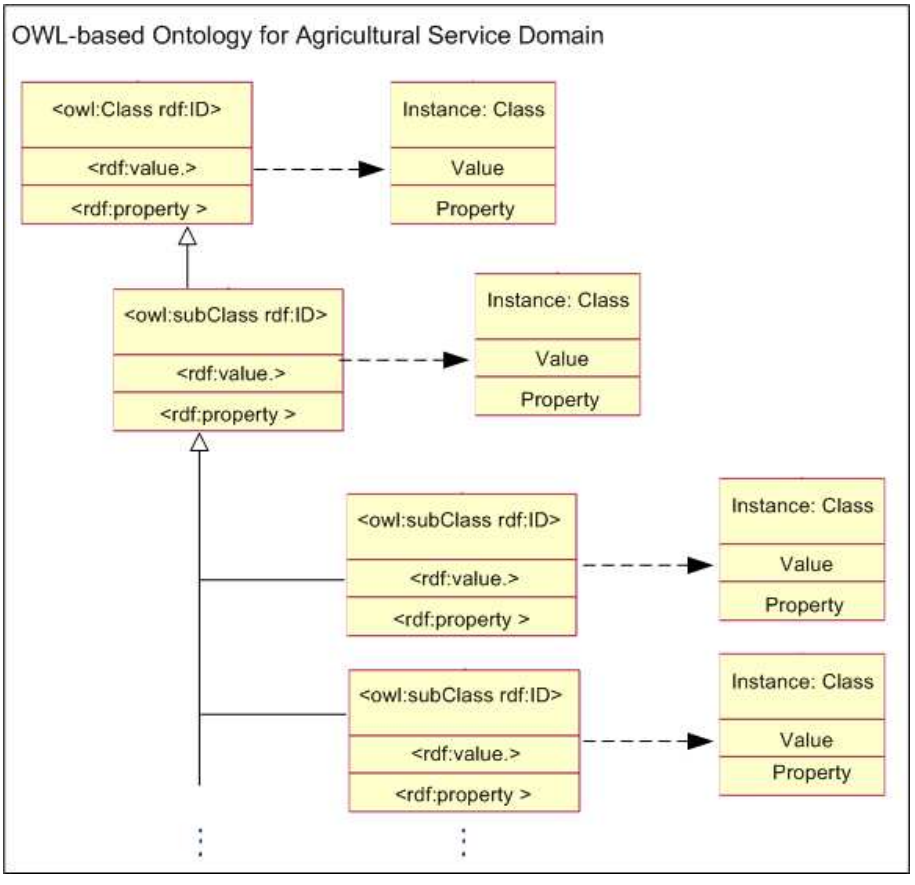


Fig. 5. The suggested OWL-based context model

In Figure 5, a <owl:Class> can hierarchically be constructed with several <subClass>s. And, each <owl:Class> consists of <rdf:value> and <rdf:property>, and includes more than one instance of the class type. In this time, a <owl:Class> become a type of the instance, and the instance can be a specific individual as a subject or a object of RDF-based context triplet. For example, to consider a context for the temperature of the greenhouse no. 3, the instance of the class type of Greenhouse, which may be a subclass of the Building or Organization class, is the greenhouse no. 3 as a individual. In the same way, the instance of the class type of Fahr., which may be a subclass of the Temperature class, is the 50, or any number, Fahr. as a real sensed data.

4 Experiments and Results

In this section we will represent real sensed data from various agricultural sensors as contexts through the suggested OWL-based context model and a reasonable agricultural ontology. First of all, we need a reasonable ontology for u-agricultural domain. To do this, we design it with protege [9,10] which is one of the useful tools to develop an OWL-based ontology. Figure 6 shows an agricultural ontology using OWL.

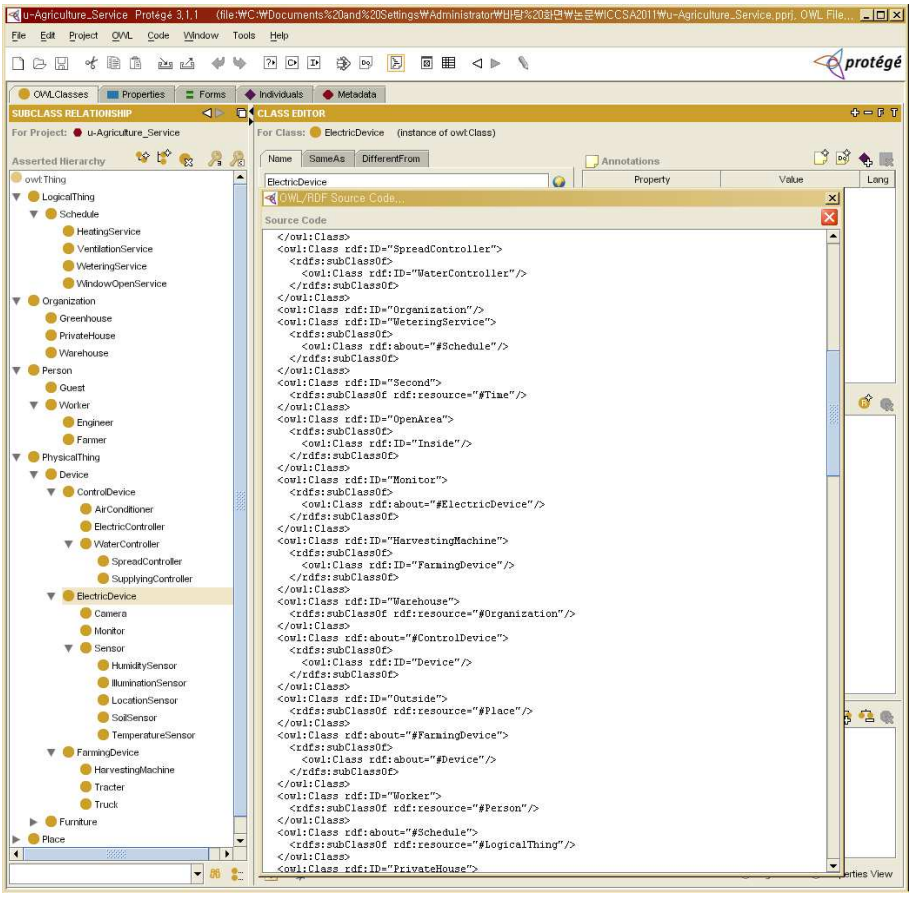


Fig. 6. An agricultural ontology using OWL

In figure 6, the Sensor class includes the TemperatureSensor class, IlluminationSensor class, LocationSensor class, HumiditySensor class, SoilSensor class as subclasses. Those sensor classes have several individuals and properties that are needed when a high-level context is composed with RDF-based contexts.

Figure 7 shows context a high-level context meaning that the temperature of Greenhouse no. 3 is 50 Fahr. using the agricultural ontology in Figure 6.

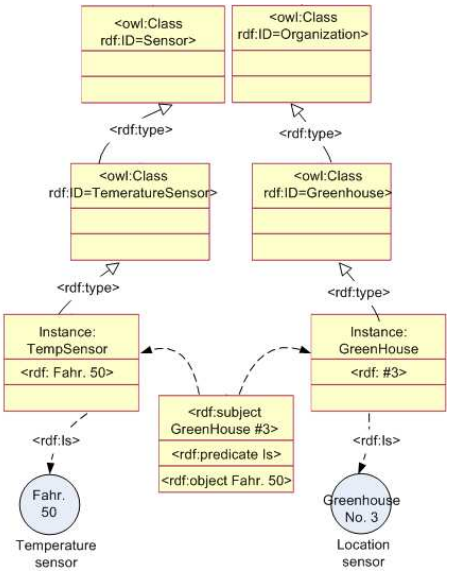


Fig. 7. A sample high-level context using the agricultural ontology

5 Conclusion

In this paper, we proposed an OWL-based context model for u-agricultural environment. For this, we introduced RDF-based context model to represent real sensed data from agricultural environments with a triplet. And, we designed an OWL-based agricultural ontology. Because the suggested context model is based on RDF context model and OWL ontology, it has a high scalability for the expansion of the shared knowledge in the agricultural service domain. By using the suggested context model, a developer can easily and efficiently develop a various context-aware service application in u-agricultural environments. Especially, the suggested context model can be useful for service automation in the types of device-to-device or service-to-service. In future works, we are going to research the implementation of agricultural context-aware services using the suggested context model.

References

1. Dejene, E., Marian, S., Lionel, B.: An Ontology-Based Approach to Context Modeling and Reasoning in Pervasive Computing. In: PerComW 2007 (2007)
2. Cho, Y., Yoe, H., Kim, H.: CAS4UA: A Context-Aware Service System Based on Workflow Model for Ubiquitous Agriculture. In: Kim, T.-h., Adeli, H. (eds.) AST/UCMA/ISA/ACN 2010. LNCS, vol. 6059, pp. 572–585. Springer, Heidelberg (2010)

3. Hsien-Chou, L., Chien-Chih, T.: A RDF and OWL-Based Temporal Context Reasoning Model for Smart Home. Information Technology Journal 6(8), 1130–1138 (2008)
4. Xiao, H.W., Da, Q.Z., Tao, G., Hung, K.P.: Ontology Based Context Modeling and Reasoning using OWL. In: Second IEEE Annual Conference on Pervasive Computing and Communications Workshops (2004)
5. Chen, S., Bu, Y., Li, J., Tao, X., Lu, J.: Toward context-awareness: a workflow embedded middleware. In: Ma, J., Jin, H., Yang, L.T., Tsai, J.J.-P. (eds.) UIC 2006. LNCS, vol. 4159, pp. 766–775. Springer, Heidelberg (2006)
6. Tang, F., Guo, M., Dong, M., Li, M., Guan, H.: Towards Context-Aware Workflow Management for Ubiquitous Computing. In: Proceedings of ICESS 2008, pp. 221–228 (2008)
7. Ardissono, L., Di Leva, A., Petrone, G., Segnan, M., Sonnessa, M.: Adaptive Medical Workflow Management for a Context-Dependent Home Healthcare Assistance Service In: Proceedings of the CWS 2005, pp. 59-68 (2005)
8. Li, J., Bu, Y., Chen, S., Tao, X., Lu, J.: FollowMe: On Research of Pluggable Infrastructure for Context-Awareness. In: Proceedings of AINA 2006 vol. 1, pp. 199–204 (2006)
9. Choi, J., Cho, Y., Choi, J.: The Design of a Context-Aware Workflow Language for Supporting Multiple Workflows. Journal of Korean Society for Internet Information 11(1), 145–158 (2009); pp. 30–39 (2006)
10. Stanford University, Ptotege 4.0, http://protege.stanford.edu/
11. Matthew, H.: Protege OWL Tutorial, http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/

A Point-Based Inventive System to Prevent Free-Riding on P2P Network Environments

Jongbae Moon[1] and Yongyun Cho[2,*]

[1] Supercomputing Center, Korea Institute of Science and Technology Information, 52-11, Eoeun-dong, Yuseong-gu, Daejeon, 305-806, Korea
[2] Information and Communication Engineering, Sunchon National University, 413 Jungangno, Suncheon, Jeonnam 540-742, Korea
jbmoon@kisti.re.kr, yycho@sunchon.ac.kr

Abstract. Unlike Client-Server system, P2P network systems are based on active file sharing communications among the users. Free-Riding syndrome is a serious problem in P2P systems because free-riders, who use free resources from others but don't share their own, make the communications passive and pessimistic. Free-Riding syndrome interrupts P2P network systems from workload decentralization, causes the performance degradation, and finally undermines the foundation of P2P network systems. This paper proposes a point-based incentive system that activates file sharing communications and prevents free-riding syndrome in advance. This system can lead users to aware their status and take part in more active communications using real-time point monitor.

Keywords: P2P, distributed system, free-riding, incentive, Gnutella.

1 Introduction

P2P (Peer-to-Peer) network system is a distributed application architecture that partitions tasks or workloads between peers. Peers are equally privileged, equipotent participants in the application. Peers make a portion of their resources, such as processing power, disk storage or network bandwidth, directly available to other network participants, without the need for central coordination by servers or stable hosts. Peers are both suppliers and consumers of resources, in contrast to the traditional client–server model where only servers supply, and clients consume.

In P2P networks, clients provide resources, which may include bandwidth, storage space, and computing power. As nodes arrive and demand on the system increases, the total capacity of the system also increases. In contrast, in typical client–server architecture, clients share only their demands with the system, but not their resources. In this case, as more clients join the system, less resource are available to serve each client.

* Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 462–471, 2011.

As a result, P2P system performance is highly dependent on the amount of voluntary resource contribution from the individual nodes. However, there is a problem named Free-Riding phenomenon. Users who attempt to benefit from the resources of others without offering their own resources in exchange are termed “free-riders.” In 2000, a measurement study of the Gnutella file-sharing network found that approximately 70% of peers provide no files and that the top 1% of the peers provide approximately 37% of the total files shared. Similar patterns have been observed in subsequent studies of Napster and Gnutella networks [1]. In 2005, Hughes found free-riders have increased to 85% of all Gnutella users [2].

Free-riding phenomenon causes the following two problems. First, Free-riding reduces the performance of P2P network system. If large majority of P2P users are free-riders, none of users want to share their own files, and then P2P does not mean anything. Second, only some participants, who provide so much information and share a lot of interesting files, cause network traffic because many users try to connect them. Then, P2P network system is also same as centralized server system. Free-riding phenomenon is often occurred because P2P system could not motivate the participants such as proper rewards and punishment [3].

In this paper, we propose a point-based incentive P2P network system to promote file sharing each other participants. The proposed point-based incentive system leads the participants to voluntarily participate in sharing their files and to prevent free-riding behavior by measuring and estimating the contribution. We implemented a point-based incentive P2P network system by using Gnucleus which is open source and one of pure P2P network systems.

2 Pure P2P Network System and Related Works

In this section, we describe pure P2P network system and one of its implementation which named Gnutella.

Pure P2P network system, which is shown in Fig. 1, does not have the notion of clients or servers but only equal peer nodes that simultaneously function as both "clients" and "servers" to the other nodes on the network. This model of network arrangement differs from the client–server model where communication is usually to and from a central server. A typical example of a file transfer that does not use the P2P model is the File Transfer Protocol (FTP) service in which the client and server programs are distinct: the clients initiate the transfer, and the servers satisfy these requests. Gnutella [4] and BitTorrent [5] are the representative pure P2P network system.

In this paper, we implemented Gnutella based P2P network system. Gnutella Protocol is a distributed P2P model and is for distributed searching. Once connected, the client requests a list of working addresses. The client tries to connect to the nodes it was shipped, as well as nodes it receives from other clients, until it reaches a certain quota. It connects to only that many nodes, locally caches the addresses it has not yet tried, and discards the addresses it tried that were invalid.

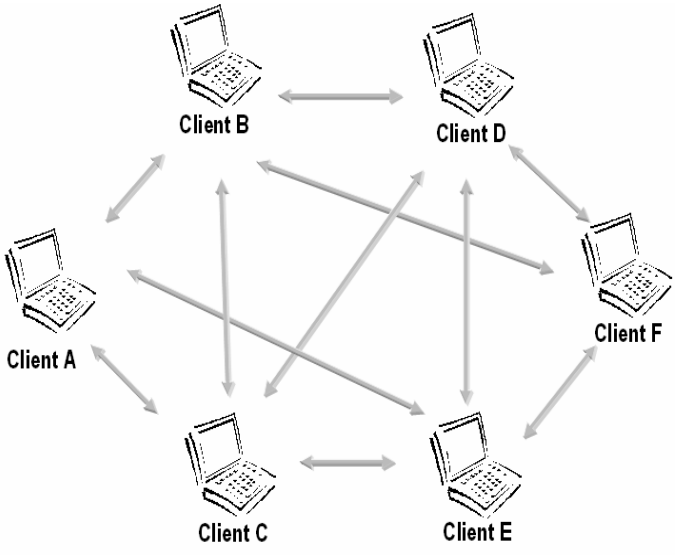


Fig. 1. Pure P2P Network System

When the user wants to do a search, the client sends the request to each actively connected node. In version 0.4 of the protocol, the number of actively connected nodes for a client was quite small (around 5), so each node then forwarded the request to all its actively connected nodes, and they in turn forwarded the request, and so on, until the packet reached a predetermined number of "hops" from the sender (maximum 7). Since version 0.6, gnutella is a composite network made of leaf nodes and ultra nodes (also called ultrapeers). The leaf nodes are connected to a small number of ultrapeers (typically 3) while each ultrapeer is connected to more than 32 other ultrapeers. With this higher outdegree, the maximum number of "hops" a query can travel was lowered to 4 [wiki].

Leaves and ultrapeers use the Query Routing Protocol to exchange a Query Routing Table (QRT), a table of 64 Ki-slots and up to 2 Mi-slots consisting of hashed keywords. A leaf node sends its QRT to each of the ultrapeers it is connected to, and ultrapeers merge the QRT of all their leaves (downsized to 128 Ki-slots) plus their own QRT (if they share files) and exchange that with their own neighbours. Query routing is then done by hashing the words of the query and seeing whether all of them match in the QRT. Ultrapeers do that check before forwarding a query to a leaf node, and also before forwarding the query to a peer ultra node provided this is the last hop the query can travel.

If a search request turns up a result, the node that has the result contacts the searcher. In the classic gnutella protocol, response messages were sent back along the route the query came through, as the query itself did not contain identifying information of the node. This scheme was later revised, so that search results now are

delivered over User Datagram Protocol (UDP) directly to the node that initiated the search, usually an ultrapeer of the node. Thus, in the current protocol, the queries carry the IP address and port number of either node. This lowers the amount of traffic routed through the gnutella network, making it significantly more scalable

There are some incentive P2P systems that have central server to manage query traffic. However, most of the central server P2P system is weak in SPOF (Single Point of Failure) that the system could not provide service when the central server is down.

Maze's centralized architecture means all control and query traffic is logged and available to us [10]. Maze uses a simple incentive system where user points increase with uploads and decrease with downloads. The central server audits file transfers and adjust user points accordingly. However, nasty users can avoid the incentive system by logging in another user account.

In XREP Protocol [11] using polling algorithm, a participant requests resources to an appropriate user who has the most reputation by sending a query to many participants. This protocol guarantees quality of the sharing information but does not consider Free-Riding phenomenon.

3 Proposed Point-Based Incentive P2P Network System

In this section, we propose a point-based incentive P2P network system shown in Fig. 2. The system consists of 4 components which are Gnutella protocol, Point Policy-based Incentive module, P2P File Sharing module, and File Sharing Management Monitoring module. The system provides easy way to know users' point and to prevent Free-Riding behavior by leading users to aware their status and take part in more active communications using real-time point monitor.

In this paper, we propose a point based incentive policies according to the following five elements to prevent free-riding and encourage cooperation between nodes on the Gnutella based P2P system.

The first element is file transfer such as uploading and downloading files. File sharing is a basic element of P2P network system. The participants who provide more files get more incentive, but the other participants who do not want to share their files get less incentive. The second element is bandwidth for file upload. In P2P network systems, storage space and bandwidth are the most constrain elements. The participants who guarantee a certain level of upload bandwidth get more incentive according to the bandwidth. The third element is query or query reply. The user sends query packet to neighbor nodes to request and reply. The query send or receive is a payload descriptor that means Query and QueryHits defined in the Gnutella protocol. The fourth element is a number of file sharing. In Napster system, all the files are stored at upload directory, and all the files must be shared by default for the other participants. However, gnutella users can easily move the files to non shared directory. This will be considered no intention of sharing files. The last element is elapsed time after connecting the P2P system. In Maze P2P network system, regular users keep connecting

about 2.89 hours, but Free-Riders only keep connecting about 1.15 hours. In this paper, we provide more incentive when the participants keep connecting in the P2P network system and prevent Free-Riding behavior.

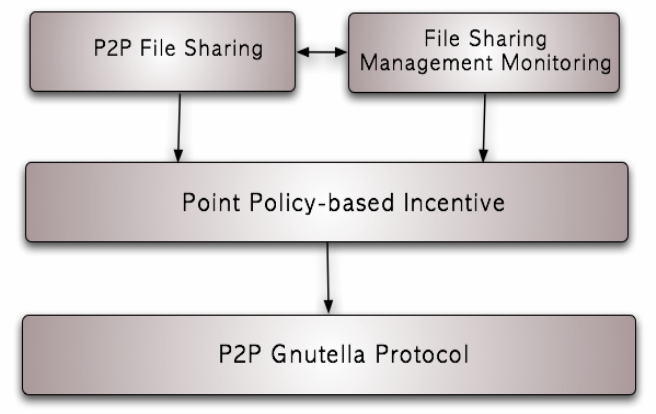


Fig. 2. Proposed Point based Incentive P2P Network System

We assigned a different weight to the five elements mentioned above, and this is shown in Fig. 3.

File Transfer 40%	Query 20%	Bandwidth 20%	File Count 10%	Time 10%

Fig. 3. Point rate of the five elements

We defined a point that a user transfers data amount of one megabyte. A participant receives initial points as default incentive when the user connects the P2P system. This initial incentive point plays an important role to remove repulsion of the users.

As shown in Fig.4, regular users get the initial point (1,000P) at the first time to connect the P2P system, and the incentive will increase or decrease according to the use of resources such as storage, bandwidth, file sharing, and keeping time. At first, when node 1 sends a query that search AAA.ZIP file to node 2 and 3. Node 2, which received the query and has the AAA.ZIP file, sends a QueryHits to node 1. Node 3, which received the query from node 1 and do not have the AAA.ZIP file, forwards the query to node 4. Node 1 with requesting a query for searching a file decrease points, node 2 and node 4 gets points. Node 1 receives information about the same data file through node 2 and node 4's QueryHits.

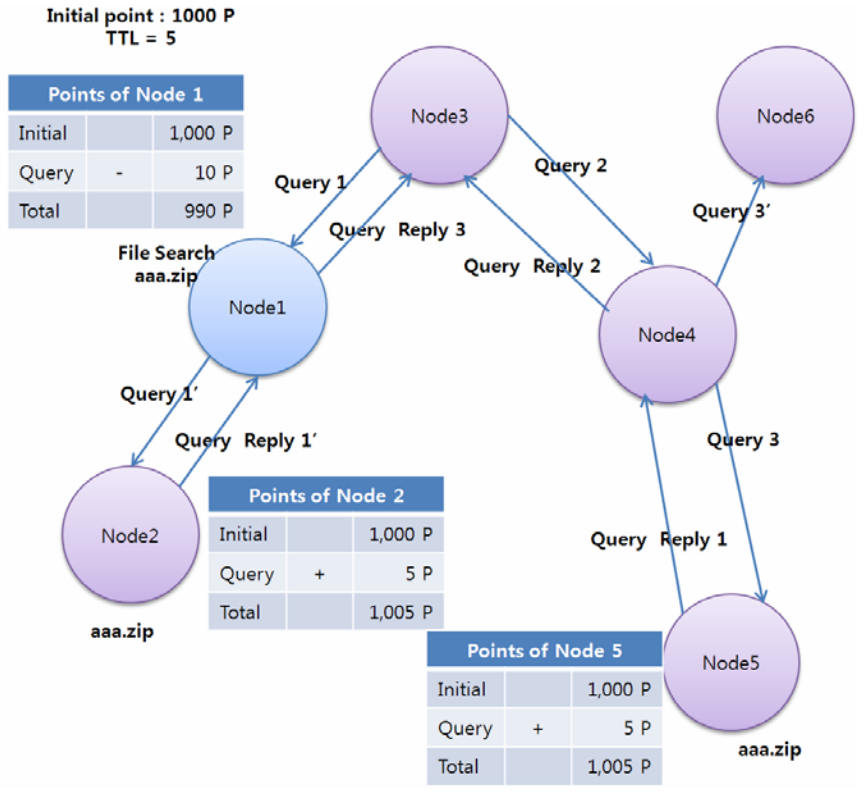


Fig. 4. Points when participants request query and reply

Table 1. Notations

Discriptor	Description
Qp_i	Points for Peer i's Query Sending and Receiving
Fp_i	Points for Peer i's File Transferring
Bp_i	Points for Peer i's Upload Bandwidth
FOp_i	Points for Peer i's Number of Files
TOp_i	Points for Peer i's Login Session Time
Tp_i	Total Points for Peer i
$\sum_{i=1}^{m} T_i$	Total Points for Peer i until occurring event "m"
$\alpha, \beta, \gamma, \delta, \epsilon$	Weight Value

$$Qp_i = \sum_{i=1}^{m} T_i' + Qr \times \alpha$$

$$Qp_i = \sum_{i=1}^{m} T_i' - Qs \times \alpha$$

(Qr, Qs mean Query receiving and Query sending.)

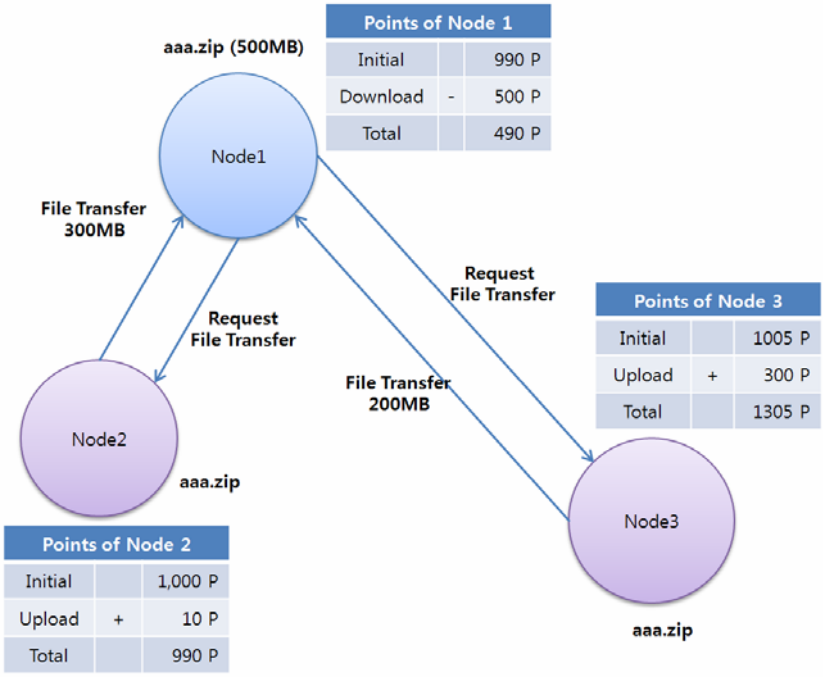


Fig. 5. Points when participants upload or download a file

As shown in Fig. 5, node 1 now knows the participants' location and information about the file after requesting and receiving the query and queryhits. Node 1 sends a request to transfer the AAA.ZIP file. Node 2 and node 4 get points amount of (total upload * β). If the size file of AAA.ZIP is 500MB, and node 1 downloads 500MB, node 1 decrease its point amount of 500MB * β. And if node 2 uploads 300MB and node 4 uploads 200, node 2 gets points amount of 300 * β and node 4 gets points amount of 200 * β.

$$Fp_i = \sum_{i=1}^{m} T_i + Fs \times \beta$$

$$Fp_i = \sum_{i=1}^{m} T_i - Fr \times \beta$$

(Fs, Fr mean file uploading and downloading)

When a participant increases the upload bandwidth, the user gets more points. It is important element to keep the upload bandwidth. It is sure that the file sharing client can control the bandwidth. Most of users using e-Donkey system are keeping the bandwidth less than 10KB/s. Therefore, participants have to keep the bandwidth like the following in the proposed system.

$$Bp_i = \sum_{i=1}^{m} T_i + 1h \times \gamma$$

(1h means 1 hour.)

When a participant has more shared files, the user gets more points.

$$FCp_i = \sum_{i=1}^{m} T_i + 1h \times FC \times \delta$$

(FCi means the number of sharing files during 1 hour.)

When a participant keeps connection longer in the system, the user gets more points.

$$TCp_i = \sum_{i=1}^{m} T_i + TC \times \epsilon$$

(TC means connection time per 1 minute.)

The total amount of points is like the following.

$$Tp_i = \sum_{i=1}^{m} (\alpha Qp_i + \beta Fp_i + \gamma Bp_i + \delta FCp_i + \epsilon TCp_i)$$

(α =20%, β =40%, γ=20%, δ=10%, ϵ = 10%)

In the proposed incentive system, when a user connects the system, the system checks the point first. And if the total point of Tp_i is less than ($\sum_{i=1}^{m} T_i / 2$ or $\sum_{i=1}^{m} T_i / 4$ or $\sum_{i=1}^{m} T_i / 8$), the user is considered as a Free-Rider. Then the system executes an appropriate Free-Riding policy to limit a function of the user. The restriction of the Free-Rider is the following.

1) If $Tp_i \leqq \sum_{i=1}^{m} T_i / 2$ then the system pops up a warning message that the user has lower point for getting shared files from other participants.

2) If $Tp_i \leqq \sum_{i=1}^{m} T_i / 4$ then the system reduces TTL value by 1 to limit the query.

3) If $Tp_i \leqq \sum_{i=1}^{m} T_i / 8$ then the system reduces TTL value by one again to limit the query.

4 Implementation

We implemented the incentive P2P system based on Gnucleus [13] which is an open source and one of pure P2P systems. We do not modify the core modules of the protocol to keep other clients compatible and use search and upload/download modules provided in the Gnucleus. We added some modules for monitoring and controlling users' points and Free-Riding behavior.

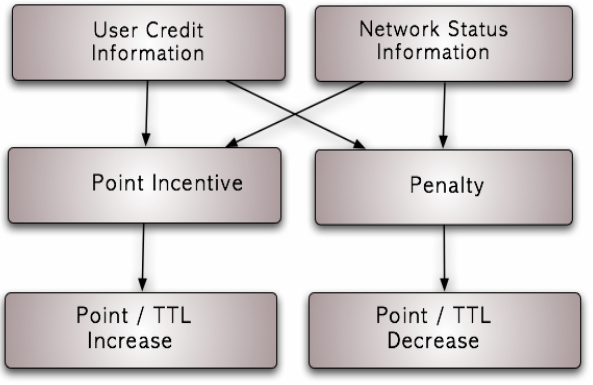


Fig. 6. Modules in the proposed incentive system

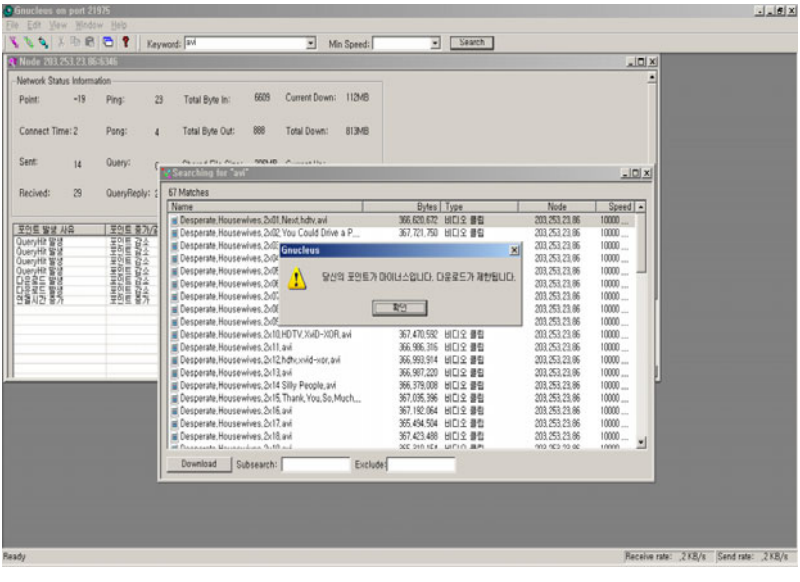


Fig. 7. The implemented system based on Gnucleus

As shown in Fig. 6, we implement a user credit information module which is for monitoring user's actions and logging the data, a network status information module which is for monitoring network status and bandwidth, and point based incentive module which are consist of point incentive, penalty, TTL increase, TTL decrease module. And the implemented system is shown in Fig. 7.

5 Conclusions and Future Works

In this paper, we propose a point based incentive P2P network system to prevent free riding in Gnutella based P2P network systems. We use five elements for calculating participants' point: File Transfer, Query, Bandwidth, File Count, and Elapsed Time. We determine a criterion for Free-Riding value and define restriction policies for the Free-Riders. We implemented a point-based incentive P2P network system which is based on Gnucleus. The system considers various elements to lead users to aware their status and take part in more active communications using real-time point monitor.

For the future works, we need to use more elements for calculating users' point and need more studies about restriction policy. We also need more methods for keeping popular files and keeping high bandwidth for uploading files.

References

1. Saroiu, S., Gummadi, P.K., Gribble, S.D.: A Measurement Study of Peer-to-Peer File Sharing Systems. In: Proceedings of Multimedia Computing and Networking (MMCN 2002) (2002)
2. Hughes, D., Coulson, G., Walkerdine, J.: Free Riding on Gnutella Revisited: the Bell Tolls? Published in IEEE Distributed Systems Online 6(6) (2005)
3. Jun, S., Ahamad, M.: Incentives in BitTorrent Induce Free Riding. Applications, Technologies, Architectures, and Protocols for Computer Communication, 116–121 (2005)
4. Ripeanu, M.: Peer-to-Peer Architecture Case Study: Gnutella Network. In: First International Conference on Peer-to-Peer Computing (2001)
5. Pouwelse, J.A., Garbacki, P., Epema, D.H.J., Sips, H.J.: A Measurement Study of the BitTorrent Peer-to-Peer File-Sharing System. Technicla Report PDS-2004-007, Delft University of Technology (2004)
6. http://en.wikipedia.org/wiki/Gnutella
7. Yang, M., Zhang, Z., Li, X., Dai, Y.: An Empirical Study of Free-Riding Behavior in the Maze P2P File-Sharing System. In: van Renesse, R. (ed.) IPTPS 2005. LNCS, vol. 3640, pp. 182–192. Springer, Heidelberg (2005)
8. Damiani, E., Paraboschi, S., Samarati, P., Violante, F.: A Reputation-based Approach for Choosing Reliable Resources in Peer-to-Peer Networks. In: Proceedings of the 9th ACM conference on Computer and communications security, pp. 207–216 (2002)

A Probability Density Function for Energy-Balanced Lifetime-Enhancing Node Deployment in WSN

Subir Halder[1,2], Amrita Ghosal[1,2], Amartya Chaudhuri[2], and Sipra DasBit[2]

[1] Dept. of Comp. Sc. & Engg, Dr. B. C. Roy Engineering College, Durgapur, India
[2] Dept. of Comp. Sc. & Tech., Bengal Engineering and Science University, Shibpur, India
subir_ece@rediffmail.com, ghosal_amrita@yahoo.com,
{amartyaamp,siprad}@hotmail.com

Abstract. Energy is one of the scarcest resources in wireless sensor network (WSN). So the issue of preserving energy requires utmost attention. There are many ways to conserve energy in such a network. One primary way of conserving energy is judicious deployment of sensor nodes within the network area so that the energy flow remains balanced throughout the network. This prevents the problem of occurrence of 'energy holes' and ensures prolonged network lifetime. This work first proposes a probability density function (PDF) and derives its intrinsic characteristics. We have shown the PDF's suitability to model the network architecture considered for the work. Next a node deployment algorithm is developed based on this PDF. Further, we have identified necessary constraints involving different network parameters for ensuring energy balance of the entire network. Performance of the deployment scheme is evaluated in terms of energy balance and network lifetime. Finally the scheme is compared with two existing deployment schemes. Simulation results confirm our scheme's supremacy over the two existing schemes in terms of all the two performance metrics.

Keywords: Probability density function, Coverage, Connectivity, Energy balance, Network lifetime.

1 Introduction

A wireless sensor network (WSN) [1] consists of several hundreds of sensor nodes which collect data from their surroundings and send the collected data to their neighbouring nodes in single hop. The neighbouring nodes in turn send the data to the sink either directly or via their one hop neighbouing nodes. The sink processes and transmits the received data to the outside world. Sensor nodes are equipped with battery whose charge cannot be replaced easily after deployment and so the need to conserve energy is a major concern of WSN. The rate of energy depletion in the network primarily depends on the deployment nature of the nodes that further depends on the application environment.

Deployment can be random or pre-determined. In random deployment, nodes are randomly deployed generally in an inaccessible terrain. For example, in the application

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 472–487, 2011.

domain of disaster recovery or in forest fire detection, sensors are generally dropped by helicopter in random manner [2]. In pre-determined deployment, number of nodes in a unit area is known apriori and is used in applications where sensors are expensive or their operation is significantly affected by their positions. These applications include placing imaging and video sensors, populating an area with highly precise seismic nodes, monitoring manufacturing plants etc [2].

One important way of conserving energy is through uniform energy or load distribution all over the network. Non-uniform energy dissipation in any part of the network may result in non-functioning of that part leading to the phenomenon of energy hole problem [3] that effects the network lifitime. The non-uniform energy dissipation arises due to uneven data transmissions by certain nodes in the network resulting in extra energy dissipation of those nodes. This problem also causes a substantial amount of energy to remain in the nodes even after network lifetime ends leading to significant wastage of energy [4]. To avoid this, nodes should be deployed in such a manner that the energy dissipation of all nodes takes place uniformly ensuring load balancing throughout the network. A good sensor deployment strategy is one that achieves both energy balance and energy efficiency [5].

Many works reported so far deal with the deployment issue. In [6], [7] authors have proposed a deployment strategy with a target to cover the area of interest. In [8] authors have proposed a deployment scheme to minimize energy consumption in the whole network so that the network lifetime is prolonged. In [9], [10] authors have proposed the deployment scheme for efficient energy usage throughout the network, thereby enhancing the network lifetime. In [11], [12] authors have used standard distribution functions for node deployment showing their capabilities for enhancing the network lifetime. Most of the distribution-function/scheme based deployment strategies have not addressed all the issues of energy balance and network lifetime simultaneously. This motivates us to propose a probability density function based on which the pre-determined node deployment strategy is proposed. The proposed scheme is targeted to achieve energy balance and enhancement of network lifetime.

The rest of the paper is organized as follows. In section 2, literature review is elaborated. The network model considered for the present work is presented in section 3. Section 4 presents the proposed node deployment scheme along with the proposed probability density function based on which the scheme is developed. In section 5, the performance of the scheme is evaluated based on both qualitative and quantitative analysis. Finally in section 6, the paper is concluded with some mention about the future scope of the present work.

2 Related Work

Y. Zou *et al.* [6] have formulated a problem on uncertainty-aware sensor node placement when nodes are dropped from airplanes. They have proposed two pre-determined node placement algorithms- maximum miss probability and minimum

miss probability for determining minimum number of required nodes and their locations (mean positions) such that coverage is ensured. However, no attempt has been made for prolonging the network lifetime and balancing energy consumption in the network.

P. K. Agarwal *et al.* in [7] have also proposed a node placement algorithm which requires minimum number of nodes to cover a region. The approach is landmark based where landmarks are the set of finite points in a 2-D space. The algorithm is proposed based on greedy approach to compute the location of sensors. Although authors guarantee coverage of a given region but fail to ensure connectivity amongst the sensors. They are also silent about network lifetime.

D. Ganesan *et al.* [8] have formulated an optimization problem for node placement and transmission structure of data gathering to minimize communication energy. The node placement strategy is first studied in 1-D network and is then extended to 2-D circular network. The algorithm in 2-D network considers the circular area partitioned into wheel like structures where wheels are comprised of a number of spokes. But, the authors have not mentioned the energy balancing criterion and network lifetime of the proposed node placement algorithm.

Wu *et al.* [9] have explored the theoretical aspects of energy hole problem in sensor networks in layered architecture. They have proposed a non-uniform node distribution strategy which ensures maximum energy efficiency in the network. The number of nodes distributed in a layer is determined based on the minimum number of nodes required in the upper adjacent layer. However, the authors are silent about the minimum number of nodes required to be placed in the farthest layer from the sink to maintain connectivity and coverage.

C. Y. Chang *et al.* [10] have proposed two node deployment schemes- distance-based and density-based for balancing power consumption among the sensor nodes. In distance-based scheme, deployment positions of nodes are adjusted such that the nodes' neighbors towards sink are located relatively closer compared to other neighbor nodes. The density-based scheme partitions the network into a number of equal-sized zones, adjusts the density of nodes in each zone by controlling the switching mode as on/off and balances the load of each zone. However, the scheme requires various control mechanisms that are difficult to implement in resource-constrained WSN.

Olariu *et al.* [11] have given a network design guideline for maximizing lifetime and avoiding energy hole with uniform node distribution. They show that uneven energy depletion due to energy hole is unavoidable for free-space model, but can be prevented for multipath model. The authors have provided the design guideline for multipath model with corona architecture in terms of widths and the number of layers. However, they have not explored the potential of non-uniform node deployment.

D. Wang *et al.* [12] propose an analytical model for coverage and network lifetime of sensor networks using two dimensional Gaussian distribution showing more nodes get deployed closer to the sink using this model. They have proposed two deployment algorithms- one for circular and another for elliptical network area using which larger coverage and longer network lifetime are achieved. But the deployment scheme does not ensure energy balancing in the network.

3 Network Model

In this section, we describe the network architecture along with sensing, communication and energy models considered for this work.

3.1 Architecture

We consider a square shaped network area a×a which is covered by a set of annuli. Each such annuli is designated with width r as layer. The sink is considered to be located at the centre of the network area. Nodes are placed in different layers surrounding a single sink. A layer is identified as L_i where i = 1, 2, . . . , N. Here i=1 indicates the layer nearest to the sink and i=N indicates the layer farthest from the sink.

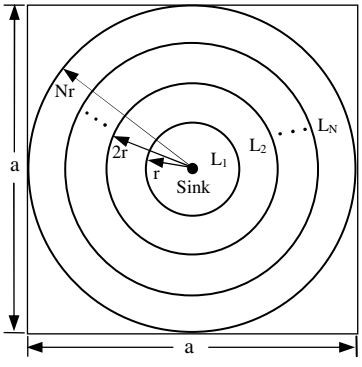


Fig. 1. Layered network area

We assume all the sensor nodes are homogeneous with respect to their initial energy, sensing and communication ranges. The nodes are static and distributed within the network with a given node density. Node density is defined [3] as the ratio of the number of nodes in a layer and the area of the layer. Further we consider that a unit area generates data at ρ bits/sec which is collected by the nodes and sent to the sink after a fixed time-interval q(t). We assume that data is transmitted from layers towards sink following the greedy forwarding routing policy [13] where a node in a layer chooses a neighbor as next-hop when the neighbor is located closest to the destination in comparison with the other neighbors. As a result of this greedy approach, during data transmission from a layer to the sink the data traverses each of the layers only once along the transmission path.

3.2 Sensing and Communication Model

3.2.1 Communication Model

We define a network as connected, if any active node can communicate with the sink node either in single hop or in multiple hops. We assume two nodes can directly

exchange messages if their Euclidean distance is not larger than the communication range R_c. The relationship between r (width of a layer) and R_c must satisfy the condition $r \le R_c$ [12] for ensuring connectivity in the network area (Figure 1).

Lemma 1: For a given network area a×a, in order to maintain connectivity of the network, the number of layers (N) stands in relation with R_c as $N \ge \frac{a}{2R_c}$.

Proof: If the radius of each layer in the layered architecture is r, then the distance between the centre of the inner most layer and the farthest edge of any other layer is given by ir (Refer figure 1), where i is the layer number. If the distance between the centre of the inner most layer and the farthest edge of any other layer in the network area is $\frac{a}{2}$, then replacing i by N, we get $Nr \ge \frac{a}{2}$ *or,* $N \ge \frac{a}{2r}$.

Replacing $r \le R_c$ in the above relation, we have $N \ge \frac{a}{2R_c}$.

3.2.2 Sensing Model

We define a unit area to be covered if every point in that area is within the sensing range of at least one active node. The nodes perform observation [14] at an angle of 360°. The maximal circular area centered around a node v, that can be covered by the node is defined as its sensing area S(v). The radius of S(v) is called the v's sensing range [14] R_s. We assume the relationship between r and R_s must satisfy the condition $r \le 2R_s$ [12] for covering the network area (Figure 1) under consideration.

Corollary 1: For a given network area a×a, in order to maintain network coverage, the number of layers (N) stands in relation with R_s as $N \ge \frac{a}{4R_s}$.

Proof: From lemma 1, the relationship between a and N is evaluated as, $N \ge \frac{a}{2r}$. Replacing $r \le 2R_s$ in the relation $N \ge \frac{a}{2r}$, we have $N \ge \frac{a}{4R_s}$.

3.3 Energy Model

We have considered the first order radio model [12] as our energy model where energy consumption of a node is dominated by its wireless transmissions and receptions; so the other energy consumption factors such as for sensing and processing are neglected. According to this radio model, energy consumed by a node for transmission and reception are as follows:

Energy consumption for transmitting (e_{tx}) n-bit data over a distance d is

$$e_{tx}(n, d) = e_{elec}\, n + e_{amp}\, n\, d^2 \tag{1a}$$

Energy consumption for receiving (e_{rx}) n-bit data is

$$e_{rx}(n) = e_{elec}\, n \tag{1b}$$

4 Probability Density Function Based Node Deployment (PDFND)

In this section the node deployment strategy based on probability density function and the implementing algorithm is presented.

4.1 Proposed Probability Density Function (PDF)

The mathematical domain under consideration is divided into a number of concentric circles centered at (0, 0) and having radii increasing arithmetically from R to NR with a difference of R. In the mathematical domain, if (x, y) be a point lying between circle-(i-1) and circle-i, then the probability density at that point is

$$f(x,y;N,i,R)=\frac{k(2i-1)}{N^2 i^4},\ \forall\,(i-1)^2R^2<x^2+y^2\le i^2R^2 \tag{2a}$$

where $i=1,2,\ldots,N$ and k is a constant given by $k=\dfrac{N^2}{\pi R^2\left[1+\dfrac{3^2}{2^4}+\dfrac{5^2}{3^4}+\cdots+\dfrac{(2N-1)^2}{N^4}\right]}$.

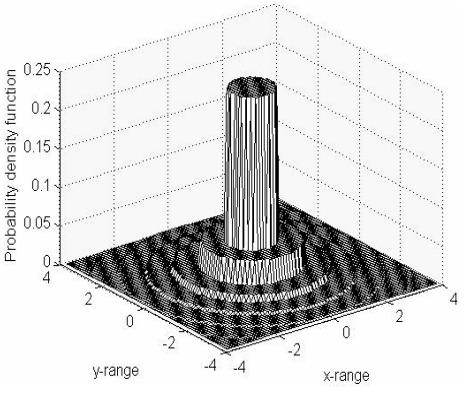


Fig. 2. Surface plot of the PDF

Figure 2 is the 3-D graph of the proposed PDF. The characteristics of the PDF show fall in the functional value with rise in the value of i implying lower probability and vice versa.

Theorem 1: The value of constant k is given by $k=\dfrac{N^2}{\pi R^2\left[1+\dfrac{3^2}{2^4}+\dfrac{5^2}{3^4}+\cdots+\dfrac{(2N-1)^2}{N^4}\right]}$.

Proof: Let p_i denotes the probability at point (x, y) and the point lies between circles (i-1) and i. From the proposed PDF, the probability p_i is

$$p_i=\frac{k(2i-1)}{N^2 i^4}\iint f(x,y)\,dx\,dy.$$

In the above relation $\iint f(x,y)\,dx\,dy$ is the domain area. For circular domain, area is $\iint f(x,y)\,dx\,dy = (2i-1)\pi R^2$. The probability of x and y is given as $p_i = \frac{k(2i-1)^2\pi R^2}{N^2 i^4}$. By fundamental rule of probability, $\sum\sum f(x,y) = 1$

$$\sum_{i=1}^{N} p_i = 1 \; or, \; \sum_{i=1}^{N} \frac{k(2i-1)^2\pi R^2}{N^2 i^4} = 1$$

$$k = \frac{N^2}{\pi R^2 \sum_{i=1}^{N} \frac{(2i-1)^2}{i^4}} = \frac{N^2}{\pi R^2\left[1+\frac{3^2}{2^4}+\frac{5^2}{3^4}+\cdots+\frac{(2N-1)^2}{N^4}\right]}.$$

Theorem 2: If two random variables X and Y follow a proposed PDF with parameters N and i, then the cumulative distribution function (CDF) of X and Y is given as

$$F[X \le x, Y \le y] = \frac{k\pi R^2}{N^2}\left[\sum_{j=1}^{i}\left[\frac{(2j-1)^2}{j^4}\right] + \frac{\left(\eta^2 - i^2\right)}{i^4}\right]$$

where (x, y) such that $0 \le x^2 + y^2 \le \eta^2 R^2$, where $i \le \eta \le i+1$.

Proof: The probability of two discrete random variables X and Y for a particular value within a given range of i is considered as

$$\frac{k\pi R^2}{N^2}\sum_{j=1}^{i}\frac{(2j-1)^2}{j^4}. \tag{2b}$$

The probability of the variables X and Y between a given domain area iR and ηR, where $\eta R > iR$ is given as

$$\frac{k}{N^2 i^4}\left[\pi(\eta R)^2 - \pi(iR)^2\right]. \tag{2c}$$

So, the CDF of X and Y using equations (2b) and (2c) is obtained as

$$F[X \le x, Y \le y] = \frac{k\pi R^2}{N^2}\left[\sum_{j=1}^{i}\left[\frac{(2j-1)^2}{j^4}\right] + \frac{\left(\eta^2 - i^2\right)}{i^4}\right].$$

Theorem 3: If the two random variables X and Y follow the proposed PDF with parameters N and i, then the expectation of X and Y is given as

$$E[XY] = \frac{kR^4}{N^2}\sum_{i=1}^{N}\left[\frac{2}{i} + \frac{2}{i^3} - \frac{3}{i^2} - \frac{1}{2i^4}\right].$$

Theorem 4: If the two random variables X and Y follow a proposed PDF with parameter N and i, then the covariance of X and Y is given as

$$\mathrm{Cov}(X,Y)=\frac{kR^4}{N^2}\sum_{i=1}^{N}\left[\frac{2}{i}+\frac{2}{i^3}-\frac{3}{i^2}-\frac{1}{2i^4}\right]-\frac{4kR^3}{3N^2}\left[\sum_{i=1}^{N}\left[\frac{3}{i^2}+\frac{3}{i^3}-\frac{1}{i^4}\right]\right]^2.$$

Due to page limitation, the proofs of Theorems 3 and 4 could not be incorporated.

4.2 Proposed PDF-Based Deployment

The probability density function (PDF) proposed in the previous sub-section is discrete in nature. Our objective is to deploy sensor nodes in the layered network area (Figure 1) with the proposed PDF. The PDF is mapped with the node deployment in a layered network area as follows: the parameter i corresponds to layer number where i=1,2,…,N. Here i=1 indicates the layer nearest to the sink and i=N indicates the layer farthest from the sink. The parameter R corresponds to the width r of the annuli/layer. The density function is designed as a non-uniform one such that higher value of PDF implies a node deployed around the sink and the lower value of PDF is observed as one moves away from the sink. The probability density of deploying a node at point (x, y) for a value of i is

$$\frac{k(2i-1)}{N^2i^4}, \text{ where } i=1,2,\ldots,N. \tag{3a}$$

In equation (3a), k is constant and i is the number of layers in the network. The probability density for nodes deployed within layer-i is

$$\frac{k(2i\text{-}1)A_i}{N^2i^4}$$

where A_i is the area of layer-i and

$$k=\frac{N^2}{\pi r^2\left[1+\frac{3^2}{2^4}+\frac{5^2}{3^4}+\cdots+\frac{(2N-1)^2}{N^4}\right]}$$

where r is the width of a layer. The area of layer-i is $A_i=\left[(2i-1)\pi r^2\right]$. The probability of deploying nodes at layer-i is $p_i=\frac{k(2i\text{-}1)A_i}{N^2i^4}$. Replacing the value of A_i, the probability of deployment of nodes at layer-i is

$$p_i=\frac{k(2i-1)^2\pi r^2}{N^2i^4}. \tag{3b}$$

The number of nodes in layer-i (T_i) is equal to the probability of deploying nodes at layer-i (p_i) multiplied by the total number of nodes (T_{total}) that are to be deployed within the network area i.e.,

$$T_i=p_i\times T_{total}. \tag{3c}$$

The probability expression implies that nodes in a layer are uniformly distributed with equal probability, but this probability varies in different layers.

4.3 Algorithm for Node Deployment

1: input a, R(r), T_{total} /* area parameter, width of layer, and total number of nodes to be deployed */

/* assume r= R_c ; section 3.2.1 */

2: compute $N = \frac{a}{2 \times r}$ /* N: no. of layers; section 3.2*/

3: compute k /* k: constant; section 4.2 */

4: **for** i=1; $i \leq N$; i + +

5: compute p_i /* p_i : probability of deploying nodes at layer-i; equation 3(b) */

6: compute $T_i = p_i \times T_{total}$ /* T_i: no. of nodes to be deployed in layer i */

7: **end for**

4.4 Illustrative Example

Let us consider a square area of 200×200 sq unit where 100 nodes are deployed employing the PDF. The area is covered by layers with width (r) of 25 units. The number of layers is $N = \frac{a}{2 \times r} = \frac{200}{2 \times 25} = 4$ (section 3.2). Replacing the values of N and R (r) in equation (2a), we get k=0.004. The probability of node deployment at layer-i is computed using equation (3b), as $p_i = \frac{0.49 \times (2i-1)^2}{i^4}$. Probability of node deployment at layer-1, $p_1 = 0.49$. Probability of node deployment at layer-2, $p_2 = 0.27$. Similarly $p_3 = 0.15$, $p_4 = 0.09$. Using equation (3c), the number of nodes deployed in the 4 layers is: in layer-1 $T_1 = 0.49 \times 100 = 49$, in layer-2 $T_2 = 27$, in layer-3 $T_3 = 15$ and in layer-4 $T_4 = 9$.

We observe that the number of nodes deployed in each layer conforms to the non-uniform nature of the PDF. Therefore, it fulfils our objective of deploying more nodes towards sink and lesser nodes at locations away from the sink.

5 Performance Analysis

Performance of the present node deployment strategy is measured based on two parameters such as energy balance and network lifetime. Both qualitative and quantitative analysis is presented here.

5.1 Qualitative Analysis

In this sub-section, parameters involved in maintaining energy balance and enhancing network lifetime are identified.

5.1.1 Energy Balancing

Our objective is balancing energy consumption among all the network layers so that network lifetime is maximized. A WSN with layered architecture is said to be energy balanced when all nodes of the network use up their energy at the same time [9].

Nodes of all the layers except those belonging to the farthest layer from the sink, spend their energy for transmitting their own data, receiving data from nodes of adjacent layers farther from the sink and forwarding the received data. Nodes of the farthest layer from the sink spend their energy only for transmitting their own data.

Energy required for a node to transmit (e_{tx}) n-bit of data over the distance R_c is- $e_{tx}(n, R_c) = e_{elec}\, n + e_{amp}\, n\, R_c^2 = e_t n$, where $e_t = e_{elec} + e_{amp}\, R_c^2$ [using equation (1a)].

Similarly, energy required for a node to receive (e_{rx}) n-bit data is- $e_{rx}(n) = e_{elec}\, n = e_r n$, where $e_r = e_{elec}$ [using equation (1b)].

Let ECR_i denotes energy consumption rate i.e., energy consumption per unit time-interval q(t) (section 3.1) of layer-i. As the last layer consumes energy for transmitting its own sensed data, for i=N,

$$ECR_N = e_t \times \rho \times A_N \tag{4a}$$

where $\rho \times A_N$ is the number of bits transmitted for the layer N per second. As the rest of the layers consume energy both for transmitting its data and for receiving and forwarding the other outer layers data, for i=1,2,…,(N-1)

$$ECR_i = e_t \times \rho \times A_i + \left[(e_t + e_r) \sum_{j=i+1}^{N} (\rho \times A_j) \right] \tag{4b}$$

where $e_t \times \rho \times A_i$ for transmitting its own data and $(e_t + e_r) \sum_{j=i+1}^{N} (\rho \times A_j)$ for receiving & forwarding the other outer layers data.

All the nodes of the network use up their energy at the same time [9], means that the ratio of total initial energy content of a layer and ECR of that layer is same for all layers in the network. So for energy balancing, the following condition must be satisfied-

$$\frac{T_1 \times E_{Initial}}{ECR_1} = \frac{T_2 \times E_{Initial}}{ECR_2} = \cdots = \frac{T_i \times E_{Initial}}{ECR_i} \tag{4c}$$

where T_i is the number of nodes in layer-i and $E_{Initial}$ is the initial energy in each node.

From equation (4c), the condition required for balancing energy throughout the network is-

$$\frac{ECR_i}{T_i} = \frac{ECR_{i+1}}{T_{i+1}}$$

$$\frac{T_i}{T_{i+1}} = \frac{e_t \times \rho \times A_i + \left[(e_t + e_r) \sum_{j=i+1}^{N} (\rho \times A_j) \right]}{e_t \times \rho \times A_{i+1} + \left[(e_t + e_r) \sum_{j=i+2}^{N} (\rho \times A_j) \right]}$$ [using equation (4b)].

Simplifying the above relation by dropping ρ and replacing A_i by $(2i-1)\pi r^2$ in sequence results in-

$$\frac{T_i}{T_{i+1}} = \frac{e_t \times (2i-1)\pi r^2 + \left[(e_t + e_r) \sum_{j=i+1}^{N} (2j-1)\pi r^2 \right]}{e_t \times (2i+1)\pi r^2 + \left[(e_t + e_r) \sum_{j=i+2}^{N} (2j-1)\pi r^2 \right]}$$

$$\frac{T_i}{T_{i+1}} = \frac{e_t \times (2i-1) + (e_t + e_r)\sum_{j=i+1}^{N}(2j-1)}{e_t \times (2i+1) + (e_t + e_r)\sum_{j=i+2}^{N}(2j-1)}. \tag{4d}$$

From the proposed PDF (section 4 & Appendix) LHS of equation (4d) is evaluated as

$$\frac{T_i}{T_{i+1}} = \frac{(2i-1)^2 (i+1)^4}{(2i+1)^2 i^4}.$$

The RHS of equation (4d)

$$\frac{e_t \times (2i-1) + (e_t + e_r)\sum_{j=i+1}^{N}(2j-1)}{e_t \times (2i+1) + (e_t + e_r)\sum_{j=i+2}^{N}(2j-1)}.$$

Both the LHS and RHS are ratios where denominator and numerator have same power of i and therefore, these two terms are approximately equal. So if the nodes are deployed employing the proposed PDF, it fulfils the objective of energy balancing.

5.1.2 Network Lifetime

Network lifetime is defined in terms of network coverage. It is the time till the proportion of dead nodes exceeds a certain threshold, which may result in loss of coverage of a certain region, and/or network partitioning [4]. Energy consumption rate by the nodes of a layer can be calculated using equations (4a) and (4b). Energy consumption per unit q(t) by each node in layer-i (ER_i) is given as

$$ER_i = \frac{ECR_i}{T_i \times q(t)} \quad \text{for i=1, 2, …, N [from equation (4c)].}$$

As our scheme is energy balanced, the lifetime of a node is same as lifetime of a layer or network lifetime. The lifetime of each node in layer-i is

$$LT_i = \frac{E_{Initial}}{ER_i}.$$

Putting the value of ER_i in the above relation, we have

$$LT_i = \frac{E_{Initial} \times T_i}{ECR_i} q(t)$$

$$LT_i = \frac{E_{Initial} \times T_i}{e_t \times \rho \times A_i + \left[(e_t + e_r)\sum_{j=i+1}^{N}(\rho \times A_j)\right]} q(t). \tag{5}$$

The parameters- $E_{Initial}$, T_i, q(t), e_t, e_r, ρ, A_i affect the lifetime of a node or layer. As the values of e_t, e_r, $E_{Initial}$, A_i are constant, we have concentrated on rest of the three parameters as mentioned below.

5.1.2.1 Number of Nodes in Each Layer (T_i). From equation (5) it is observed that with the increase in number of nodes in each layer, keeping the other parameters fixed, the network lifetime increases. Therefore, network lifetime is directly proportional to the number of nodes in each layer.

5.1.2.2 Interval of Periodic Data Collection $(q(t))$. From equation (5) it is inferred that as the interval of periodic data collection rate increases, the lifetime of the network also increases keeping the other parameters unchanged. Increase of interval of periodic data collection refers to less data collection, thereby resulting in reduced energy consumption.

5.1.2.3 Information Generation rate(ρ). In equation (5) lifetime of the network is inversely proportional to the information generation rate which means that with increase in information generation rate, node has to sense more data. So, more energy consumption takes place leading to shortening of network lifetime.

5.2 Quantitative Analysis

The effectiveness of the proposed node deployment scheme, reported in section 4.2 is evaluated through simulation. Moreover all the theoretical claims made through qualitative analysis presented in section 5.1 are justified by simulation results.

5.2.1 Simulation Environment

The simulation is performed using MATLAB (version 7.1). Simulation results of PDFND are compared with two existing schemes namely node deployment with Gaussian distribution (NDGD) [12] and node deployment with Uniform distribution (NDUD) [15]. We assume perfect MAC layer issues while considering this work. Extensive simulation has been performed and average results of 2000 independent runs have been taken while plotting the simulation graphs.

5.2.2 Simulation Metrics

To evaluate the performance of PDFND, energy balance and network lifetime as defined in sections 5.1.1 and 5.1.2 respectively have been considered as performance metrics. The number of deployed nodes is varied from 120 to 700. We define two more parameters-energy consumption rate per node in a layer and average residual energy of each layer for evaluating the extent of energy balance in the network**.** Though the concept of the parameter energy consumption rate per node is used in section 5.1, the same is formally defined here.

Energy consumption rate per node (ER): It is defined as energy consumption by a node per unit time. It is evaluated as $\text{ER}=\dfrac{\text{Energy consumption of a layer}}{\text{Number of nodes in the layer}}$

Average residual energy per node (Avg RE per node): It is defined as the residual energy of a node in a layer after network lifetime ends. It is evaluated as

$$\text{Avg RE per node}=\frac{\text{Sum of residual energy of nodes in a layer}}{\text{Number of nodes in the layer}}$$

Two sets of experiments are conducted for evaluating the performance of the present scheme and two other competitor schemes. One set of experiment measures energy balancing in the network and the last set verifies the enhancement of network lifetime. For each set, experiments have been conducted for two different network

sizes viz. network with 3 and 7 layers deploying 120 and 700 nodes respectively. The parameters and their corresponding values used for simulation are listed in Table 1.

Table 1. Simulation Parameters

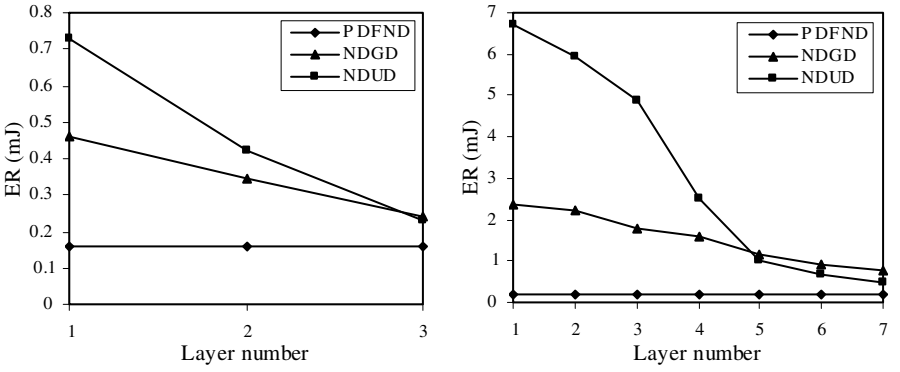
Parameters	Value
Initial energy $(E_{Initial})$	50 J
e_{elec}	50 nJ/bit
e_{amp}	10 pJ/bit/m^2
Communication range of a node (R_c)	160 m
Sensing range of a node (R_s)	80 m
Information generation rate (ρ)	0.1 bits/sec
Interval of periodic data collection (q(t))	1 sec
Network area	1 km^2 ~ 5 km^2

5.2.3 Energy Balancing

In this sub-section energy balancing of the scheme is evaluated in terms of the following two parameters.

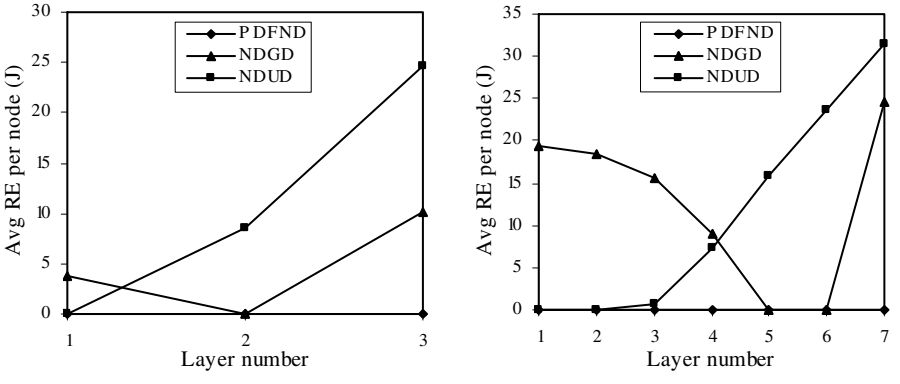
5.2.3.1 ER. Figure 3 shows ER for different network sizes. We observe that for PDFND the ER for a particular network size is constant for all layers and this rate varies with network sizes. For example, for network with 3 and 7 layers the ER is 0.16 mJ and 0.21 mJ respectively. These results imply that the ER for PDFND increases with increase in network area. In NDGD, ER for different layers varies considering a particular network size. So for NDGD, irrespective of network size, node in layer-1 has the maximum ER and node in farthest layer has the lowest ER. Therefore, nodes deployed in layers nearer to the sink drain out their energy much more quickly in comparison to nodes deployed in layers farther away from the sink. Similar observation holds for NDUD. This justifies our claim that PDFND is energy balanced whereas this is not true for both NDGD and NDUD.

5.2.3.2 Avg RE Per Node. Figure 4 illustrates the comparison of PDFND with NDGD and NDUD considering avg RE per node as a performance metric. Node deployment using NDGD or NDUD, results in an abrupt change in avg RE per node in each layer, independent of network size. In NDGD, energy of nodes in layer-2 (Figure 4(a)) and nodes in each of the layers-5 & 6 (Figure 4(b)) is drained out completely, though the nodes of other layers in the network retain sufficient energy for carrying out normal network operation that causes the phenomenon of energy hole. Similarly for NDUD, energy of nodes in layer-1 (Figure 4(a)) and nodes in each of layers-1, 2 (Figure 4(b)) is drained out completely though the nodes of other layers in the network have adequate energy for normal network operation. So NDUD also suffers from energy hole problem.



(a) 3-layer network (b) 7-layer network

Fig. 3. Energy consumption rate per node for various network sizes

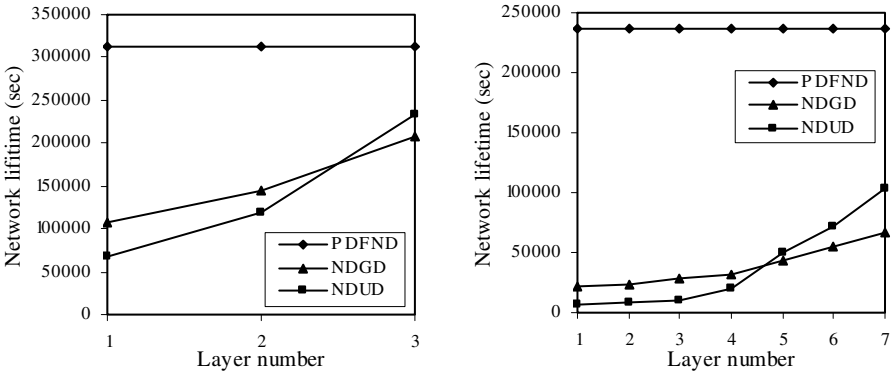


(a) 3-layer network (b) 7-layer network

Fig. 4. Avg RE per node for various network sizes

5.2.4 Network Lifetime

We consider data collection interval (q(t)) as 1 sec. The graphs in Figures 5(a) and 5(b) represent the network lifetime for two different network sizes. The network lifetime of PDFND is 50.65% and 55.02% more than that of NDGD and NDUD respectively for 3-layer network. For 7-layer it is 83.91% and 83.61% more than that of NDGD and NDUD respectively. Moreover, in PDFND the flat nature of the plot ensures that network lifetime terminates in more or less same time in all the layers as compared to NDGD and NDUD. This ensures energy in PDFND is balanced to a greater extent than both the competent schemes.



(a) 3-layer network (b) 7-layer network

Fig. 5. Network lifetime for various network sizes

6 Conclusion

In this work we have proposed a pre-determined node deployment scheme in wireless sensor network using a probability density function defined by us. The target of the scheme is to achieve energy balancing and enhance network lifetime while maintaining coverage and connectivity. We have provided theoretical formulation of energy balancing and network lifetime. Based on this analysis we have derived certain constraints, involving network parameters, to be satisfied to achieve the target. An algorithm is also developed to implement the scheme. We claim that our scheme successfully achieves the target. The claim is substantiated by performing both qualitative and quantitative analysis. Finally the results of quantitative analysis are compared with two existing works [12] [15] of node deployment that clearly demonstrates our scheme's dominance over the existing works.

As a future extension of our work, the deployment strategy may be made more realistic by considering 3-D environment. Moreover, the scheme may be analyzed for further improvement considering various QoS parameters.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Communications Magazine 40(8), 102–114 (2002)
2. Younis, M., Akkaya, K.: Strategies and Techniques for Node Placement in Wireless Sensor Networks: A Survey. Ad Hoc Network Journal 6(4), 621–655 (2008)
3. Li, J., Mohapatra, P.: Analytical Modeling and Mitigation Techniques for the Energy Hole Problem in Sensor Networks. Pervasive and Mobile Computing Journal 3(3), 233–254 (2007)
4. Lian, J., Naik, K., Agnew, G.: Data Capacity Improvement of Wireless Sensor Networks using Non-uniform Sensor Distribution. Journal of Distributed Sensor Networks 2(2), 121–145 (2006)
5. Cheng, Z., Perillo, M., Heinzelman, W.B.: General Network Lifetime and Cost Models for Evaluating Sensor Network Deployment Strategies. IEEE Transactions on Mobile Computing 7(4), 484–497 (2008)

6. Zou, Y., Chakrabarty, K.: Uncertainty-aware and Coverage-oriented Deployment for Sensor Networks. Journal of Parallel and Distributed Computing 64(7), 788–798 (2004)
7. Agarwal, P.K., Ezra, E., Ganjugunte, S.: Efficient Sensor Placement for Surveillance Problems. In: Krishnamachari, B., Suri, S., Heinzelman, W., Mitra, U. (eds.) DCOSS 2009. LNCS, vol. 5516, pp. 301–314. Springer, Heidelberg (2009)
8. Ganesan, D., Cristescu, R., Lozano, B.B.: Power-Efficient Sensor Placement and Transmission Structure for Data Gathering under Distortion Constraints. In: Proceedings of IPSN, pp. 142–150 (2004)
9. Wu, X., Chen, G., Das, S.K.: On the Energy Hole Problem of Nonuniform Node Distribution in Wireless Sensor Networks. In: Proceedings of IEEE MASS, pp. 180–187 (2006)
10. Chang, C.Y., Shih, K.P., Chang, H.R., Liu, H.J.: Energy-Balanced Deployment and Topology Control for Wireless Sensor Networks. In: Proceedings of IEEE GLOBECOM, pp. 1–5 (2006)
11. Olariu, S., Stojmenovic, I.: Design Guidelines for Maximizing Lifetime and Avoiding Energy Holes in Sensor Networks with Uniform Distribution and Uniform Reporting. In: Proceeding of IEEE INFOCOM, pp. 1–12 (2006)
12. Wang, D., Xie, B., Agrawal, D.P.: Coverage and Lifetime Optimization of Wireless Sensor Networks with Gaussian Distribution. IEEE Transactions on Mobile Computing 7(12), 1444–1458 (2008)
13. Karp, B., Kung, H.T.: GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In: Proceedings of ACM MOBICOM, pp. 243–254 (2000)
14. Tian, D., Georganas, N.D.: Connectivity Maintenance and Coverage Preservation in Wireless Sensor Networks. Ad Hoc Networks Journal 3(6), 744–761 (2005)
15. Liu, B., Towsley, D.: A Study of the Coverage of Large-scale Sensor Networks. In: Proceedings of IEEE MASS, pp. 475–483 (2004)

Appendix

From equations 3(b) and 3(c), the number of nodes deployed in layer-i and layer-(i+1) is given as

$$T_i = p_i \times T_{total} = \frac{k(2i-1)A_i}{N^2 i^4} \times T_{total} \text{ and } T_{i+1} = p_{i+1} \times T_{total} = \frac{k(2i-1)A_{i+1}}{N^2 i^4} \times T_{total}$$

respectively.

Dividing T_i by T_{i+1} and canceling k, T_{total}, and N^2 from denominator and numerator in sequence, we have

$$\frac{T_i}{T_{i+1}} = \frac{p_i}{p_{i+1}} = \frac{(2i-1)(i+1)^4 A_i}{(2i+1)i^4 A_{i+1}}.$$

The above relation is simplified by replacing the value of $A_i = (2i-1)\pi r^2$ and $A_{i+1} = (2i+1)\pi r^2$ as

$$\frac{T_i}{T_{i+1}} = \frac{(2i-1)^2 (i+1)^4}{(2i+1)^2 i^4}.$$

Towards Combining Individual and Collaborative Work Spaces under a Unified E-Portfolio

Hugo A. Parada G., Abelardo Pardo, and Carlos Delgado Kloos

Department of Telematic Engineering,
University Carlos III of Madrid, Spain
{hparada,abel,cdk}@it.uc3m.es
http://gradient.it.uc3m.es

Abstract. E-portfolios in learning environments have been attributed numerous benefits and their presence has been steadily increasing. And so has the variety of environments in which a student participates. Collaborative learning requires communication and resource sharing among team members. Students may participate in multiple teams throughout a long period of time, sometimes even simultaneously. Conventional e-portfolios are oriented toward showcasing individual achievements, but they need to also equally reflect collaborative achievements. The approach described in this paper has the objective of offering students an e-portfolio as a local folder their personal computer containing a combined view of their individual and collaborative work spaces. The content of this folder can be synchronized with a remote server thus achieving resource sharing and publication of a clearly identified set of resources.

Keywords: e-portfolio, collaborative learning, personal learning environment.

1 Introduction

Although the presence of information and communication technology in the realm of learning experiences has been increasing significantly over the years, an equally significant increase has also been perceived in the orchestration of any experience. Scenarios in which students were simply given a set of resources with some basic (and typically fixed) organization have evolved to learning scenarios in which students interact in groups, search for additional resources, communicate using different structures and patterns, etc.

This emerging scenario requires highly flexible learning spaces to provide the right environment for each stage of a learning experience. Learning Management Systems (LMS) were initially conceived to solve administrative issues and their structure relies heavily in the concept of "course" or "classroom". But this approach is now showing clear signs of fatigue. There are several forces that are re-shaping the landscape. Two of them are relevant to the approach presented in this document. The first one is that learning is relying more in multi-modal

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 488–501, 2011.

communication. The exchange of information between students and instructors or among students themselves supported by different tools and technologies, is shown to have a positive effect in the learning process. Collaborative learning, for example, proposes the creation of students groups in which the interaction among team members is the basis for achieving a set of objectives.

The second changing force is "student-centered learning". The student is now the central part of the strategy to achieve a successful learning experience. Activities now may require students to individually collect, organize and create new resources as a consequence of the learning process.

When combining these two forces, the need for flexible learning spaces emerges, and this need is not fully covered by conventional LMSs. Tools such as for example e-portfolios or social learning platforms are emerging as solutions to accommodate this type of scenarios. A student may participate in numerous groups. Some of them may be embedded within a course, and others may step out of the course boundaries. From the point of view of the instructors, an efficient and flexible group management would help cope with the increase of complexity when orchestrating non-trivial collaborative experiences.

In this document a platform is proposed to offer students a folder in the space of their personal computer with a structure to combine a personal area, a public area that is sent to a remote server and available in the net, and as many shared spaces as the number of groups in which they participate. The work focuses on the resource management aspect and how to manage efficiently a large number of student groups.

2 Related Work

Portfolios have been used for years in different disciplines to support personal development. Due to the adoption of information and communications technologies to implement it, they were called E-portfolios; which were initially defined as a selective and structured collection of digital resources gathered for the purpose of documenting experiences, storing intellectual assets, showcasing achievements, assessing personnel, and retrievable using a medium such as the Web, DVDs, etc [3,8]. But such an ample definition, when used in the context of a learning experience, has given rise to an equally large set of potential benefits. A review of the research literature shows that e-portfolios promote more effective learning, can be used as a form of assessment, may facilitate employment, encourage life-long learning, allow reflection on the artifacts included, obtain feedback from peers, and more generally, allow a different mode of interaction.

But as some authors have pointed out [2,7], much effort has been devoted to describe and emphasize the advantages for institutions, instructors or "learning" in general, but not so much to the impact on the students.

In [12] Jafari describes the e-portfolio role in higher education and points out the danger derived from the variety of meanings that the term represents for different actors of the educational arena. Educators, career center directors, department chairs and students all have different views of what an e-portfolio

is and how it should be used. This variety of views has caused confusion and hindered the adoption of e-portfolios. The proposed solution is the use of different identifiers preceding the term "e-portfolio" to help clarify its functionality. If this recommendation were to be used in this work, the proposed functionality could be called "local folder-based e-portfolio".

In the education area e-portfolios have been used to follow the students progress as to how they are achieving the course goals; and to become aware of student performance during the course based on it. Highlighting this line, in [4] Bhattacharya *et. al.* propose an e-portfolio aimed at raising awareness of the set of skills acquired by the students by selecting and analyzing the most significant resources. This task is complicated when students will have a mixture of resources created individually and collectively in various teams. A central repository combining this resources is the strategy proposed in this paper.

One important aspect of the use of e-portfolios in an educational environment is how easy knowledge from different sources can be integrated. In [5] the use of student e-portfolios is proposed in science and technology both for group and individual tasks. The authors claim that students working in groups experienced a richer environment when compared to those working individually. However, no technical details are given as to how the e-portfolio is implemented.

A critical stage when using e-portfolios in a learning environment is their deployment and adoption by the students. In [16], Murray and Currant point out that engagement by students is paramount for an e-portfolio to have an impact. Typical web-based platforms may encounter some resistance from the students due to their "rigidity". The approach proposed in this paper is based on managing a folder stored in the user local space (for example, a personal computer). On the other hand regarding e-portfolio adoption in [11] Hämäläinen *et. al.* consider the following two aspects as key: first, that the integration of information created during the course into an e-portfolio from different sources should be made as simple as possible; and secondly, the e-portfolio has to be well-organized. Our proposal considers these two aspects by allowing the planning and creating of a flexible reference structure involving links to external folders by instructors in order to host the student's artifacts.

Collaborative learning is conceived as an effective strategy to improve learning, and it can be improved by using tools supported by pedagogical goals. In [7], Carrol *et. al.* present a tool to encourage both reflective thinking and collaborative reflective learning. The work describes *Dotfolio*, a tool designed and used to support collaborative learning which includes functionality to encourage the exchange of ideas among teachers and students, collaborative reflection about course topics, etc. This tool highlights the importance of collaborative activities as the result of individual efforts rather than individual isolated activities. Our work proposes a solution oriented towards scenarios in which students participate in multiple group activities, some of them simultaneously, for a long period of time.

Most e-portfolio implementations used to be mainly oriented to sharing individual knowledge and experiences (after all, that was the essence of the classic

portfolio). However this trend is changing. In [14] Limanauskiene *et. al.* claim that the success of e-learning depends not only on the individual's abilities but also on how individuals share these abilities with others for building innovations. Based on this premise they propose a group e-portfolio implementation aimed at fostering the sharing of knowledge among learning community members. This work proposes the creation of a shared space for individual and collaborative tools so that learners can have a group environment from which individual e-portfolios can be accessed. The group portfolio is conceived as a portal integrating environments such as social networks, Learning Management Systems, learning community portals and personal portfolios. However, the strategy adopted is based on users sharing their individual portfolios within a group. Our contribution aims at maintaining a clear yet intuitive distinction between different collaborative spaces and the individual portfolio.

Our work is similar in spirit to that presented in [10]. The process of design an e-portfolio tool is described in which collaborative learning has a paramount presence. In [19], Wang describes a collaborative learning experience in which students use the cloud application "DriveHQ" to share all the documents within the teams.

The work presented in this paper is similar in the sense that it extends the e-portfolio functionality by combining the student individual work space (his/her personal computer) with the various collaborative environments in which he/she takes part for an arbitrarily long learning experience. The focus of the contribution is in offering a single folder to be used in a personal computer in which students may instinctively work on private documents, contribute resources to a group activity, or select a subset of this resources to make them publicly available on the net. The proposed approach can be easily complemented with current platforms to offer virtual spaces to interact containing other conventional tools such as forums, chats, etc.

The revision of the literature on e-portfolios, shows that most works propose mechanisms to support the individual work of the students. So far they do not go into great detail as to an e-portfolio contribute to collaborative work; and how they can be applied to a specific scenario. It is also evident that aspects related to how to combine individual e-portfolio with results derived from collaborative were not addressed.

Taking the advantages of individual e-portfolio as starting point in this work we extend its functionality by integrating it with resulting artifacts from group work. It is based on a "reference work space" which allows designing an initial structure of both the individual and shared work space. This reference work space contains the combination of individual and groups spaces and can then be easily replicated to be used by the students in a course. In addition our proposal facilitate planning the course activities, monitoring student progress, and make changes to the work space structure when new requirements arise.

3 E-Portfolio for Multi-collaborative Environments

As described in Section 2, e-portfolios are typically used as a main tool to support the personal learning process and showcase individual student achievements. By preparing these resources and sharing them with peers, students may also work on reflection and obtain valuable feedback. The work proposed in this paper tries to embrace these advantages and combine them with a collaborative setting. In order to simplify its presentation, a specific scenario has been created. Let us consider a one semester course which is part of an engineering degree program. The course adopts a collaborative team strategy for a software development project. The objective is to expose students to professional practices. Several instructors are assigned to different sections of the course. During the semester students must take part in both individual and group activities. Each activity is related to a project with the objective of producing an artifact. These artifacts are stored in a work space which is organized into several folders.

The artifacts may be of two types: those generated as results of the individual work, and those derived from team work. Therefore the student work space is divided into two areas: personal and team work spaces. Students that are members of the same team have access to the same content in a folder. Individual spaces remain private. Students may submit new artifacts at any time to a central server. For those submissions, they choose to place them on the private or team work areas. The type of feedback received from the tutors on the artifacts may also be individual or from the team members. Teams are created based on the instructor's observation during the first weeks of the course and work is carried out independently for each team. Team structures may vary slightly during the semester depending on unforeseen circumstances (for example, severe conflicts). At any time, each student has an area where the individual as well as team artifacts are stored.

This environment offers a more flexible space for students to manage the information available during the course so that they can make a more productive use of it. Additionally both students and teams can host additional material and links to external information sources in their work space; and a set of artifacts in a specific area can also be distinguished as openly available through a web page. With this approach, we want the students to make a more efficient use of the course resources, and for tutors to simplify the management of team management and supervision. Figure 1 illustrates the scenario we have just described.

But managing this multi-collaborative scenario translates into a set of tasks for the instructors: define the initial structure of these work spaces, decide team sizes, folder structure, access policy, permissions, initial project documentation, etc. Ideally, the combination of individual and group spaces should be created for each student. The proposed tool provides support to lower the administrative tasks to manage a set of students which are given access to an individual folder in which they can produce any artifact, plus a team space in which these artifacts are created collaboratively. The system offers primitive operations to define access policies. Based on these primitives the instructors can describe the team definition rules by assigning the users to each work space. During the course,

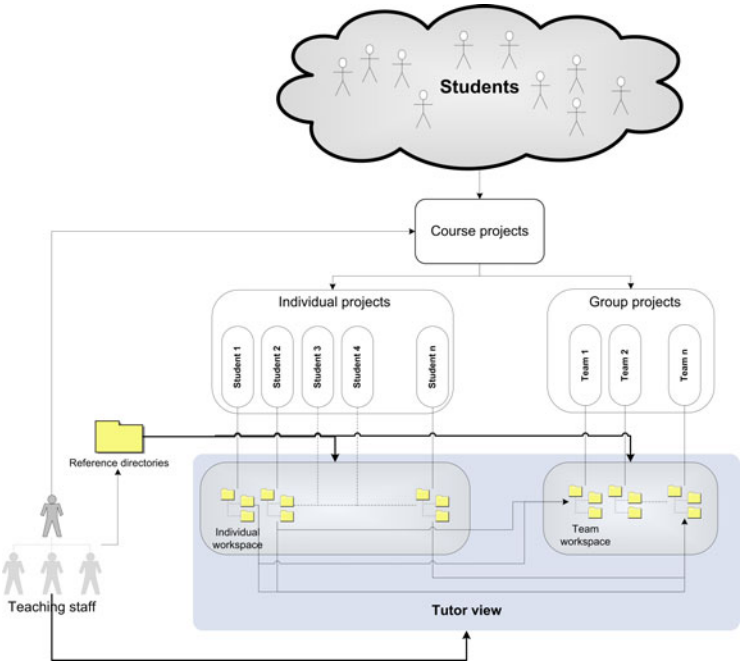


Fig. 1. E-portfolio scenario

tutors can interact with student either by reading, modifying and adding new information to the content folders or by creating new folders. This interaction can occur at any time during the course or can be regulated through the definition of a set of timestamps defining a window during which operations are allowed in a folder. This mechanism is used to enforce submission deadlines. Tutors may also create additional teams during the course enactment.

In order to manage the complexity derived from this scenario we propose an e-portfolio that is managed directly by the student just like any other regular folder in his/her personal computer. This e-portfolio expands from the individual space to a multi-space. It can be created with an initial basic structure. Tutors can modify the initial structure during the course and these changes are then transferred to the student spaces. Tutors can create different views of these spaces by selecting just the relevant parts at each moment. Within a single space, tutors may compare all the resources that a student is using in the different scenarios. Analogously, students are offered a folder in their individual computers containing all the resources in which they worked for the different scenarios.

The process of building and managing these new folders is structured into the following steps: design the initial structure of the portfolio, assign students to the teams, and create the e-portfolio. First, instructors define the basic elements from which the student spaces will be instantiated. Then a set of templates that represent folders and sub-folders are defined. We have called these templates the

reference directories. After that the links pointing to the collaborative folder are also defined.

Once these spaces are defined, students are then assigned to the teams. For each grouping, the instructor defines a set of rules to replicate the team's shared spaces withing the individual spaces. All course students will have assigned a directory reference where both the individual and team sources are available.

Once the folder structure and the group rules are created, the reference directories with the initial content are replicated. Each student starts with the same initial local folder with the same structure with as many links to the team shared spaces are required.

4 Proposed Implementation

In the specific context of engineering studies, some recent studies suggest [1] that team projects are important components of undergraduate curricula to expose students to professional practices. The same tendency is observed in the context of Internet where numerous web-based portals for software projects have appeared (see [6] for a survey). Version control systems (henceforth simply VCS) are tools used in these scenarios to exchange all types of resources among team members and maintain the history of changes. There are numerous VCS tools available and their functionality has been increasing steadily over the years; among the more popular we highlight Subversion [9], CVS[1], Mercurial[2], and recently Git[3].

Although our proposed digital portfolio structure could be implemented using any VCS, we have selected Subversion because it provides a more fine-granularity access control policy. Thus our approach allows a group of users to share a set of directories stored in a centralized server. With the current approach, the use of the digital portfolio has the added value of exposing students to the use of these industry-type tools typically used in team based projects.

In educational institutions, however, these systems are typically introduced (if at all) in the last courses of a program. But as stated by Reid [18], there is clearly a pedagogic value in using them also in the early courses. Other authors even consider the use of these tools "imperative" [15]. In the portfolio application, though, the objective is allow the creation of as many shared and individual spaces as needed for a large set of students.

The implementation of the proposed digital portfolio is based on three data representations: a set of reference work spaces, a set of connecting rules, and a set of group definitions.

4.1 Reference Work Spaces

The first concept on which this application is based is the so called "reference work space". It is simply a folder containing the initial structure of a work space

[1] www.cvshome.org
[2] mercurial.selenic.com
[3] git-scm.com

to be offered to a set of students. The application allows the manipulation of a set of these reference work spaces to be later cloned as many times as specified by the group definitions.

The most important feature of these work spaces is that they are replicated into multiple instances while maintaining the relation with the initial structure. This relation, together with the capability that VCS allow to "merge" changes introduced by the instructors on the references for a fast propagation of these changes to the replicated instances.

The change propagation is specially useful when, half way down the life of a shared space, an initial resource (copied in all instances) needs to be modified, or the folder needs to be internally reorganized. Instructors may apply these changes to the reference work space and the tool propagates them to the different student spaces. The process is automatic as long as the underlying VCS is capable of merging these changes. When this is not possible, there are mechanisms to solve these conflicts, although they require the intervention of the instructors.

The idea behind maintaining several reference work spaces is to create one for each different "realm" where students are supposed to work. For example, a reference work space would define the initial structure of the individual portfolio, whereas each group in which the student participates would have its corresponding reference space as well.

4.2 Connecting Rules

But typically, students participate in an ever increasing variety of communities. Sometimes these communities are within one single course, some others span across course boundaries. In order to offer the combined view of all their work spaces, a set of rules needs to be defined such that all the student spaces are connected to offer an initial identical folder.

The notion of "connecting" two folders is a well known technique used internally by operating systems to simplify file management. The terms usually used are "symbolic links", "hard links" or "shortcuts". The idea is quite simple. At an arbitrary point in a file hierarchy a different folder is attached to appear as if it is physically at that point.

The proposed approach follows this analogy to allow the definition of a set of rules to combine a subset of the reference work spaces defined in Section 4.1. Figure 2 shows an example of the configuration files containing these two ingredients. In the left side of the figure three reference work spaces (with names "tree1", "tree2", and "tree3") are defined pointing to three locations in the central repository. In the left side, two connecting rules are defined. The first one connects the space in "tree2" as an additional child at the top level of "tree1". Analogously, the second rule defines the space in "tree3" also as another child at the top level of "tree1". As a consequence of these definitions a single individual work space with the structure of "tree1" is defined that has as two sub folder in the top level the copy of the structures "tree2" and "tree3" respectively.

This resulting folder is then cloned using the group definitions explained in the following section.

```
# Reference repositories
[reference.tree1] # Individual
root = IndSpace
[reference.tree2] # Pair space
root = Pairs
[reference.tree3] # Team space
root = Teams
```

```
# Connection rules
[connect.IndHasTree2] # Pair to
ind. root = IndSpace
path = .
child = tree2
[connect.IndHasTree3] # Team to
ind. root = IndSpace
path = .
child = tree3
```

Fig. 2. Reference Repositories and Connecting Rules Configuration

4.3 Group Definitions

In a generic learning scenario, a set of users may be organized into a set of groups with different arrangements depending on multiple variables. For example, within the same course, activities can be carried out in teams or in groups formed by one representative of each team. If we widen the scope, students may be grouped differently in several courses of the same degree program.

The discussed approach assumes a global user community in which at least two roles are defined: students and staff. Students can be divided following several grouping policies. Each grouping policy is captured with the following data:

- Name of the grouping policy to be used in grouping commands.
- Reference to one of the reference repositories (as described in Section 4.1).
- A prefix to be used to generate the names of the instances derived from the policy.
- The number of digits to use to create the group number to append to the previous prefix.
- A space separated list of staff users that may access the instances of the group spaces.
- A comma separated lists of groups. Each group is a space separated list of strings in which the first one is the group name, and the following are the user names of the group members.

Figure 3 shows the definition of three grouping policies. The "group.Individual" policy states that the reference work space "tree1" should be cloned six times, and each of them should be accessed by a single user. Also, the prefix policy is defined such that the user name is used as the name for the work space. Furthermore, the access to these work spaces is allow to all three staff members.

The second grouping policy (with name "group.Pairs") states that the "tree2" reference repository should be replicated for the three pairs of students described. These work spaces will be named with the prefix "Pair_" followed by correlative numeric values represented by two digits. The access to these work spaces is allowed only to two of the three instructors.

The third and last grouping policy (with name "group.Triplets") states that two team instances should be created with names starting with the prefix "Team_"

```
# Group descriptions
[group.Individual]
reference = tree1
prefix = @USERID@
staff = staff1 staff2 staff3
instances =
  user1 user1,
  user2 user2,
  user3 user3,
  user4 user4,
  user5 user5,
  user6 user6
```

```
[group.Pairs]
reference = tree2
prefix = 'Pair_'
fill_prefix = 2
staff = staff1 staff2
instances =
  Pair_01 user1 user2,
  Pair_02 user3 user4,
  Pair_03 user5 user6
[group.Triplets]
reference = tree3
prefix = 'Team_'
fill_prefix = 2
staff = staff2 staff3
instances =
  Team_01 user1 user3 user5,
  Team_02 user2 user4 user6
```

Fig. 3. Group Descriptions for individual, pairs and team cloning

followed by a two digit number. Two other instructors are allowed to access these work spaces.

4.4 Portfolio Creation and Manipulation

For the sake of clarity, an example of how the proposed approach can be deployed is described in detail in this section. Let us assume that three instructors `inst1`, `inst2`, and `inst3`, are in charge of creating, deploying and supervising the work of a set of six students (with user names `student1`, through `student6`).

For this user community, three working spaces are considered:

Individual. Each student has an individual work space with an initial structure. Such space contains a clearly identified folder the content of which will be available in the net. All instructors have access to this area.

Pair Group. Students are grouped in pairs (three pairs in the course) to carry out a set of activities. There is one work space for each pair of students shared among the pair members. Each instructor supervises on of the pairs.

Teams. Students are also grouped in teams of three (two teams in the course) to carry out a second set of activities. There is one work space shared among the team members. Instructors `inst1` and `inst2` supervise the first team, and `inst 2` and `inst3` supervise the second team.

Figure 4 shows the three reference work spaces initially created by the instructors. Only the folders are shown to simplify the view. Instructors place in these areas all the initial files to be given to the students. For example, folder `public` in the `tree1` may include a file stating that the folder is meant to contain documents to publish in the net. Thus, any document uploaded in that area becomes public once it is committed to the central repository.

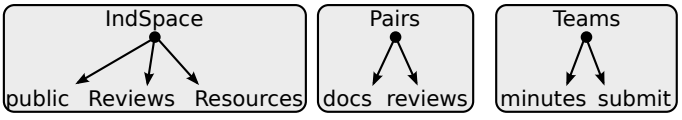


Fig. 4. Three reference work spaces created by the instructors

The definition of these three directories as reference work spaces is accomplished with the configuration shown in the left side of Figure 2. Next, a monolithic work space is created by defining how the three work spaces are combined. In this case, we want that every instance of the individual folder contains two instances of the pairs and team work spaces at the top level. This accomplished with the configuration rules shown in the right side of Figure 2.

Finally, the way in which reference spaces are to be replicated is captured by the configuration shown in Figure 3. The individual work space is replicated for each student, and at the same time combined with a space shared by the pairs and another space to be shared by the teams. Additionally, the right permissions are defined for each folder in the space.

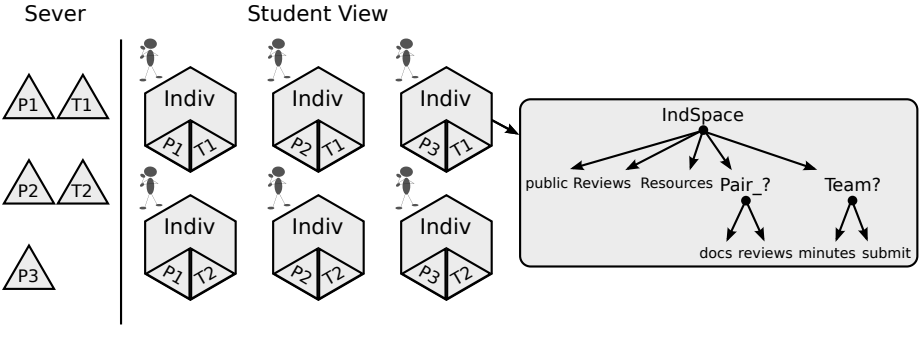


Fig. 5. Final set of work spaces in server and student view

Figure 5 shows the resulting global working environment once the configuration rules are interpreted and executed. Six individual spaces are created by cloning the reference, and their corresponding working spaces for the work in pairs and teams appear as another regular folder.

5 Discussion and Future Work

With the proposed approach, several important features are combined into a single solution. In its simplest form, instructors may create an initial structure of a folder where students collect the material related to the learning experiences in

which they are participating. This initial structure contains a public area which is publicly available to showcase the achievements. Instead of designing a page in a web-base platform, the work in their regular documents and the work in a page can be intermixed.

When students become members of a team as part of a collaborative experience, their personal space is modified with the appearance of a folder shared among all team members. Access rights for the documents is given only to the team members and staff. The exchange of documents within the group and with the tutor is significantly simplified. These documents are sharing a space which is stored remotely. When, either within the same course, or an entirely different experience, students participate in other groups, their personal space is again extended with a new folder. Access rights can be properly adjusted even to offer a window of validity during which changes are allowed in the contained resources.

A concrete example of a collaborative activity will help to clarify the proposed functionality. Let us assume a course with 20 students in which a "jigsaw" [13] pattern is used. Initially students are divided into four "expert profiles". There are five students in each profile. Initially all students have an initial folder with their private files. When divided into these four profiles, four groups are created. Each group is given a different set of resources to review and exchange their remarks. For each student, a new folder appears with the appropriate resources, and the folder is shared among the peers with identical profile.

In the second step, 4 member team work spaces are created taking one student from each profile. A second work space appears for each team in which a new set of resources is exchanged. At the same time, instructors are provided full access to these spaces to be able to monitor their progress as well as to provide feedback about the different tasks.

With the proposed solution, the creation/deletion of these new spaces can be done easily, for an arbitrarily large number of students, fully embedded into their personal learning environment, and with the possibility of easily applying changes to any of the resources.

A first set of pilot experiences were conducted in which students were given access to a set of work spaces not yet integrated under the proposed single folder view [17]. Students were given a first shared space in which a set of activities were carried out in pairs. At a second stage, with the same cohort, teams of four students were created, and a second shared space was created from an initial structure given by the instructors.

Students adopted the working paradigm with no problem. As a matter of act, the proposed scheme complemented nicely the "web disk service" offered by the university. Such disk can only be used for storing individual files, so the newly created spaces were ideal for exchanging documents within the teams.

The next step is to deploy a single unified view of the multiple work spaces. Together with this deployment, the course will include activities specifically oriented to the publication of certain resources in a place publicly available on the net. Once this deployment is finished, the more challenging task of assessing its impact in a learning experience will be tackled.

Finally, our approach was designed to allow its integration with other platforms of the learning ecosystem. Hence these topics have been keep out of scope of this paper. Then, several open issues should be taking into account in order to integrate our solution to this "ecosystem"; specifically we consider face the following: providing an access point to the contents generated by students in collaborative an social platforms; definition of mechanism to allow integration with learning management systems.

Acknowledgment

Work partially funded by the Learn3 project, "Plan Nacional de I+D+I TIN2008-05163/TSI", the *Consejo Social - Universidad Carlos III de Madrid*, the Acción Integrada Ref. DE2009-0051, and the "Emadrid: Investigación y desarrollo de tecnologías para el e-learning en la Comunidad de Madrid" project (S2009/TIC-1650).

References

1. Association for Computing Machinery. Computer Science Curriculum 2008: An Interim Revision of CS 2001. IEEE Computer Society, Los Alamitos (2008)
2. Ayala, J.: Electronic Portfolios for Whom? Educause Quarterly 29(1) (2006)
3. Barrett, H.: ICT support for electronic portfolios and alternative assessment: The state of the art. In: World Conference on Computers and Education (WCCE), pp. 1–11 (2001)
4. Bhattacharya, M., Heinrich, E., Rayudu, R.: Work In Progress: E-portfolios in Computer Science and Engineering Education. In: Proceedings of the Frontiers in Education. 36th Annual Conference, pp. 11–12 (2006)
5. Bhattacharya, M., Novak, S.: An Integrative Model for the Evaluation of E-portfolios. In: International Conference on Advanced Learning Technologies, July 2007, pp. 215–216. IEEE Press, Los Alamitos (2007)
6. Cabot, J., Wilson, G.: Tools for Teams: A Survey of Web-Based Software Project Portals. Dr. Dobb's (October 2009), http://www.drdobbs.com/tools/220301068 (last accessed December 2010)
7. Carrol, N.L., Markauskaité, L., Calvo, R.A.: E-Portfolios for Developing Transferable Skills in a Freshman Engineering Course. IEEE Transactions on Education 50(4), 360–366 (2007)
8. Challis, D.: Towards the mature ePortfolio: Some implications for higher education. Canadian Journal of Learning and Technology 31(3) (2005)
9. Collins-Sussman, B., Fitzpatrick, B., Pilato, C.: Version control with subversion. O'Reilly Media, Cambridge (2004)
10. Gerbic, P., Maher, M.: Collaborative self-study supporting new technology: The Mahara e-portfolio project. In: Proceedings of ASCILITE Melbourne 2008, pp. 320–324 (2008)
11. Hämäläinen, H., Ikonen, J., Porras, J.: Developing Technical E-portfolio Construction Process. In: 2009 Ninth IEEE International Conference on Advanced Learning Technologies, July 2009, pp. 711–712 (2009)
12. Jafari, A.: The "Sticky" ePortfolio System: Tackling Challenges and Identifying Attributes. Educause Review 39(4) (2004)

13. Johnson, D., Johnson, R.: Learning together and alone. Cooperative, competitive, and individualistic learning. Allyn and Bacon, 160 Gould Street, Needham Heights, MA 02194 (1994)
14. Limanauskiene, V., Blazauskas, T., Binkis, M.: Knowledge Sharing with Group Portfolio. In: International Conference on Advanced Learning Technologies, pp. 428–430 (July 2009)
15. Linder, S., Abbott, D., Fromberger, M.: An instructional scaffolding approach to teaching software design. Journal of Computing Sciences in Colleges 21(6), 238–250 (2006)
16. Murray, C., Currant, N.: E-portfolios along the Lifelong Learning Cycle: Differences between Use, Pedagogy and Context. In: International Conference on Advanced Learning Technologies, pp. 491–493. IEEE, Los Alamitos (2006)
17. Pardo, A., Delgado Kloos, C.: SubCollaboration: Large Scale Group Management in Collaborative Learning. Software: Practice and Experience 41(4), 339–465 (2011)
18. Reid, K., Wilson, G.: Learning by doing: introducing version control as a way to manage student assignments. In: Proceedings of the 36th SIGCSE technical symposium on Computer science education, pp. 274–276. ACM, St. Louis (2005)
19. Wang, Q.: Design and evaluation of a collaborative learning environment. Computers & Education 53(4), 1138–1146 (2009)

A Scenario Editing Environment for Professional Online Training Systems

José Luis Aguirre-Cervantes and Jean-Philippe Pernin

Laboratoire d'Informatique de Grenoble,
961 rue de la Houille Blanche,
38402 Grenoble Cedex
{Jose-Luis.Aguirre,Jean-Philippe.Pernin}@imag.fr

Abstract. This paper presents a methodology and a graphical software environment dedicated to support the activity of various actors (instructional designers, training managers, pedagogical engineers, developers, learners, etc.), involved in the production of professional online training systems. After having presented the basis of our specific "wheel methodology" for building learning scenarios for those systems, we detail the models on which the methodology relies. The process behind the methodology follows a goal-oriented and a template-based approach. Finally, a prototype environment and the first experimentations carried out with this environment are described.

Keywords: learning scenario, professional training, goal-oriented process, learning templates.

1 Introduction

In previous works, we have focused our research on developing methods and tools aimed to model learning scenarios according to an *"intention-based"* approach [1]. Particularly, the ISiS (Intention, Strategies, interactional Situations) model assists a scenario designer (typically a teacher in secondary school) by providing him with a methodology based on the elicitation of actors' intentions (teachers, learners, domain specialists, etc.) for expressing a learning scenario. This one is considered as a hierarchical plan composed of strategies, interactional situations and learning components. After having experimented with this model in academic situations, we propose to adapt it to specific industrial situations where the needs and design contexts are different.

Our work was mainly developed inside the Learning Game Factory project (LGF, 2009-2011), granted by the European Regional Development Fund - Rhône-Alpes. LGF gathers four laboratories and five industrial partners around the question of how to use innovative objects in learning systems. Its goals are to integrate, control and

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 502–517, 2011.

pilot, based on a pedagogical point of view, a large variety of components within professional online training systems. Concerned components can be learning games, simulations, interactive learning applications, etc. which can be developed in different languages or techniques. The aim of this paper is to present some of the main contributions resulting from this project.

Following this introduction, section 2 presents a specific methodology, the *"wheel iterative process"*, adapted for the building of professional training learning systems. We focus on providing involved actors with "agile methods" allowing an incremental definition and management of learning scenarios, by situating at the center of the process a goal-oriented and a template-based approach. We thus define a series of steps (device formulation, scenario design, scenario implementation, scenario execution, analysis and re-engineering), where the scenario is progressively enriched by different actors. For each step, the scenario can be characterized by a certain level of modeling. We propose then to consider those different models to constitute the basis upon which we have developed our environments.

In section 3 we present more precisely the core concepts of the two main models concerned with the scenario design step: the Context/Intention model and the Skeleton model. The elaboration of those models takes into account results of works about goal-oriented-design and intelligent agents' architectures.

Section 4 introduces Scenedit-II, the prototype of the scenario editor we have developed for the LGF Project. This environment provides the designers with three main *workspaces* from which they can progressively create the learning scenario. The first one, the *intentions workspace*, is dedicated to the definition of context and intentions, in terms of concerned knowledge, of associated constraints to the training context or of expected motivations for the learner and/or other actors (game motivations particularly). The second one, the *library workspace*, proposes a set of libraries where components (plans, strategies, interactional situations, game components, etc.) in the form of templates can be picked up or stored. The third one, the *scenarisation workspace*, allows to organize, by using different visual representations, the learning scenario activities. At every moment of design, the users can explore the *library workspace*, in order to look for components adapted to the intentions or contexts formulated in the *intentions workspace*.

In section 5, we present the first experimentations carried out with our tools, which partially implement the "wheel iterative process" for the production of a learning scenario. The first steps, device formulation and scenario design, are almost completely covered; scenario implementation and execution in a dedicated player are briefly considered for simple sequential scenarios. In this section, we also stress the fact that our architecture allows to handle different visual representations for the same learning scenario, where each representation is suited to different needs coming from different designers. We finish the paper with conclusions and perspectives on some future works.

2 A Methodology for Building Professional Training Learning Scenarios

2.1 Methodology as a "Wheel Iterative Process"

As stated in the previous section, we are concerned with the engineering process of learning scenarios for professional training. As figure 1 shows, several actors of different nature and accomplishing different tasks take part in this process. In the rest of the paper, we will refer to these scenarios as *training scenarios* or simply *TS*.

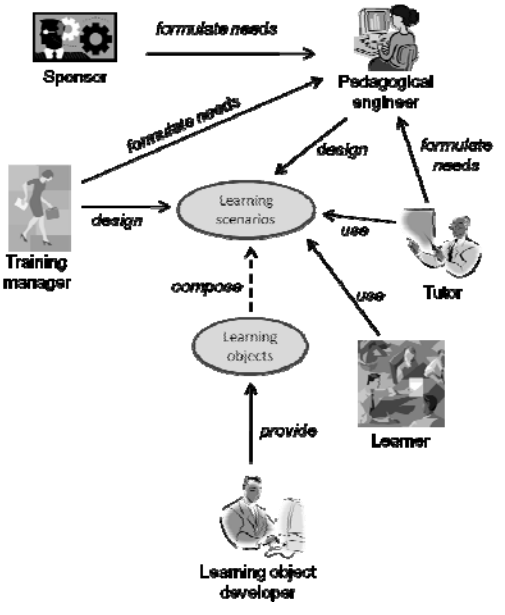


Fig. 1. Actors involved in the engineering of learning scenarios for professional training

Sponsors have the main responsibility of formulating the TS needs; TS needs, which can also be formulated by *training managers* and *tutors*, are captured and formalized by *pedagogical engineers*. Scenario design is the main responsibility of *pedagogical engineers*, *training managers* and *tutors*. *Learning object developers* are responsible of providing scenario components; finally, the learning scenario is mainly used by *learners* and *tutors*.

For this process, we propose a development methodology that follows an incremental iterative approach. This methodology, which is shown in figure 2, is inspired on results of previous research works on pedagogical scenario design [2] emphasizing the need of involving learning scenario designers in the elaboration of models and tools being proposed [3,1]. The methodology is strongly based on reuse concepts at several levels of granularity as such ones presented in [4].

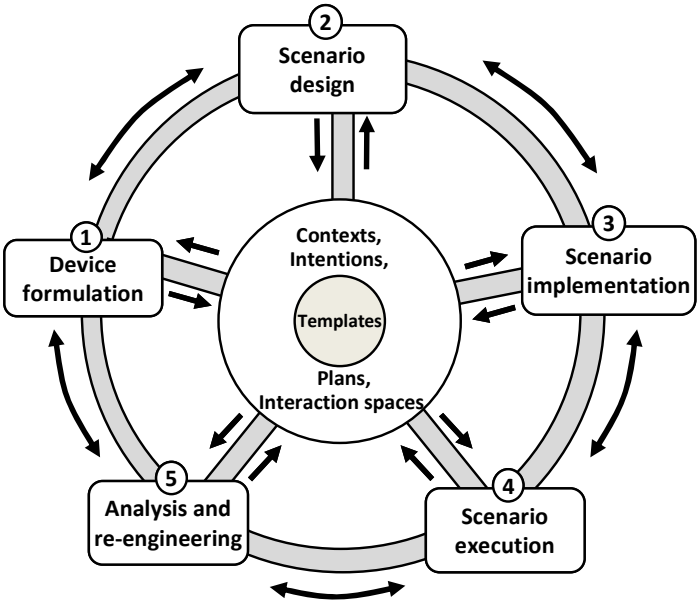


Fig. 2. Schema of an incremental iterative methodology for building professional TS

The process is directed by a collection of models situated in the center of the figure, as nodes surrounding the center represent the main phases constituting the methodology. Models can be originated from existing templates; they are elaborated by the actors appearing in figure 1, tackling different problems and being the main responsibles for the production of one or more specific models. For example, in the *Device formulation* phase, sponsors, training managers, tutors and pedagogical engineers must produce most of the *Context* and *Intentions* models. Moreover, production of models on later phases is guided by models defined in earlier phases. We will go deeper on that in section 2.2.

The figure shows that phases are not necessarily performed in linear order; on the contrary, the methodology could be seen as analogue to that one for developing software systems in a spiral and incremental way [5]. The arrows introduce a "wheel metaphor" reflecting the fact that, like in a wheel, sometimes we need to move forward (i.e. to get partial implementations of the scenario based on initial formulations), sometimes to move backward (i.e. to review the formulation based on partial implementations). In all cases, wheel movements are always dictated by the wheel hub, i.e. the models' contents in the center of the figure.

The proposal of this methodology is strongly influenced by the fact that the context of professional training learning scenarios development has some specificities which distinguish it from a more traditional academic context. In particular, the process is highly dynamic, i.e. the TS is constantly modified, and interactions between participants can last very short. Pedagogical engineers and sponsors can outline initial scenarios from the very first interview coming soon to initial model specifications. Some of these specifications can even directly concern advanced phases of the

methodology, and this is true all along the development process. This means that changes can be done to the models at any stage of the development process and following several directions.

One of the implications of such a methodology is that it can be considered as a workflow where different phases must be related by mediation of the models. This implies as well that we can *"short-circuit"* traditional linear flow, i.e. from *Device Formulation* to *Scenario Design* to *Scenario Implementation* and so on, permitting flows from one phase to any other phase. For example, scenario implementations could be modified without going first by changing initial formulations. As a result, the methodology is very *reactive,* so modifications to the implementation must be back propagated to modify the design of the scenario and what was set up in the initial formulation. To better clarify these ideas, suppose that the actors involved in the *Device Formulation* phase have established the intention to use a *competition simulation game*; if during scenario implementation, the actors access a template of a *role-playing board game* and decide to integrate it as one of the scenario's activities, the Context and Intention models defined at the beginning will reflect these changes.

In the next section we will briefly explain the different models coming out when the methodology is applied.

2.2 Methodology Models

In the *Device Formulation* phase, sponsors of the training device along with pedagogical engineers, and possibly training managers and tutors, set up the contexts where training will take place, and the intentions pursued by the training device with respect to those contexts. Typical examples of parts of the context are the targeted audience, the possibly roles played by that audience, the frames of knowledge and competences related to the training device, locations and resources used during training, etc. Intentions are next related with information included in the contexts. We will call the output of the first phase the Training Scenario Context/Intention Model (C/I Model); the form of the C/I Model is explained in section 3.1.

In the *Scenario Design* phase, pedagogical engineers in collaboration with training managers and tutors analyze the C/I Model for designing the skeleton of the TS. Taking into consideration the C/I Model, they specify what will be the main stages of the scenario and how those stages will be related, i.e. the logic of the scenario. This logic is defined in terms of operators indicating the way the stages will be chained: as a sequence, as a conditional alternative, as a single choice between several alternatives, as parallel branches, etc. General templates of common or specific activities can be used to define the skeleton, and also some specific activities can be attached to final stages. The output of this stage will be called the Training Scenario Skeleton Model (SkModel); the form of the SkModel is explained in section 3.2.

Next, the SkModel is refined by the pedagogical engineers and other actors like training managers, resulting in an execution model where more details are added about the use of the scenario. Specific training activities must be attached to all atomic stages of the SkModel. Attached activities can be created from scratch, or they can be recovered from libraries and repositories in order to be adapted to the specific needs of the training device. Based on what could happen in previous stages of the scenario, more subtle chaining conditions than those established in the SkModel are

defined for some of the atomic stages. For example, on the basis of resulting values of some parameters for precedent stages, one can specify what must be the initial state of another stage. Also, at this time training managers define the agenda for participants that will use the device during the training sessions: specific conditions about time and places, when and how the activities will be accessible, how the roles previously defined will be accorded to participants, evaluation rules for some of the stages, etc. The output of this stage is called the Training Scenario Execution Model (ExModel).

During the scenario execution phase, participants will work with the activities included in the TS. They will interact with players that can read and interpret the ExModels, and other actors like tutors or training managers, will follow the work of participants, analyze their performance, and possibly regulate at runtime some activities or adjust the scenario logic itself as a result of that analysis. As scenarios will be played by several actors, multiple outputs are the outcome of this phase. They will be called the Training Scenario Executed Models (ExedModel). Those models will include overall information about the main actions executed by the participants as well as some final results obtained by the evaluation means previously specified.

Finally, when the TS has been played by all the participants, a broader analysis can take place to evaluate the scenario itself. Validations can be done to verify, for example, if intended objectives were reached by the participants or if the inclusion of some activities has been valuable for enhancing participants' performance. Later on, the conclusions of those analysis can be taken into account for the adjustment of the same scenario to new training situations, or for defining completely new TS. The outcome of this phase is the Training Scenario Evaluated Model (EvModel).

3 Models and the Library Space

We present in this section models for the first two phases of the methodology: C/I Model and SkModel. The ExModel is by now partially developed as an extension and a transformation of the SkModel. The rest of the models are not considered for the moment. This section also includes the structure of a library of templates that can be reused for editing new TS.

3.1 Training Scenario Context/Intention Model

Following our approach, learning scenarios are first defined with respect to various contexts and intentions specific to the domain for which the TS is designed. In general, contexts are defined by all the objects and information that could play an important role for defining some part of the scenario, either a structural part or a behavioural part. For example, it is important, from the training perspective, to know where and how training will take place; what will be the places (physical or virtual) where learners will be trained; what kind of resources learners will access; what are the goal competences for which the training device is intended, etc. In our case, we identify three different types of context and intentions: organizational, training and knowledge. Figure 3 shows the UML class diagrams that represent the C/I Model.

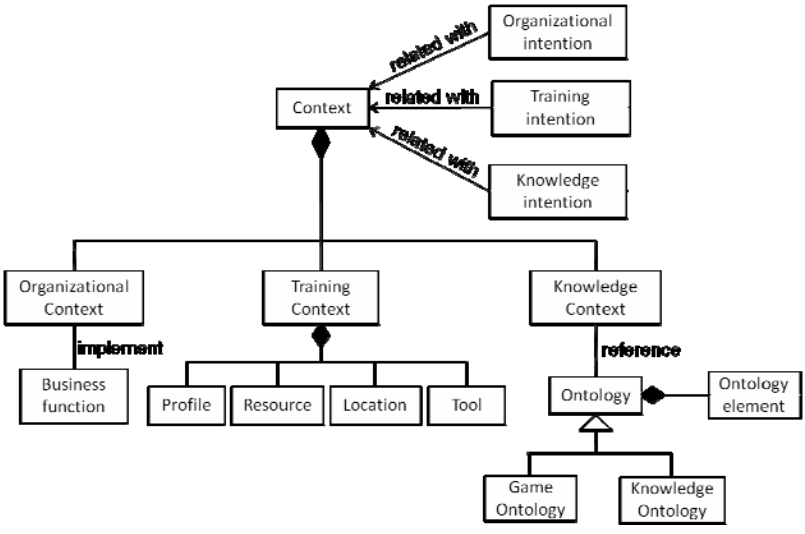


Fig. 3. C/I Model

The organizational context is restricted to some fields providing information about the enterprise, as its activity sector, a general description, and the business functions and services targeted by the training device. Organizational intentions are general statements related with the vision and strategies of the enterprise, but restricted to the business functions defined in the context. For example, *"to improve the quality of ... function"* or *"to reduce the waiting time of users for ... function providing ... service"*.

Training context enumerates and describes the targeted audience, the locations where the training device will be used, and resources and tools needed by the users of the training devices. Targeted audience is specified in the form of profiles referencing the business functions. Training intentions refer to general organisational and implementation aspects of the training device: organisation of sessions for the targeted audience, type of training device (distance learning, blended learning), tutor requirements, etc.

Knowledge context is provided by several ontologies, some of them of general nature, some others specific to the domains addressed by the training device. They include concepts and relationships between those concepts that are relevant to the training device. Knowledge intentions referred to knowledge and skills intended to be acquired by users of the training device. Those intentions can be expressed in structured way following the pattern:

`(<formulator>+,<target-public>+,<competence verb>+,<items of knowledge>+)` where

- `<formulator>+` refers to the persons who provide the intention (at least one); it could be any of the actors taking part in the process shown in figure 1.

- `<target-public>+` refers to the profiles specified in the training context (at least one).
- `<competence verb>+` denotes a competency list of verbs selected from competencies ontologies (at least one).
- `<items of knowledge>+` refers to one or more nodes of the ontologies constituting the knowledge context. A special kind of general knowledge and context intentions are for example those ones related with games. Game context furnishes general information about the use of games inside the enterprise and game intentions specify what are the intended forms of game planning to be used in the training device under the form of game motivations (an extended list of Caillois [6]) and game mechanisms [7].

3.2 Training Scenario Skeleton Model

The purpose of the training scenario skeleton model is to specify the scenario's logic. More specifically, TS logic is defined in terms of the chaining of the activities permitting the targeted audience to attain the intentions established in the previous C/I Model. The activities are supposed to be held in the contexts contained in this C/I Model.

Various educational modeling languages (EML) have been proposed for designing pedagogical scenarios. Certainly the most referenced in the literature is IMS Learning Design [8], which has been created as a "neutral language" for covering two complementary needs: (1) to be able to express a large variety of learning situations, and (2) to implement created pedagogical scenarios on a large range of EML (interoperability). Our objective is not to propose an alternative to replace such languages, but as underlined in [9], we think that it is necessary to develop more high level authoring or graphical environments adapted to the needs of specific designers. In our specific context of industrial development, we have identified a design step, where designers must manipulate easily the scenario's logic without taking into account implementation details. In further steps, it is possible to translate this logic towards others more "interoperable" notations, such as IMS-LD.

Inspired on the Belief-Desire-Intention agent architecture (BDI) [10], we propose to represent scenario skeletons as hierarchical plans, where the logic of the scenario is based on the use of plan operators. Beliefs, desires and intentions denote mental attitudes, representing respectively the information, motivational and deliberative states of an agent. These mental attitudes determine in fact the agent behavior. As beliefs represents in some sort what the agent knows about the world, desires represent objectives or situations that the agent would like to accomplish or bring about; goals represents consistent active desires. Finally, intentions represent what the agent has chosen to do, i.e. desires to which the agent has to some extent committed. With respect to our proposal, we can consider that the C/I Model encompasses some of the designer beliefs (what they know), and all of the desires and intentions (what they want).

In most of the BDI agents implementations, intentions are achieved by the execution of plans, maybe partially conceived at the beginning, and with details being filled in as the execution progress. A plan is defined as a sequence of actions achieving a goal [11]. We claim that the problem of building a learning scenario can

be considered as a partial-order planning problem. As a consequence, the solution of that problem, i.e. the learning scenario, can be represented as a graph of actions instead of a sequence. The graph is organized by means of operators whose arguments can be sub-plans or primitive actions.

To support our claim, we state that following an intention based approach, learning scenario designers must formulate the skeleton model selecting the activities that are considered to be the most compatible with the intentions, and arrange them on the base of chaining operators considering some partial order restrictions defined by the intentions. On the one hand, this means that some intentions must be attained before others, and logically, the activities addressing the first ones must be considered before the activities addressing the second ones. On the other hand, this means that in some cases, some activities can be placed in the scenario without specifying which one comes first.

Examples of plan operators for pedagogical purposes are the sequence itself, the alternative, the free choice, etc. Figure 4 shows the graphical representation of a scenario skeleton where terminal nodes correspond to specific activities and all of the internal nodes correspond to sub-plans of the main plan represented by the root of the structure shown on the figure. In the figure, sequence operators are represented by circles, alternative operators are represented by diamonds, and free choice operators are represented by triangles. As it will be presented in section 4, other visual representations are possible.

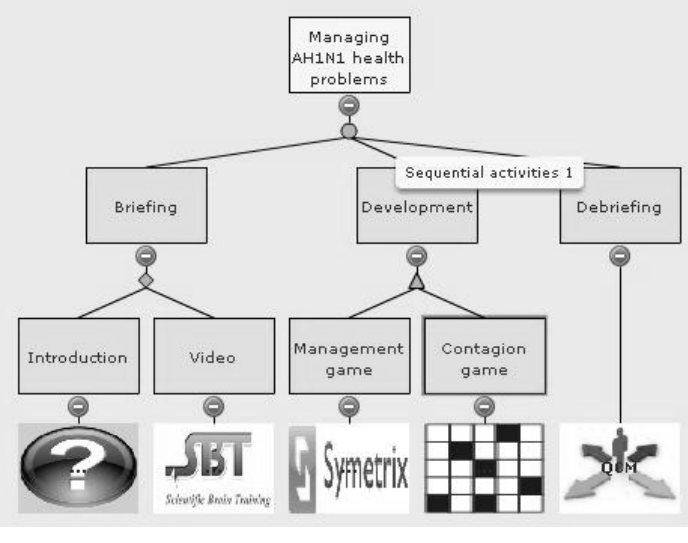


Fig. 4. Example of Skeleton Model

Skeleton terminal nodes correspond to conceptual objects denoted as *interaction spaces (IS)*. They reflect the fact that inside them, learners are confronted with the execution of the specific activities intended to work for achieving the intentions established in the scenario. In IS, learners can interact with dedicated interfaces, for

example to participate in a blog, to play a game, to work on the elaboration of a report, etc.

Figure 5 shows the UML class diagrams for the skeleton model, which is in fact an instance of the Composite design pattern [12].

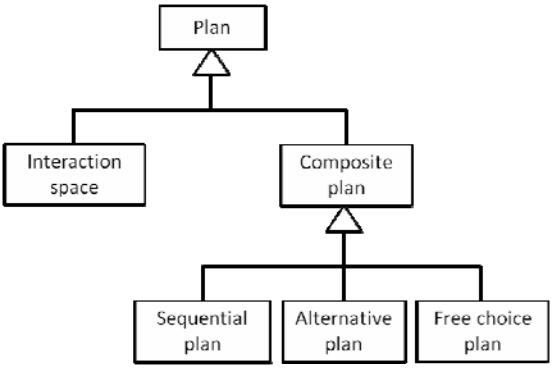


Fig. 5. UML Skeleton Model

Even if currently models are built by human agents, we believe that the fact of dealing with scenarios as plans can bring in many benefits in order to envision the development of artificial agents for assisting the tasks done when applying the wheel methodology. We will discuss about this in section 6 when talking about conclusions and future works.

3.3 Library Space and Interoperability

One of our concerns is the reuse of learning components at several levels of granularity, from basic non interactive components (videos, texts, etc.) to complex interactive objects (multiple choice questions, learning games, simulations, complete scenarios or part of scenarios, etc.). For this doing, we make use of learning object templates.

We must satisfy several requirements to incorporate the use of templates in all phases of the methodology presented in section 2. Specifically, during design and implementation phases, templates must be accessible to designers for being integrated through reuse and adaptation into the scenario's specification.

For design purposes, a template is a black box whose interface is contained in an information package expressed in the form of metadata. Metadata describes templates' properties for making possible their interoperability. By the moment, we propose an initial protocol allowing patterns to interoperate in terms of four capacities: *harvestability*, *configurability*, *observability* and *adaptability* (see figure 6).

Harvestability is the template capacity of being harvested, i.e. searched on according to filtering criteria; *configurability* is its capacity to be given an initial setting; *observability* is its capacity to be traced at runtime; *adaptability* is its capacity to be adjusted at runtime, i.e. to change its actual setting. The four capacities can be used at design time. For *harvestability*, templates are annotated using LOM format

[13]; for the other capacities we use a complementary XML file specification. For the LGF project, game classifications has been added to the <classification> section of the LOM specification.

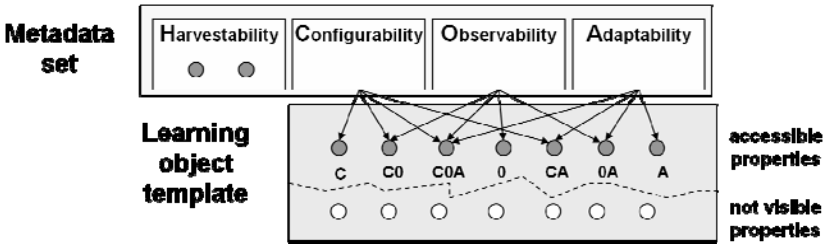


Fig. 6. Picture of a learning object template

4 ScenEdit-II: A Working Prototype

One goal within the LGF project was to offer a tool where different actors could work on the first phases of the methodology: from *Device Formulation* to *Scenario Implementation*. The ideas exposed in sections 2 and 3 were implemented in a prototype called *ScenEdit-II*. The tool's front-end is shown in figure 7.

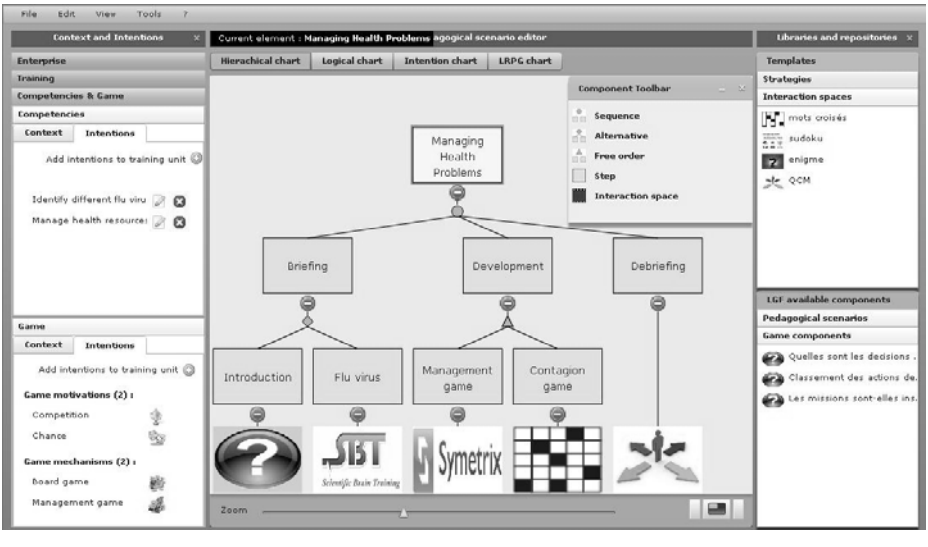


Fig. 7. An example of a training scenario developed with our editing tool

The first two phases of the methodology are completely included in the prototype and the Scenario Implementation phase is partially implemented. This will bring the

possibility to build the specification of TS that could be later executed. The interface is organized in three vertical workspaces from which one can progressively create the TS. Each space deals with the construction of one or more models produced by the methodology. The left part of the screen is dedicated to the construction of the C/I Model. Scenario designers can enter all the characteristics describing the 3 different types of contexts: Enterprise, Training and Knowledge (Competencies). Knowledge intentions can be entered in the structured format of section 2.1 by the use of graphical interfaces that reference ontologies' nodes that constitute the knowledge context (see figure 8).

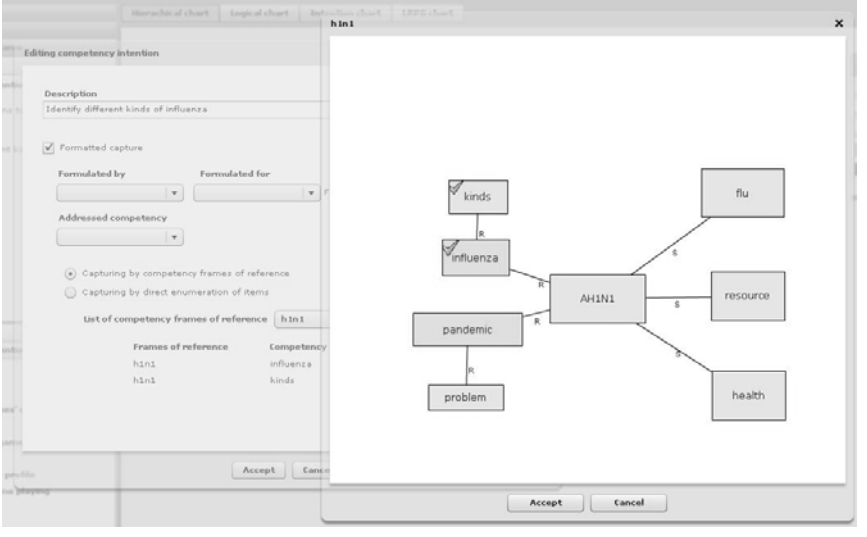


Fig. 8. Graphical interface for edition of knowledge intentions showing the graphical representation of an ontology

The central part of the screen is dedicated to the edition of the Skeleton and the partial Execution Models. Figure 7 illustrates the skeleton of a training scenario where the terminal nodes correspond to specific components based on different templates, while the internal nodes correspond to groups of activities organized by operators. In the figure, circles represent sequence operators, diamonds represent alternative operators, and triangles represent free choice operators. Thus, the scenario is composed of a sequence of three stages: the first stage is an alternative between two activities, and the second stage includes two activities that can be "played" in any order. This visualization mode is well suited for scenario designers as it shows in a very structured fashion the composition of the TS. Other visualization modes are possible like the one shown in fig. 9 representing the same scenario of fig. 7; this visualization correspond to a temporal view of the activities included in the TS, and is best suited for other actors as tutors or learners. In any case, edition of the TS is accomplished by simple drag-and-drop operations by accessing a component toolbar, or by context menus attached to the graphical representations of the scenario components.

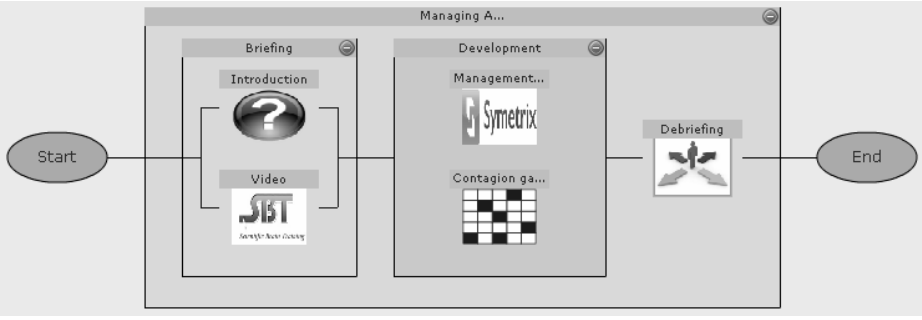


Fig. 9. Temporal graphical representation of a training scenario

Intentions of any kind can be attached to operator nodes. A propagation schema of intentions is established either in a top-down or a bottom-up direction. The former one means that we can attach to a node intentions coming from any of its ancestors; the later one means that if a new intention is attached to a node, that intention is considered automatically as an intention of all of its ancestors.

The right part of the editor shows the library of learning object templates. Templates can be gathered from local or external repositories by the application of special algorithms matching scenario intentions with harvesting metadata. Gathered components are ranked according to the result obtained by application of those algorithms; they can be easily integrated into the scenario skeleton by simple drag-and-drop manipulations.

For gathering purposes we consider the set I of scenario intentions as the union of three disjoint sets, $I = KO \cup GO \cup CV$ where

- KO is the set of nodes from knowledge context ontologies being referenced in the knowledge intentions,
- GO is the set of nodes from the special game context ontology being referenced in the game intentions,
- CV is the set of competency verbs appearing textually in the knowledge intentions

The actual matching algorithm implements a weighted function

$f(T) = w_1 m(KO) + w_2 m(GO) + w_3 m(CV)$ where

- T is a learning object template,
- KO, GO and CV are as explained before,
- w_1, w_2 and w_3 are constants such that $0<w_i<1$ and $w_1+w_2+w_3 = 1$. This constants are the weights of the matching criteria taking into account by the function. Weights denote the relative importance assigned to each criterion.
- m is a function which calculates in what extent the template matches a specific criteria. This is done by looking at particular template LOM tags, such as the <keyword> tag in the <general> LOM section or an <educational-objectif> tag included by us in the <classification> LOM section.

For each criteria the function `m` returns a result between 0 and 10, so the matching function `f` will return a result between 0, meaning that the template does not fill any of the TS expected intentions, and 10, meaning that the template fills all of the TS expected intentions.

The editor also allows the users to save partial and complete scenarios, to create new templates to be added to the library, and to transform the scenario's skeleton model into a partial execution model for which specific platform interpreters can be developed.

5 Current Experimentations

We have technically experimented with success the different phases of our methodology in the context of the Learning Game Factory project (LGF, 2009-2011). The protocol of our experimentation was the following:

- development by each of the five "component providers" of generic "learning game components" able to be harvested, configured, adapted and observed. For example, a component may be a generic "memory game", whose objects' properties, rules and interface can be adapted according to designers' needs;
- integration and indexation of the developed learning components in different repositories managed by each partner;
- in the scenario editor: harvesting of available components according to the design context;
- in the scenario editor: importation and adaptation of selected harvested components by accessing the metadata package that implements the interoperability protocol mentioned in section 2.3;
- at runtime, in a specific player developed for the experimentation by a partner: execution of sequential scenarios with dynamic settings and observation of harvested components.

We are currently working on the specialization of the tool in the context of training devices based on the use of role games [14]. The basic models of section 2.2 have been extended for this purpose.

6 Conclusions and Future Works

We have presented in this paper an iterative “wheel” methodology and a tool implementing this methodology to support the activity of the actors involved in the production of professional online training systems. The methodology is sustained in the enrichment of various models in order to progressively produce more detailed definitions of the training scenarios. Models for the two initial phases, formulation and design of the training scenario, have been introduced and explained. The first one captures the contexts and intentions of the device, as the second one uses a partial-order planning approach to define a structure that satisfies the intentions in the defined contexts.

The methodology is strongly based on the reuse of previously designed components expressed in the form of templates. A protocol has been defined allowing the recovering, filling and adaptation of those templates to new learning situations.

The first results on the use of the developed tool, ScenEdit-II, have shown that our approach was relevant and allowed to design a training scenario from the recovering and adaptation of objects coming from different sources based on the specification of an initial Context/Intention Model. A scenario design platform independent model is generated by the tool; this model is later interpreted by a specific player which executes the scenario.

As one of our future works, we want to refine and complete the models of our methodology following a Model-Driven Engineering (MDE) approach, in particular the execution model. MDE is proposed to overcome the difficulty inherent to the development of complex systems in order to obtain domain specific tools and applications through the clear expression of domain specific concepts [15]. The idea is to build a system through the successive construction of models moving from higher to lower levels of abstraction through well defined transformations. We consider that the task of creating learning scenarios is a complex situation justifying the utilization of this approach, as it is already proposed in other works like [16] and [17].

We want also to experiment with intelligent agents' technologies for assisting the users in the elaboration of entire or partial scenario structures, in particular with common and current planning algorithms. Artificial planners could propose solutions to designers based on various sources of knowledge. For example, a partial-order graph of the intentions could be used by partial-order planners to obtain initial solutions, which could be later refined by the use of knowledge about actors' preferences (tutors, learners, etc.), knowledge about general pedagogical rules, etc.

References

1. Emin, V., Pernin, J.-P., Guéraud, V.: Model and tool to clarify intentions and strategies in learning scenarios design. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 462–476. Springer, Heidelberg (2009)
2. Hotte, R., Godinet, H., Pernin, J.-P.: « Scénariser l'apprentissage, une activité de modélisation. » Numéro Spécial, Revue Internationale des Technologies en Pédagogique Universitaire, Montréal (2008)
3. Martel, C., Vignollet, L., Ferraris, C., Villiot-Leclercq, E.: A design rational of an editor for pedagogical procedures. In: 9th International Conference on Computer Supported Collaborative Learning, vol. 2 International Society of the Learning Sciences (2007)
4. Paquette, G.: Apprentissage sur Internet: des plateformes aux portails d'objets à base de connaissance. In: Pierre, S.(ed.) Innovations et tendances en technologies de formation et d'apprentissage, pp. 1–30. Presses de l'école polytechnique de Montréal (2005)
5. Boehm, B.: A Spiral Model of Software Development and Enhancement. ACM SIGSOFT Software Engineering Notes 11(4), 14–24 (1986)
6. Caillois, R.: Man, Play and Games. Free Press of Glencoe (1961)
7. Mariais, C., Michau, F., Pernin, J.P.: The Use of Game Principles in the Design of Learning Role-Playing Game Scenarios. In: 4th European Conference on Games Based Learning, pp. 181–184. Academic Publishing Limited, Copenhagen (2010)

8. IMS Learning Design Specification, IMS Learning Global Consortium (consulted on April 2011), http://www.imsglobal.org/learningdesign/
9. Koper, R., Tattersall, C.: Learning Design: A Handbook on Modeling and Delivering Networked Education and Training. Springer, Verlag (2005)
10. Wooldridge, M.: Reasoning About Rational Agents. MIT Press, Cambridge (2000)
11. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Englewood Cliffs (2009)
12. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Reading (1995)
13. Learning Technology Standards Committee of the IEEE. Draft Standard for Learning Object Metadata. IEEE, New York (2002), http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf (last consulted on April 2011)
14. Mariais, C., Michau, F., Pernin, J.P., Mandrau, N.: Learning Role-Playing Games: méthodologie et formalisme de description pour l'assistance à la conception. In: Environnements Informatiques pour l'Apprentissage Humain, Mons (2011), (to be published)
15. Schmidt, D.C.: Guest Editor's Introduction: Model-Driven Engineering. Computer 39(2), 25–31 (2006)
16. El-Kechai, H., Choquet, C.: Reusing Pedagogical Scenarios at a Knowledge Level: a Model Driven Approach. In: Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), IEEE Press, NY (2007)
17. Laforcade, P., Choquet, C.: Next Step for Educational Modeling Languages: The Model Driven Engineering and Reengineering Approach. In: Sixth IEEE International Conference on Advanced Learning Technologies (ICALT 2006), pp. 745–747. IEEE Press, NY (2006)

Standardization of Game Based Learning Design

Sebastian Kelle, Roland Klemke, Marion Gruber, and Marcus Specht

Open University of The Netherlands,
Center for Learning Science and Technology,
Valkenburgerweg 177,
Heerlen, The Netherlands
{ske,rkc,mgr,spe}@ou.nl

Abstract. The standardization of the design of learning games is a contradictory topic: The existence of a rich variety of domains and applications is in conflict with the desire for unification that would result in improved reusability, interoperability and reduction of design complexity. In this paper, we describe the use of the ICOPER Reference Model (IRM) specification as foundation layer for the design of digital learning games. This reference model incorporates design and development processes as well as standards such as IMS Learning Design, a framework for presenting content according to logical rules like conditions and properties. The paper reports about exemplary learning games that make use of e-learning standards the IRM consists of, and explains about potential and limitations both from the game and e-learning design perspective, resulting in suggestions how to close missing links.

Keywords: IMS-LD, IRM, ICOPER, Game Based Learning, Standardization, Serious Games.

1 Introduction

Ever since the advent of e-learning, various working groups, committees and bodies have been working on achieving standards and specifications for enhancing quality, interoperability and the reuse of learning contents and designs. Examples for such standardization bodies are CEN, IEEE, ISO, ADL, ANSI, DIN, BSI, and NEN, only to name a few [30].

One of the realities of different standardization bodies creating different standards can be a lot of overhead in coordination. As Duval reports [11], as a consequence of this, one of the key problems in e-learning standardization is the lack of experimental validation of the actual usefulness especially of interoperability standards: They are theoretical constructs that are often of premature value, when it comes to practical application. However, there is still a high interest in common standards, amplified by the fact that large parts of the e-learning market are covered by schools and universities that generally support the exchange and sharing of knowledge across institutional or cultural barriers.

Inspired by the successes of the video gaming industry, as well as a trend in pedagogy, e-learning providers are increasingly incorporating game-based learning

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 518–532, 2011.

approaches. Due to the gaming industry taking the role as technology innovator for learning game incentives, relevant standards are often of a proprietary nature and closely tied to particular pieces of hardware, e.g. game consoles and game controllers. As a consequence of these marketing strategies that seek to preserve unique selling points, digital learning games go with a diversity of formats and file types, involving many different sub-standards relating to technology, content, and subcategories thereof. Nevertheless, similar to other e-learning formats, a digital learning game requires learning goals, learning contents, trajectories through the learning contents, and a structural framework that ties together all these components.

Therefore it seems plausible that game-based learning could benefit from existing work on e-learning standards. In this paper we will explore how e-learning standards could play a role in aligning the different elements that make up a digital learning game. We will analyze a recently developed reference model (the so-called ICOPER Reference Model) that was created from best practice experiences in e-learning for its potential to be used as conceptual framework for the design of learning games.

2 Problem Analysis

Various standards exist in the fields of e-learning and game design, however, little work has been done to connect both fields. With respect to the e-learning part of our scope, recently a big effort has been undertaken to find a coherent model that unites technical and conceptual standards available for the design of technology enhanced learning solutions: The ICOPER Reference Model [29].

In this problem analysis, we will first describe the current situation of learning- and game design standards that are most relevant for interoperability, reusability and reduction of design complexity. Then we will cover the combined perspective of learning game design and point out some problematic aspects that result from the lack of bilateral standards.

2.1 Standards in Game Design

In digital gaming, technical standards have a high relevance that even can be of reciprocal character because many commercial games take the innovation role for technology, spearheading the latest developments and "setting" new standards at a fast pace. These modern technology standards encompass multimedia technologies for input, audio and (3D) graphics and are manifested as "game engines" that serve as mostly proprietary production models in professional game design and development. Examples are the DirectX standard [36], Microsoft XNA [40] for developing Xbox console games, the "Vision Game Engine" [42] for developing multi-platform games, and as final example the CryEngine [35] for developing videogames with the highest cinematic realism of what is possible today. These standards are technical standards, rather than design standards, but in gaming it is often difficult to differentiate between the design and implementation, therefore these "engines" come with documentation on how to design and develop games for them.

Modern digital games also tend to more and more make use of network features and provide added functionality by connecting to the internet, which requires the

inclusion of a stack of telecommunication standards in the implementation. Already in 1984, Crawford [7] mentioned the possible "connection of computer games over phone lines" as distinctive advantage of computer games over classic games. In his design methodology for computer games he describes a sequence that ranges through the initial choice of goal and topic, a preparation phase in which some research and brainstorming is needed, a structural design phase that has to be evaluated (falling back on the previous phase iteratively), and finally a programming, testing and post mortem phase. The reason for the long-lasting acceptance is that this design method resembles the most widely used software engineering models and has definitions that are sufficiently wide to leave interpretation space for the application on many different types of games. Although the creation of games relies on technical and structured software engineering methodologies, the creative aspect of the design process appears mystifying: according to Adams [1], the idea creation at the early stage of the game design process is more an artistic than an engineering process.

Salen and Zimmerman [26] have compiled a detailed description of important factors to consider for meaningful game design. They promote a systemic approach that frames a game inside a formal, experiential and cultural system that range from closed to open. In their compendium, one of the core elements of game design is the definition of game rules, which create the "game system" structurally. Rules of a game are categorized according to "constitutive", "operational" and "implicit" rules, which can be interpreted corresponding to a scale from "prescriptive" to "own choice". Also the game play as such is equally important, as it is forming the experiential parts of the system. According to Salen and Zimmerman "a game designer only indirectly designs the player's experience, by directly designing the rules". [p. 327].

An example for a game approach that makes use of "implicit" or "own choice" rules is interactive story-telling, which is found in many (especially adventure-) games. Due to the experiential nature of such games it is an approach that is often found in learning games. One of the concrete examples for such an approach has existed in a niche until the eighties, and only had some publicity in more recent times: Interactive Fiction. As described for example by Donikian and Portugal [10], this medium abolishes the difference between author, spectator, actor and character, and creates a big potential for immersion, due to identification with a role and ownership of influence on a non-linear story sequence. The technology supporting this has been evolving for decades from simplistic single-user text adventure approaches up until now where there are authoring systems (e.g. Inform 7) [38] that understand natural language. The output files are usually in a system independent package format called "BLORB" [34], which is interpretable by web-based engines (e.g. Glulx) that boast the power to render a fully-fledged multi-user adventure game to be played in a browser [24].

Another, more general effort of standardizing game design can be found in the use of game design patterns, which preserve knowledge about building elements of games and give information on how to implement them. The approach is described semi-formally by Kreimeier [17] who uses "Alexandrian" proxy patterns consisting of a problem description the pattern is going to deal with, a solution description, consequence description and examples. Björk and Holopainen [2] collected a large fundus of game design patterns, which extends the relatively informal approach of

Kreimeier onto a more detailed and complete level. There, they describe the use of game elements, and their connection possibilities with other elements, indicating links like manipulation, instantiation as well as conflicts between elements (for example: a real-time gameplay pattern cannot be combined with a turn-based gameplay pattern). This approach is also endorsed by Westera et al. [33] who stress the usefulness of patterns with respect to the reduction of design complexity.

2.2 Standards in E-Learning Design, United in the ICOPER Reference Model

With respect to e-learning, Cooper and Kraan [6] point out that standards in e-learning are important because they can avoid that information gets "locked into a supplier's product", and are able to join up different systems because they use the same data backend. While this is particular relevant for interoperability, standardization in e-learning also yields other desirable outcomes, like reusability and reduction of design complexity.

In e-learning there are big efforts to create standards that make learning content transferable between technical platforms and educational scenarios. The ICOPER project [37], funded by the European E-Content-Plus program, reflects such a standardization effort on a meta-level, analyzing different standards on different levels and their interoperability. At the core of the project resides the ICOPER Reference Model (IRM) [29] that is based on a rich pool of best practice examples for the successful use of each substandard that concerns technology enhanced learning.

It embraces all relevant standards available including content related standards, user modeling standards, interoperability standards as well as process oriented standards. The IRM, in its purpose to agglomerate various e-learning standards into a functional concept, shows promising directions, because it helps avoid the hazard of using standards that overlap and cause redundancies, as well as conflicting standards. In this paper the ICOPER Reference Model is chosen as the starting of the analysis.

Table 1. The e-learning standards used in the IRM

Standard	Description
RCD / LOD	Reusable Competency Definitions / ICOPER Learning Outcome Definition (LOD): LOD is an application profile based on RCD, a data model that
PALO	Personal Achieved Learning Outcome profiles
LOM	Learning Object Metadata, a standard to describe metadata for learning objects
OAI-PMH	Open Archive Initiative's Protocol for Metadata Harvesting, a protocol specifying the harvesting of metadata of learning objects in repositories
IMS-LD	IMS Learning Design, a standard for sequencing content according to logic (e.g. adaptive) rules, as well as user roles
IMS-QTI	Question & Test Interoperability format, defining a data format for online assessments

With the help of these standards the main components of the IRM are formulated:

- The Domain Model
- Process Models
- Service Descriptions
- Data Schemes

The domain model consists of high-level learning context scenarios, which are drawn from institutional, corporate, professional and re-skilling training practices. The domain model is developed around key concepts such as learning outcome, learning design (including teaching method), learning content, learning opportunities and assessment.

These need to be matched for the respective purpose of each learning scenario and therefore are more of exemplary value. The domain model thus becomes a context-based scaffolding for designing the necessary processes and entities defined in the IRM so that they fit the domain or context.

Also, the IRM covers instances of process models serving learners, learning facilitators, and other stakeholders in the delivery of outcome-oriented teaching. In addition, the IRM contains service descriptions for search and retrieval, publication services, user management services, recommendation services, harvesting services, registry services, and validation services. Finally, data schemes are given for providing the relevant technological frameworks for storing dynamic data, schemas for personal achieved learning outcomes (PALO) and learning designs to be included on the backend side. The data model of the IRM was prototypically implemented by the ICOPER project in the Open Icoper Content Space (OICS); it covers a recommendation how to create a competence map for learning outcomes, an incremental model for different layers to create learning processes, a concept model, a domain model and, finally, a process iteration model on how to design IRM based solutions.

2.3 Standardization in Game-Based Learning Design

For making digital games that work for learning purposes, both aspects of gaming and learning and the standards relevant to them have to be combined. According to Ebner and Holzinger [12], there are important advantages in standardization of technology such as compatibility, transferability and reusability. Also, there are advantages like social benefits, enabling standardized jargon to efficiently communicate among specialists of a specific subject. Disadvantages can be found mentioned in reduction of variety, retard of innovation as well as “excess inertia”, which Farell [13] describes as the impediment of “switching from one standard or technology to a possibly superior standard or technology”. One of the key reasons for dealing with standardization is that the creation of learning games is a very costly enterprise, as reported for example by Van Eck [32] and Moreno-Ger et al. [22]. Each time a learning game is developed it requires a hand-tailored design and implementation from scratch. As possible solution to the problem they discuss the repurposing of commercial “off the shelf” (COTS) games for learning. Although such repurposing is easier said than done, it can save a lot of design and implementation effort. A concrete

example is described by Gee and Hayes [14] in which the role-playing game "The Sims" is adapted for collaborative learning purposes.

Still, the use of standardization in game-based learning has controversial aspects. Besides the possible hazard of reducing variety, Squire [31] points out that in the information age certain fundamental principles of economic reality have changed since the dusk of industrial age: Conformity has been replaced with diversity, compliance with initiative and standardization with customization.

Therefore, standardization may come at the cost of customization and other advantages that are related to flexibility of design methods, content, user interaction and other factors. This also has consequences for gaming: The reduction of flexibility might reduce motivation, fun and the possibility for immersion in game play, which is fundamental to the success of a game.

Relating more specifically to the topic of game-based learning, however, the situation of standardization is more on a taxonomical level. The Serious Games organization [27] as well as Breuer and Bente [4] have made the effort to pool together a taxonomy for serious games, in which a wide scope of different categories are listed. Unfortunately, this does not include any technical standards or recommendations on how to design or implement serious games.

Using a classification taxonomy is nevertheless a starting point to get an overview what different types of serious games exist and what are examples. According to Breuer and Bente serious games can be classified according to platform, subject matter, learning goals, learning principles, target audience, interaction modes, application area, controls/interfaces and common gaming labels (puzzle, quiz, etc.). This can help to inspire a learning-game designer to consider all options during the early stages of the design process.

The situation of standardization in learning games seems not very systematically developed but that does not mean that there are no working examples. When it comes to the implementation of a digital learning game, as reported by Livingstone and Hollins [18], various technical standards for gaming can be put to use, such as different standards in 3D technologies (for instance, VRML, X3D, COLLADA, OpenGL and WebGL), browser languages and also different kinds of multimedia standards like flash or, more lately, HTML5, for example for the use in mobile devices.

Interactive storytelling has a specific relevance to the design of learning games, and the IMS-LD standard has been shown to have this potential [25]. This can be done by creating conditions that rely on an extended propositional logic control (also known as IMS-LD Level B) which fire upon certain user behavior. For example: if the user behaves in a certain way, the system may detect that and react adaptively by rearranging the order of content consumption. In [16] Gruber et al. describe how IMS-LD is used successfully to present an interactive course on architecture making use of such adaptive content sequencing. This enables a certain degree of "free movement" of a learner inside a coherent structure, which incorporates the challenge to solve a quest in order to advance on different paths through the learning content. In line with the principle of Open Information Access, this "free movement" can be interpreted as educational pattern that can be found in adaptive storytelling, as well as constructivist learning. Likewise, the IMS QTI specification would allow for quiz-like approaches of game-based learning [15].

This demonstrates that e-learning standards for adaptiveness and assessment have a potential to enrich game designs with functionality that is relevant for learning. In the concrete example of IMS-LD used here with the "Recourse" authoring tool, however, some limitations were detected, for example that in practical application repeating a certain activity was not possible, once a "unit of learning" had started [16].

While this is a concrete example of a learning standard that doesn't quite live up to its theoretical power, on a more general note, in the design process the initial choice of one of the available standards is highly speculative, and there is no sound argumentation to know up-front which e-learning standard is appropriate for what learning game purpose. There may be some flexibility in choice, but not every e-learning standard is going to be of equal usefulness to the design of learning games, due to different requirements. To tackle this we need a more refined approach that helps to streamline design routines without omitting to consider important standardization methods.

3 Using the ICOPER Reference Model as Bridge between Gaming and Learning

The fundamental assumption in this paper is that parts of the ICOPER Reference Model can be used to build the bridge between gaming and learning. One of the key questions about building this bridge, is how the existing ICOPER Reference Model can be exploited for use in game-based learning design, where are possible gaps, and resulting thereof, how the existing IRM can be extended.

In addition to an overview of the status quo on existing approaches we found evidence about (i.e. learning games that use e-learning standards that appear in the IRM), we also will report about our own experiences and what we could learn from them. Finally, we will give recommendations on how to use the IRM for the design of learning games.

3.1 E-Learning Standards Applied in Games

In this section, we will shortly revisit the literature on existing learning games that have been using e-learning standards. As mentioned above, there are only few examples of learning games that make use of e-learning standards. The way IMS-LD theoretically works for the use in adaptive game-based units of learning is described by Burgos et al. [5]. They explain how the Level A and B of IMS-LD can be mainly used for creating the adaptivity of content presentation: Level A consists of user-modeling (users, roles) and content related components (environments, resources, links, activities). Level B consists of logical properties, conditions, calculations, global control elements and monitoring services. The architecture that enables the game-based learning approach relies on a proxy layer for communication between "game activities" (gamelets that correspond to interactive content elements) and the learning flow.

As a practical example, of this approach Moreno-Ger et al. have created an adaptive game [21] using IMS-LD as control framework for a video game on chocolate making. In this example, the SLED-player environment [41] (a tomcat-based web server module

that interprets the XML-based units of learning that are the output of IMS-LD based authoring tools) works as an aggregator for the game content, while providing logical properties and conditions that steer the sequence of the game content, as well as user roles. It is, therefore, an example for the learning process controlling and triggering gaming elements.

For use in a virtual world environment, Livingstone and Hollins [18] explain how interoperability can be achieved between learning management systems (LMS) and Second Life, a massively multiplayer online role play game that had its zenith in 2007, which remains of interest for experimental use of 3D game learning environment research. The design here consists mainly of a proxy layer between Moodle and the virtual world, which enables communication by means of http requests and XML-RPC calls (roughly: a standard that does the opposite of http requests: sending instructions that are executed on a web server), thus providing the linkage between dynamic objects in the virtual world and the LMS. Since most of the interaction happens inside the "game" world, this is an example for the gaming side taking control over the learning process; however, the communication layer puts the two aspects in balance and enables activities in both directions.

Another approach by Minović et al. [19], [20] describes the use of a meta-model for educational games (called the "educational game meta-model") that is based on knowledge modeling theory. The proposed model makes use of a series of information channels that enable communication between Knowledge Objects (interpreted here as Learning Objects) and the actual game components. On the implementation level, the approach is realized as XSLT-based web client, providing in this example a game-authoring environment and game client presenting an online adventure game in the domain of geography. Despite the promising direction of using open technical standards for creating the meta-model as layer between learning objects and functional game parts, the system is closed in itself, and the aspect of reusable content is missing. Since both authoring and gameplay happens within the "game" prototype, the learning flow is influenced and controlled by the game component.

Del Blanco et al. [8] use a virtual learning environment game, based on SCORM, forming a connection to a Moodle LMS in the background. Similar to the approach described by Livingstone and Hollins [18], the game aspect takes the role of steering the occurrence of learning objects. In another approach by Del Blanco et. al. [9], the LAMS (Learning Activity Management System, [39]) environment was used, encapsulating video gamelets in a quasi-IMS-LD logic, here the "IMS-LD" part was enabling that the LMS took over the sequencing of the game-based content.

Also, Börner [3] describes a Flash based learning game that makes use of the SCORM standard to structure the learning content of the game. For multi-user aspects, a distributed server architecture was used. The design is strongly dependent of the overall learning trajectory; therefore, in this case the learning process takes control over the game sequence.

3.2 Own Experiences

Schmitz and Klemke [28] report on the design of the SPITKOM learning game using e-learning standards that are found in the ICOPER Refernce Model (IRM).

The SPITKOM project is a game-based learning approach to train for the European Computer Driver's License ECDL. It forms an example, how the IRM is used as "slave-standard". The main process is driven by hard-coded game logic which uses the learning outcomes, learning contents, assessments and personal achievement profiles that are explicitly modeled and stored in the Open ICOPER Content Space (OICS). Reusability in SPITKOM can thus be achieved mainly on the level of learning content, thus it would be straightforward to reuse the approach in a different content domain (by exchanging the domain model). However, since the game-logic is hard-coded, it would be difficult to create reuse with a different game-logic – the game component would have to be exchanged with a new one.

Another approach is using the IRM as "master-standard", making use primarily of IMS-LD to design the structure. An adaptive learning game on a quiz-like basis was developed for the training of first aid and basic life support. The basic procedure was, similar to the approach described in [5] and [21], to use IMS-LD Level B for creating the control structure of the adaptive story-telling used in the game. In this case, the domain model was fixed, but the control structure could be easily adapted. For practical reasons the implementation of the game-based learning design we used the Emergo toolkit [23] that provides a similar expressiveness as IMS-LD and the same range of functionality we were interested in. A screenshot is presented in figure 1.

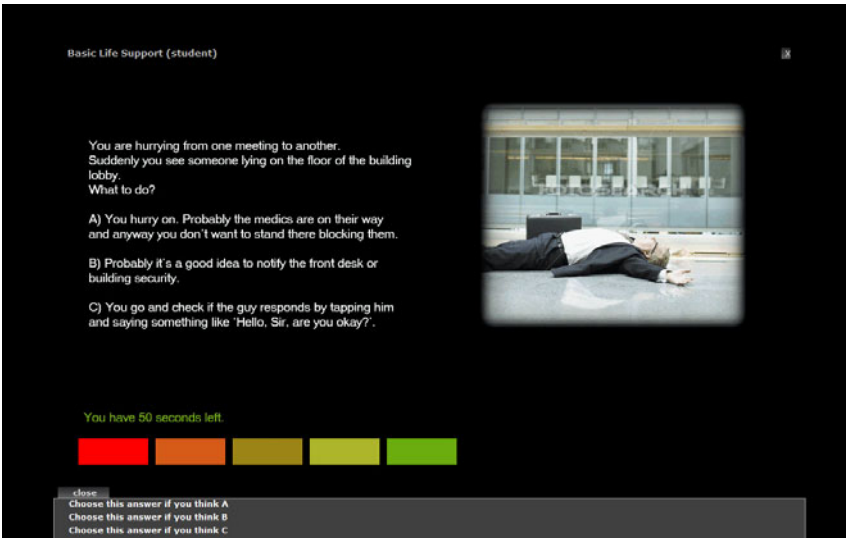


Fig. 1. The Basic Life Support training game, using EMERGO game platform

The resulting prototype was used for performing experiments on the effects of different game design patterns on learning.

These practical examples allow for a comparison. While the SPITKOM approach seemed to satisfy the preferences of gamers, the missing part was the explicit teaching method and the corresponding learning design. This was done for the benefit of a

specialized and more game-like user interface. It demonstrates that the requirement of a challenging game experience conflicts with the pedagogical requirements because there was little flexibility regarding modification requests.

In this area the IMS-LD based approach was more flexible, because there, both content and game logic can quickly be altered according to changing learning outcome definitions or learner profiles. Another advantage is that IMS-LD has the potential to use external learning material and, hence, be linked with the service architecture provided by the OICS.

It appears that e-learning driven examples for the design of game-based learning could have the disadvantage more likely to disappoint learners that expect a fully-fledged gaming experience, because they adhere to e-learning standards from the beginning, resulting in a shortcoming on the game-like behavior and feeling of the result.

This leads us to a missing link between game standards and learning standards.

3.3 Duality between Gaming and the E-Learning Design

The two different approaches reflect different design methodologies (start the design cycle from the gaming or the e-learning standards perspective). These approaches match with what we have been trying in practice. Starting the design from the side of learning, it is possible to model the educational process and then iteratively integrate game elements into the instructional design. From the game perspective, the methodology links game elements with learning activities and outcomes.

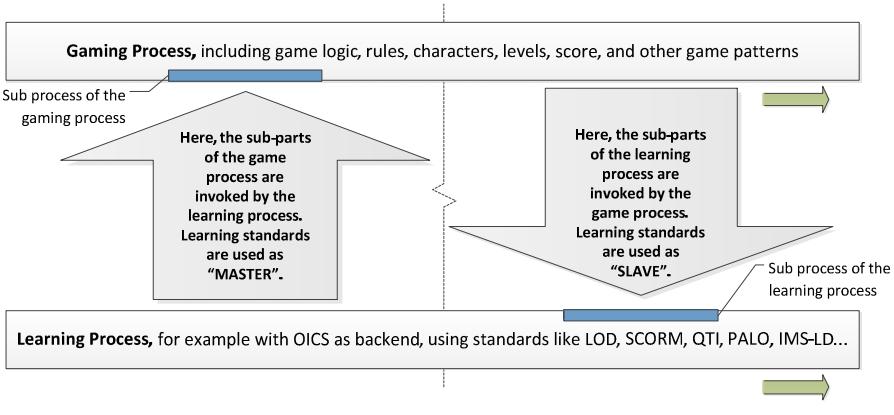


Fig. 2. The "Master" usage of e-learning standards is applied in the Basic Life Support game prototype, while SPITKOM uses the e-learning standards as "Slave" model (in this case: the OICS). The ideal situation would be to have both directions in one learning game.

The result of the two different approaches, i.e. using e-learning standards as "master" and "slave" model (figure 2), proved two main disadvantages. In the case of the SPITKOM game, the problem is that the game component is difficult to adapt and hence provides difficulties for reusability. Also the IRM/OICS needs to "satisfy" the game

requirements, which poses the encounter of rigidness with respect to interoperability questions.

The other approach, i.e. using e-learning standards to start out from, poses the repurposing of e-learning based frameworks for gaming, which turned out to have limitations with the respect to making a learning game that actually has the properties and "feel" of a real game.

Table 2. How the discussed games make use of e-learning standards

Approach	**makes use of what e-learning standard**	**Corresponds to the use of learning standards as**	**Remarks**
Moreno-Ger's Game [22]	IMS-LD	"master"	
Livingstone and Hollins' 3D game concept [18]	SCORM	"slave"	work in progress
Minović's game	Learning Objects	"slave"	
Börner's game [3]	SCORM	"master"	
del Blanco's e-adventure games [8], [9]	SCORM [8], IMS-LD [9]	"slave" [8], "master" [9]	The game design in [9] was done in IMS-LD but the implementation was using LAMS.
SPITKOM game [28]	LOD, SCORM and QTI	"slave"	
Basic Life Support game	IMS-LD	"master"	The game design was done in IMS-LD but the implementation was using EMERGO

In table 2 we summarize how the described approaches make use of e-learning standards. This is concluded by the way the design approach is described in the corresponding literature, starting out by first designing the game component or starting the design with consideration of e-learning standards, which is reflected in the system architecture as described in 3.1. It becomes visible that the list of learning games that use IRM-conform e-learning standards is indeed quite short. This indicates there is still a large gap between e-learning standards and learning game design.

4 Discussion

Although there has been proof-of-concept for different applications making use of Learning Objects, SCORM and IMS-LD for game-based learning, a more holistic design approach is desirable, especially when considering the full spectrum the IRM offers with respect to meeting requirements for learning. As a possibility, the Game and Learning aspect could be serialized in the design process, where a first idea and draft concept of a learning game is followed by the formalization of a domain model, which serves as construction scaffold and requirement specification for the remaining parts.

One suggestion is that the definition of game rules could be complemented with learning outcome definitions (LOD) on the learning side as well as control structures defined in IMS-LD. Correspondingly, the design of game play (as a consequence of the game rules) on the learning side are matched with teaching methods and learning design is reflected in adaptive content modules (e.g. SCORM). Scoring may be realized with assessment (QTI) elements.

Finally, the technical requirements engineering as well as implementation of the game design are reflected on the learning side with the content repository Open ICOPER Content Space (OICS), which forms a backend, provisioning content and metadata to fill external learning services with learning material. In addition it offers an assessment infrastructure and user modeling framework, making it a backend for learning management systems; and all the elements mentioned before as being relevant to the design process of learning games (LOD, PALO, QTI, etc.) are stored here. The OICS can therefore build instances of entire domain models, and, when matched with requirements for gaming, a game domain model. This means that it supports instantiating the domain model for game-based learning, so that it helps the design. Although there are still some unsolved issues regarding the implementation part of IMS-LD, with LAMS and EMERGO there exist practically usable authoring and deployment environments that are using virtually the same notation and functionality as IMS-LD. For both there exist working examples of learning games. Well noted, IMS-LD has its primary power to integrate diverse learning activities into Learning Management Systems and sequence them logically. Coming from the gaming side (to integrate learning processes into a game design), other approaches are more sensible such as using Learning Outcome Definitions.

The notion of game design patterns can also be reflected in the IRM by formalizing the more structural type of these patterns (such as storytelling and game-sequencing patterns) for example in BPMN notation, hence, providing important guidelines for the rule set of a game and, on the technical side, the capability to be translated into implementation stubs. In this respect much can still be learnt from the example of Interactive Fiction we mentioned, which is using its own standards that are not (yet) covered by the IRM.

5 Conclusions and Outlook

In this paper e-learning standards were analyzed for their appropriateness for game-based-learning. It is concluded that there are some issues, but also a lot of potential.

While the game industry undoubtedly has a wide variety of de-facto standards for designing games, these often lie hidden behind the walls of large corporations that have to protect their assets. On the e-learning side, it was easier to find openly documented standards relevant to design, more of the type of "de-jure" standards (see [30] for de-facto/de-jure discrepancy).

The synthesis of both gaming and learning can be considered from the game perspective, where the game logic or story components trigger learning processes, or, vice versa, from the learning perspective, where learning control structures define the gaming elements. By analyzing the IRM, missing links were identified between gaming and e-learning. While the game-driven perspective produced more convincing results regarding the user experience, the learning-driven perspective had advantages regarding reusability. The outcome of this observation is that there needs to be more harmonization between game design and e-learning design, for example a technical solution that makes it possible to use IMS-LD directly without encountering limitations as described in [16]. Vice versa, the IRM could profit from the incorporation of standards derived from game design, such as structural game design patterns that encapsulate practical experience of successful learning games, hence contributing to a corresponding domain model.

Overall, the IRM, in its purpose to agglomerate various e-learning standards into a functional concept, shows promising directions, because it helps avoid the hazard of using standards that overlap and cause redundancies, as well as conflicting standards. However, there still needs to be work done for finding a suitable domain model to be instantiated in the IRM for the use of game-based learning. To get suitable findings for this, future research needs to include a more extended testing of available e-learning standards for the use of gaming while continued work on interoperability standards is needed on the technical side, a direction the creation of the OICS points us into.

Acknowledgements

We wish to express our gratitude to the ICOPER project, as well as in particular our partners at CELSTEC, HUMANCE and University of Jyväskylä, who were all a great help for inspiration of this paper.

References

1. Adams, E.: Fundamentals of Game Design. New Riders (2009)
2. Björk, S., Holopainen, J.: Patterns In Game Design. Charles River Media (2004)
3. Börner, D.: Integration von Spiel-basierten Lernprogrammen in SCORM-konformen Lernumgebungen unter besonderer Beachtung des Multi-User Aspektes (2007)
4. Breuer, J.S., Bente, G.: Why so serious? On the relation of serious games and learning. Eludamos. Journal for Computer Game Culture 4, 7 (2010)
5. Burgos, D., Moreno-Ger, P., Sierra, J.L., Fernandez-Manjon, B.: Authoring game-based adaptive units of learning with IMS Learning Design and e-Adventure. International Journal of Learning Technology 3, 252–268 (2007)

6. Cooper, A., Kraan, W.: Assessing the business case for standards: Introduction for strategy planning and resourcing committees (2009), http://www.jisc.ac.uk/publications/briefingpapers/2009/bpbusinesscaseforstandards.aspx
7. Crawford, C.: The art of computer game design. Osborne/McGraw-Hill, New York (1984)
8. del Blanco, Á., Torrente, J., Moreno-Ger, P., Fernández-Manjón, B.: Integrating Adaptive Games in Student-Centered Virtual Learning Environments
9. del Blanco, Á., Torrente, J., Moreno-Ger, P., Fernández-Manjón, B.: Towards the Generalization of Game-Based Learning: Integrating Educational Video Games in LAMS. In: Proceedings of the 2010 10th IEEE International Conference on Advanced Learning Technologies, pp. 644–648 (2010)
10. Donikian, S., Portugal, J.N.: Writing interactive fiction scenarii with dramachina. Technologies for Interactive Digital Storytelling and Entertainment, 101–112 (2004)
11. Duval, E.: Learning technology standardization: making sense of it all. International Journal on Computer Science and Information Systems 1, 33–43 (2004)
12. Ebner, M., Holzinger, A.: Successful implementation of user-centered game based learning in higher education: An example from civil engineering. Computers & Education 49, 873–890 (2007)
13. Farrell, J., Saloner, G.: Standardization, compatibility, and innovation. The RAND Journal of Economics 16, 70–83 (1985)
14. Gee, J.P., Hayes, E.R.: Women and Gaming: The Sims and 21st Century Learning. Palgrave Macmillan, Basingstoke (2010)
15. Grant, J.: Learning needs assessment: assessing the need. BMJ 324, 156–159 (2002)
16. Gruber, M., Glahn, C., Specht, M., Koper, R.: Orchestrating Learning using Adaptive Educational Designs in IMS Learning Design. Sustaining TEL: From Innovation to Learning and Practice, 123–138 (2010)
17. Kreimeier, B.: The case for game design patterns (2002)
18. Livingstone, D., Hollins, P.: Virtual Worlds, Standards and Interoperability. International Journal of IT Standards and Standardization Research 8, 45–59 (2010)
19. Minovic, M., Milovanovic, M., Starcevic, D., Jovanovic, M.: Learning objects in educational games. International Journal of Technology Enhanced Learning 2, 336–346 (2010)
20. Minović, M., Milovanović, M., Starcevic, D., Jovanović, M.: Knowledge Modeling for Educational Games. Visioning and Engineering the Knowledge Society. A Web Science Perspective, 156–165 (2009)
21. Moreno-Ger, P., Burgos, D., Sierra, J., Manjón, B.: A Game-Based Adaptive Unit of Learning with IMS Learning Design and Creating New Learning Experiences on a Global Scale, 247–261 (2007)
22. Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J.L., Fernández-Manjón, B.: Educational game design for online education. Computers in Human Behavior 24, 2530–2540 (2008)
23. Nadolski, R.J., Hummel, H.G., Van Den Brink, H.J., Hoefakker, R.E., Slootmaker, A., Kurvers, H.J., Storm, J.: EMERGO: A methodology and toolkit for developing serious games in higher education. Simulation and Gaming 39, 338–352 (2008)
24. Nelson, G.: Afterword: Five Years Later. In: Jackson-Mead, K., Wheeler, J.R. (eds.) IF Theory Reader, pp. 189–202 (2011)
25. Richards, G.: Designing educational games. In: Koper, R., Tattersall, C. (eds.) Learning Design, pp. 227–237. Springer, Heidelberg (2005)

26. Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. MIT Press, Cambridge (2003)
27. Sawyer, B., Smith, P.: Serious Games Taxonomy,Serious Games Summit 2009, San Francisco (2008)
28. Schmitz, B., Czauderna, A., Klemke, R., Specht, M.: Game Based Learning for Computer Science Education. In: Proceedings of Computer Science Education Research Conference, Heerlen, The Netherlands, pp. 81–88. ACM Press, New York (2011)
29. Simon, B., Pulkkinen, M.: ICOPER Reference Model Specification Draft (2010)
30. Sloep, P.: Learning Technology standardization (2002)
31. Squire, K.: Game-based learning: Present and future state of the field. Masie Center E-learning Consortium (2005)
32. Van Eck, R.: Digital game-based learning: It's not just the digital natives who are restless. Educause Review 41, 16 (2006)
33. Westera, W., Nadolski, R.J., Hummel, H.G.K., Wopereis, I.: Serious games for higher education: a framework for reducing design complexity. Journal of Computer Assisted Learning 24, 420–432 (2008)
34. Blorb: An IF Resource Collection Format Standard, http://www.eblong.com/zarf/blorb/blorb.html
35. CryENGINE - Crytek Video Game Developer Company, http://www.crytek.com/cryengine
36. DirectX Developer Center, http://msdn.microsoft.com/en-us/directx/default.aspx
37. ICOPER - Interoperable Content for Performance in a Competency-driven Society, http://www.icoper.org/
38. Inform 7 Web Site, http://inform7.com/
39. LAMS: Learning Activity Management System, http://www.lamsinternational.com/
40. Microsoft, X.N.A.: Creators Club, http://create.msdn.com/en-US/
41. Sled - Bazaar Wiki, http://wiki.bazaar.org/index.php/Sled
42. Vision Game Engine - Workflow Versatility, http://www.trinigy.net/en/products/vision-engine/tools-and-workflow

Simplified Workflow Representation of IMS Learning Design*

Juan C. Vidal, Manuel Lama, and Alberto Bugarín

Department of Electronics and Computer Science
University of Santiago de Compostela, Spain
{juan.vidal,manuel.lama,alberto.bugarin.diz}@usc.es

Abstract. A way of describing the teaching, learning interactions and activities is through Educational Modeling Languages. These languages are formalized so an interpreter can automatically coordinate the activities. However, like for example IMS LD, these languages are usually difficult to understand causing that instructors do not always understand the way the designed unit of learning is coordinated. In this paper we present a graphical notation that simplifies the authoring process of new and preexisting units of learning specified in IMS Learning Design. The graphical notation extends the YAWL language and approach the design of units of learning through a set of patterns distributed in three hierarchical layer.

Keywords: Adaptive Learning, Visual Language, Learning Design, Workflows, Petri nets.

1 Introduction

In recent years, a significant effort has been invested in the definition of Educational Modelling Languages (EMLs) to facilitate the design of a course *from a pedagogic point of view*. One of the results was the IMS Learning Design (IMS LD) specification [1] that has emerged as the *de facto* standard for the representation of learning designs. From the authoring point of view, IMS LD is structurally complex and designers have difficulties to create and understand their units of learning. To deal with this issue an important effort has been devoted to the definition of visual languages that facilitate instructors the design of their units of learning (UoL) [2,3,4,5,6].

From these approaches those that represent the learning flow of IMS LD like a workflow are particularly interesting. These approaches usually coordinate the learning activities by means of a set of workflow-based control structures [7,8,9] with the aim of abstracting the edition of an IMS LD unit of learning. However, these approaches fail in not considering the learning flow of the IMS LD

* Authors wish to thank the Xunta de Galicia and the Ministerio de Educación y Ciencia for their financial support under the projects 09SIN065E and TSI2007-65677-C02-02.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 533–546, 2011.

specification like a hierarchical workflow, which is composed of a set of layers: the first layer is related to the method structure (that is, the plays); the second layer defines the acts that represent each of the plays; and the last layer models the coordination of the learning activities. With this approach, the one we followed in this paper, the creation of IMS LD units of learning consist in building a complex model from simple workflows, facilitating to users a complete view of the learning flow of a UoL.

In this paper we present a graphical user interface that provides an alternative to the design of UoLs that follow the IMS LD specification. In this GUI we address the design of UoLs through a visual model that has two objectives. On the one hand, we pretend to simplify the creation of UoLs through a set of visual patterns with a limited interaction with the user. Visual languages provide an important advantage over text-based: they are more legible. For example, it is much easier to understand a course described by a visual language than a course specified in the XML format of IMS LD.

On the other hand, we want to give to instructors a complete view of the structural characteristics of the course they are defining. Thus, the main workspace of the GUI shows the whole view of the IMS LD learning flow of a UoL. One of the main problems of IMS LD is its difficult interpretation: instructors must be experts in IMS LD to design a course. This is due to the complex structure (based on the theater metaphor) that IMS LD implements to coordinate students with the activities they must perform, which is not always well understood by instructors. For this reason, a visual representation is a major help to understand how the course is from a structural point of view.

Taking into account that a UoL can be viewed as a flow of learning where students and instructors coordinate the activities they have to perform, in [10] we presented OPENET4LD, a Petri nets-based player for UoLs specified in IMS LD. Petri nets are one of the most popular languages for modelling workflows [11,12] because they are a mathematical formalism and, furthermore, they have a graphical representation. They are also a very expressive language that can represent most of the workflow patterns that exist today [13]. However, as seen in [10], models constructed by these networks can become large and quite complex to understand. Since one of our goals is to facilitate the design of UoLs by people without experience in the modelling of workflows, we decided to use a language with a simpler visual syntax than the one provided by Petri nets. The language chosen was YAWL (Yet Another Workflow Language) [14], an extension of the Petri nets language, which provides a bunch of new connectors that represent the typical control structures of workflows. For example, it provides specific structures for selecting, splitting, synchronizing or merging of the execution thread. In a certain sense, the relationship of YAWL with Petri nets is the same as the one defined between third-generation programming languages with assembly languages, that is, they provide a higher level of abstraction that hides complexity to the designer.

Based on the syntax defined by YAWL, in this article we will describe the visual models we use to represent IMS LD and its mappings with the models

of Petri nets defined in OPENET4LD, which ultimately will confer them their operational semantics. Finally, we will describe the implementation of the GUI and how is its integration with OPENET4LD.

2 Visual Model for the Representation of Units of Learning

In this section we detail the visual models that simplifies the way in which instructors design their UoLs based on the IMS LD specification. These models or patterns are described using the visual syntax of YAWL, combined with the properties of hierarchical nets, to so facilitate their representation. Unlike Petri net-based models of OPENET4LD, that represent the complete execution of the UoL, the models described in this section are focused only on the structural properties of IMS LD. In this sense, the simpler the visual model, the easier to use. Therefore, our visual models will not control how to stop, suspend or resume the methods, plays, acts or activities of IMS LD. These details will be transparent to users and, as we will discuss in Section 3, will be provided automatically by OPENET4LD.

The model we present in this section is structured in a hierarchy of three layers. The first layer deals with methods modeling, the second with plays, while the third applies to the execution entities, that is, to activities, structure of activities, and units of learning. For each of these layers, we capture the operational semantics of the element, such as it is defined by IMS LD, and transform it into a YAWL model.

Finally, we must remark that in this article we only will describe the features of our models for the level A of IMS LD. In this level, IMS LD defines methods, plays, acts and activities, and the structural properties of these elements, which is the purpose of this article. However, we also plan to offer level B support through the GUI.

2.1 Representation of Methods

A method implies the execution of a set of plays, although some of these plays are optional, and IMS LD does not impose any requirement on the way these plays can be executed. Figure 1(a) introduces the pattern we use to capture the control that the method establishes over its plays. In this example, the figure shows the execution of three plays, although this model can be extended to represent the parallel execution of n plays. We must emphasize several components of this figure:

- A start place of the method. The circle with a triangle represents the point where the method execution begins.
- A point where the execution of all the plays are parallelized. This point is represented by the AND-SPLIT pattern of YAWL. However, the semantics of this pattern is little different since the execution of the parallel structure will not synchronize until the set of mandatory plays have been executed.

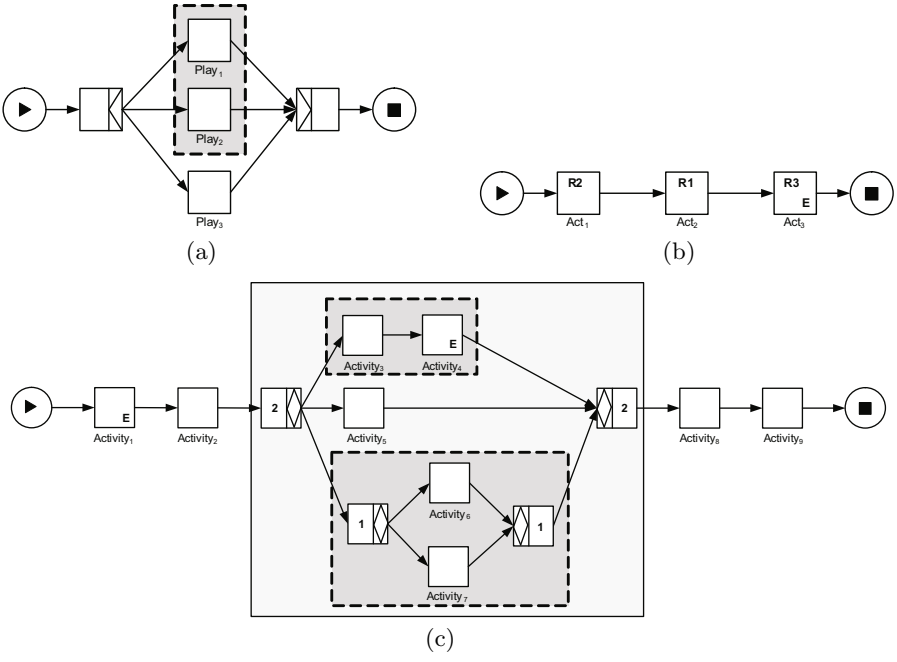


Fig. 1. Example of the patterns for IMS LD representation. The pattern (a) controls the parallel execution of plays, (b) the sequence of acts and (c) the execution of a role part of an act.

- In the central part of the figure, rectangles represent the set plays to run in parallel. The rectangles represent tasks in YAWL, which can be simple or complex. In the case of this model, each play is a complex task that have an internal structure (hidden in the model) defined by the acts of the play.
- A point where the execution of the plays is synchronized. This point is represented by the AND-JOIN pattern of YAWL.
- An end place of the method. The circle with a square represents the point where the method finishes executing.

In order to facilitate the understanding of the model, we extended the visual syntax of YAWL with a gray colored rectangle that identifies the plays of the method whose execution is mandatory.

2.2 Representation of Plays

A play consists on an ordered execution of a set of acts. YAWL does not have a specific pattern to represent sequences. In YAWL a sequence is established based on the dependencies between the elements, where dependencies are determined by the directed arcs between the visual elements. For example, Fig. 1(b) depicts this structure with a sequence of three acts, where Act_1 is the first task to be performed and Act_3 is the last one.

As in the case of plays in the method model, acts are represented by means of rectangles. These rectangles also represent complex tasks, although in this case their internal behavior will coordinate the execution of the role-parts of the act. Therefore, acts will be associated with a set of nets representing each of the role-parts to be made, unlike the case of plays that were associated with a unique net.

In addition we have extended the notation defined by YAWL to enable designers to see if an act has a role-part: their shape is annotated with a R accompanied by a number. For example, in Fig. 1(b) the Act_1 act is annotated with $R2$ which means that it is associated with two role-parts. In the case of acts 2 and 3, the notation indicates that they have one and three role-parts, respectively. The same applies with the environments associated with acts, although in this case they are denoted by an E. Therefore, when an act has any environment its shape is annotated with a E at the bottom left.

2.3 Representation of Role-Parts

The representation of a role-part is more complicated. In the case of methods and plays the structure of the net is stable and varies only when a new element is added to the model. For example, a new play in a method is just a new parallel branch within the AND-split and AND-JOIN patterns. By contrast, the design of a role-part does not have a fixed structure, its structure is defined by combining:

- *Sequence of activities.* They are defined with a sequence structure, as we have described in Section 2.2, but with the restriction that only simple activities, selection of activities, and UoLs can be part of this sequence.
- *Selection of activities.* They are defined within a structure formed by an OR-SPLIT and an OR-JOIN, although in our model both constructs take a sightly different meaning: the execution thread remains in this structure until a number of execution entities (selections, sequences, simple activities, or UoLs) have been selected. The number of execution entities that must be selected is represented as a number in the center of the OR-SPLIT/OR-JOIN control construct.
- *Simple activities.* They are the leaves of our hierarchical model and do not break down into simpler elements.
- *Units of learning.* They are represented as activities, through a rectangle, but represent a complex task. In this case, the task will execute a method model as we have described structure in Section 2.1.

An example of role-part is shown in Fig. 1(c) which represents a sequence of five activities, where the third one is a selection of activities. This figure also shows how different elements are combined, in this case one of the selection of activities is made by a sequence of activities, a simple activity and another selection of activities. As in the case of acts, environments may be associated with activities, being denoted by an E in the bottom left of the rectangle.

3 Hiding the Complexity of Our Visual Model

The visual model described in the previous section only shows part of the operational semantics that supports the execution of UoLs. Our visual model covers only structural features of UoLs. For example, it only contemplates the coordination of plays in a method, of the acts in a play, or of the role-parts in an act. If we want to have a complete model it is necessary to complement the visual model with other features more closely linked to the management of the execution.

In this paper we complemented our model with the approach followed in OPENET4LD. Behind OPENET4LD there is a set of Petri nets that support the implementation of UoLs and also provide a set of features that facilitates the management of methods, plays, events and activities. As a result, these items can be suspended, resumed, or stopped, and this action will be propagated to items that are hierarchically dependent on him.

3.1 Petri Nets-Based Workflows of OPENET4LD

OPENET4LD is a UoL player that allows uploading UoLs stored in the XML format defined by the IMS LD specification. The main difference with other IMS LD engines currently available, such as CopperCore [15], GRAIL [16], MOT+ [17] or LAMS [18], is that OPENET4LD is based on a workflow formalism, specifically on hierarchical Petri nets, so it may verify the structural properties and the correction of the implemented courses.

The transition from IMS LD is complicated and generates structurally complex Petri nets. This is because IMS LD introduce control elements that restrict the order in which activities can be executed. This control structure is based on the theater metaphor where each of the items (plays, acts, activities) requires a particular coordination. OPENET4LD implements two types of Petri nets associated with each IMS LD element. The first net was designed to unify the way in which a method, play, act, or activity is executed independently of the level in which it is situated in the hierarchy. This net is represented in Fig. 2 and is used to indicate the current state of execution of the element. As it can be seen in this figure, the net consists of four parts. The first part is focused on the transition *run-element*, and is responsible of the execution of the element. The second and third part control the suspension and resumption of the element triggered by the user or by a hierarchically higher element. Finally, the last part controls the stopping of elements motivated by the stopping of a hierarchically higher element, by a user trigger, by a time limit consumption, or by a certain condition established in IMS LD.

The second net captures the control structure for the execution of methods, plays, acts, and structure of activities. Each of these structures reflects the operational semantics defined by IMS LD and transforms it into a Petri net. For example in the case of a method, this transition is replaced by a net with a set of parallel transition representing the mandatory (defined in the when-play-completed IMS LD property) and non-mandatory plays of the method.

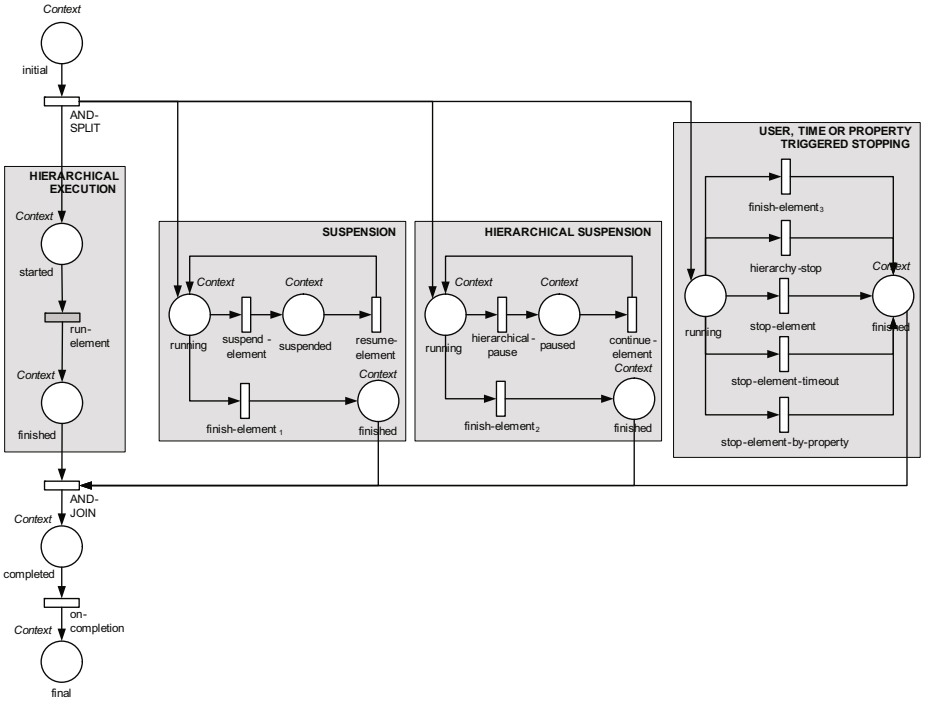


Fig. 2. Petri net that controls the execution, suspension, resumption and completion of methods, plays, acts, and activities of the unit of learning

In the OPENET4LD model, an element (such as method, play, act or activity) is therefore composed of a control net and a structure net. The union between the two nets is done through two composition mechanisms of hierarchical Petri nets: fusions and substitutions. For example, the union between a control net and its structure is done by substituting the transition *run-element* of the control net by a Petri net with the appropriate control structure. In the case of a method, this transition is replaced by a net with a similar structure as the Petri net depicted in Fig. 3. This substitution allows methods to coordinate the execution of their plays. It should be mentioned that this process is repeated for each hierarchical level of IMS LD, that is, for each method, play, and act. The union between the various components is also carried out through transitions substitution. For example, transitions *run-play*$_i$ of Fig. 3 by a control net integrates the execution of the plays with the Petri nets that represent the method execution. This feature gives to our hierarchical model a tree-shaped structure where each tree level alternates a control net with a structure net, and where tree leafs are the activities to perform.

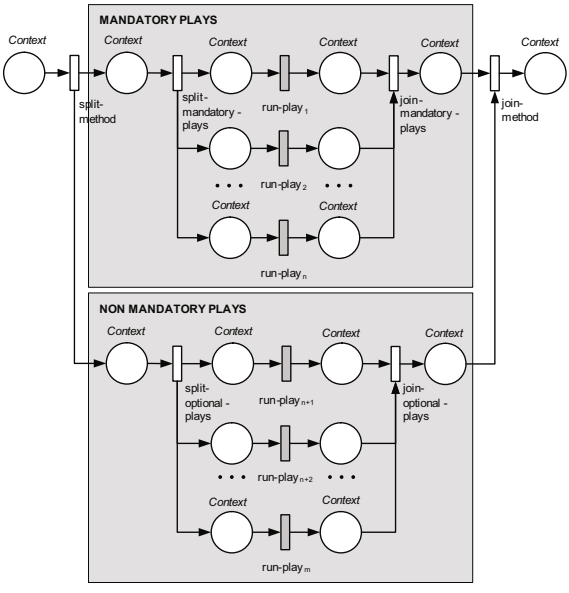


Fig. 3. Petri net that controls the execution of a set of plays

3.2 Mappings between OPENET4LD and the Visual Model

We have established a set of mappings between our visual models and the Petri nets models of OPENET4LD. Through these correspondences, our visual model is transformed into an executable model based on Petri nets. This combination is an important advantage, because (i) it takes advantage of the simplified approach for designing UoLs that our visual models provide, and (ii) it maintains the other advantages that OPENET4LD and its Petri nets-based model provides.

Certainly the implemented GUI, that supports the visual model, offers the option to save designed UoLs in XML format of IMS LD. So, is it justified the definition of correspondences between the visual model and the Petri nets considering that OPENET4LD allows loading UoLs in the IMS LD format? If the interface have the sole purpose of designing UoLs, in fact it would be meaningless. However, with this interface we intend to go further and visualize the workflows execution in the visual model.

Method mappings. Figure 4 shows an example of the correspondence established between a method described with the visual model and the Petri nets of OPENET4LD. As it can be seen in the figure, the visual representation of the method is mapped to two models of Petri nets: one with the control over the method execution and the other one with the control over the coordination of its plays. We can also see a certain similarity between the visual model and the right-sided Petri net. This is because YAWL language is an extension of Petri nets, and also because both nets were designed to control the coordination

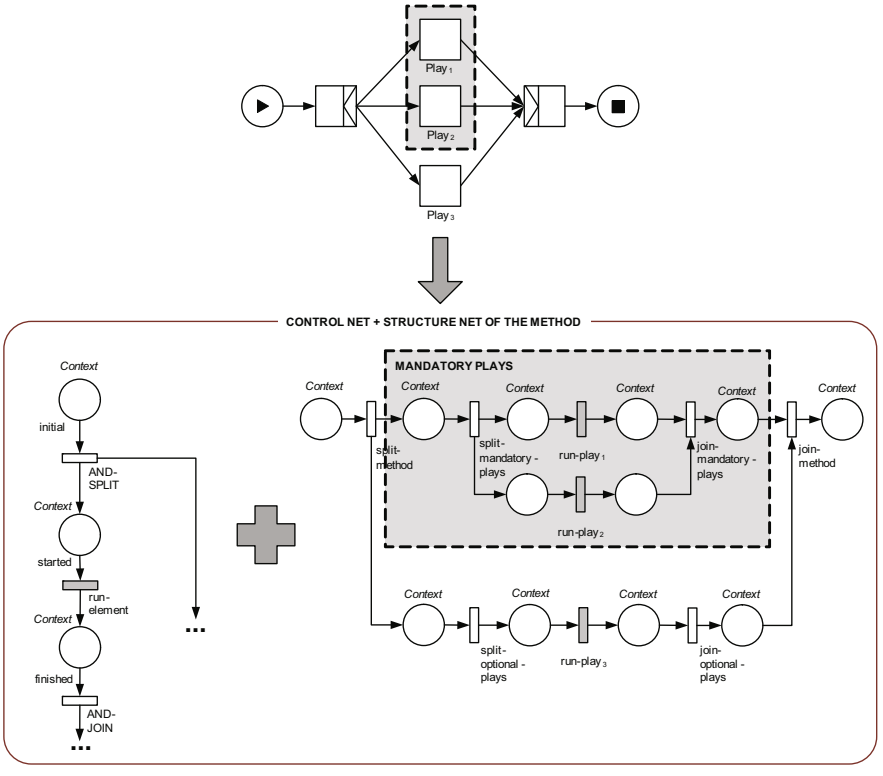


Fig. 4. Mapping between the visual representation of a method and its Petri net-based internal representation in OPENET4LD

of plays. So why do we also establish the mappings of the visual model with the right-sided Petri net? There is an important reason for this: this Petri net manages the execution of the element, and extends the IMS LD specification to support suspension, resumption or abortion of UoLs. Furthermore, it provides the means to propagate changes to the elements that depend on its execution (for example, the suspension of a play implies the suspension of all its acts, and so on). Therefore, these mappings endow our visual model with the same functionalities, so we will be able to manage the execution of the UoL from the GUI.

Play mappings. Like with methods, there is a mapping between the visual model and the two Petri nets that manage the execution of a play and coordinate its acts, respectively. In this case acts are represented in Petri net sequence.

Role-part mappings. The case of acts is different, because our visual model hides much of the structural complexity related to acts execution. In our visual model, an act is associated with a set of role-part models. However, the model does not show how these role-parts should be carried out. This association is

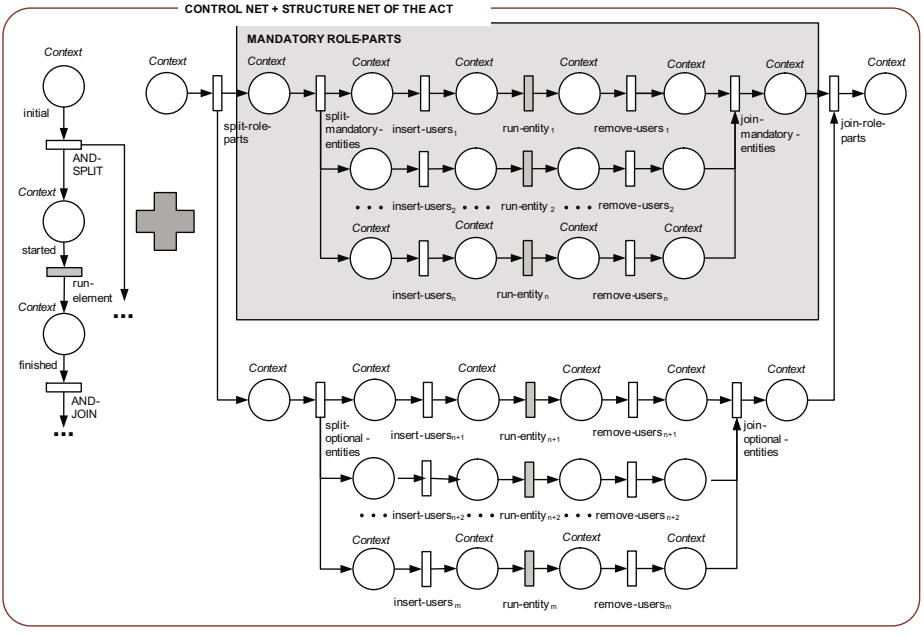


Fig. 5. Each act is mapped to these two Petri nets

done internally through the mappings, so that the Petri net that manages the execution of the act is connected to a Petri net that controls the parallel execution of each of its role-parts. For example, Fig. 5 shows the two Petri nets that control an act execution.

Each branch of Fig. 5 controls the development of the execution entity associated to the role-part. Users are inserted in the workflow execution through the *insert-users*$_i$ transitions, will run the execution entity through the *run-entity*$_i$ transition, and finally removed from the execution by means of the *remove-users*$_i$. The *run-entity*$_i$ transition represents the execution of the simple activity, structure of activities, or UoL associated wit the role-part.

The next step is the identification of the root of the last layer. For example, in Fig. 6 we show the tree of the YAWL model depicted on the top. In this tree, simple activities are the leaves nodes and the structure of activities are the branch nodes. Both node types are substituted by a control Petri net, and in the case of branch nodes, the visual element is also mapped to a net with its corresponding structure.

4 Viewer Implementation

The implementation of the GUI is depicted in Fig. 7. Its display is divided in three parts. The left part contains the panels of components and properties. The

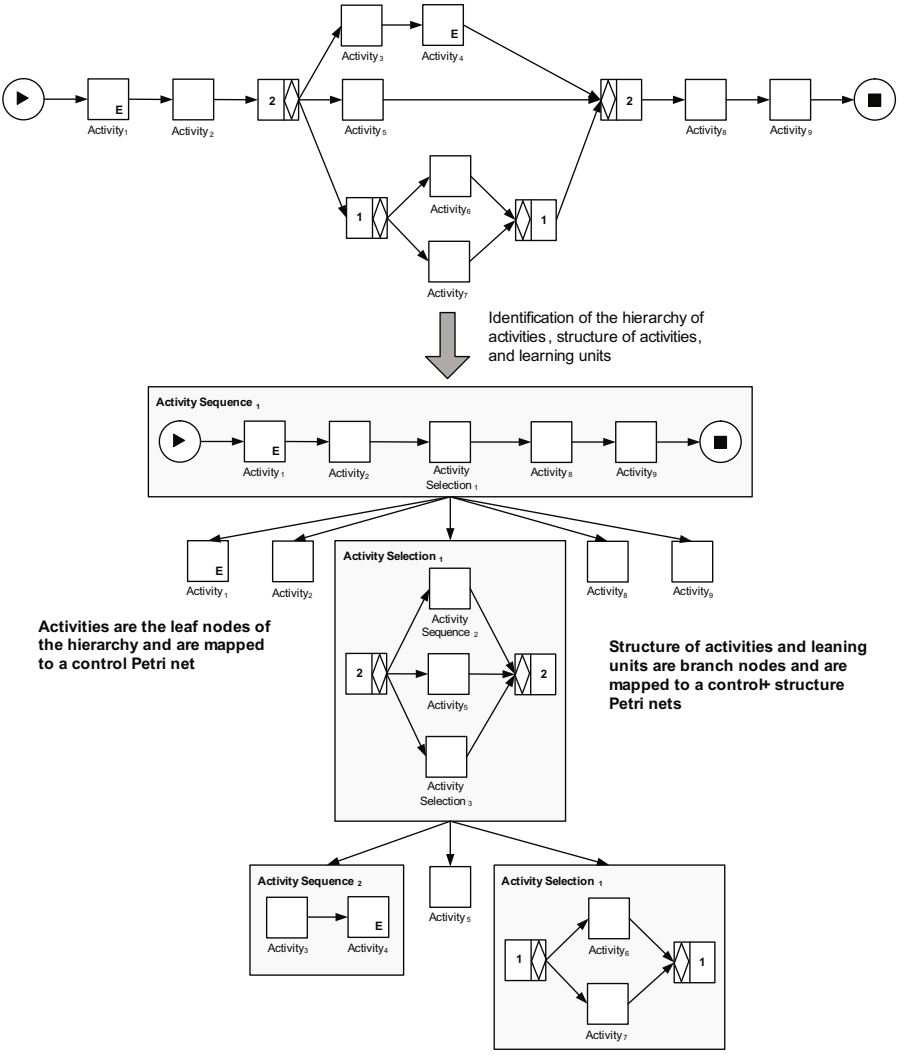


Fig. 6. Example of hierarchical decomposition of a role-part model

first contains the visual components the designer may use to create his visual models, that is, it contains shapes representing the plays, acts, activities, control structures, roles, and environments. These components can be dragged to the central panel in order to create and annotate the visual model.

As it can be seen in the figure, the central panel is again divided in three parts which tally with the three layers of our visual model. Methods are modeled at the top, plays at the central part, while each role-parts of an act is designed in a tab at the bottom. The interaction between the three parts of the model is

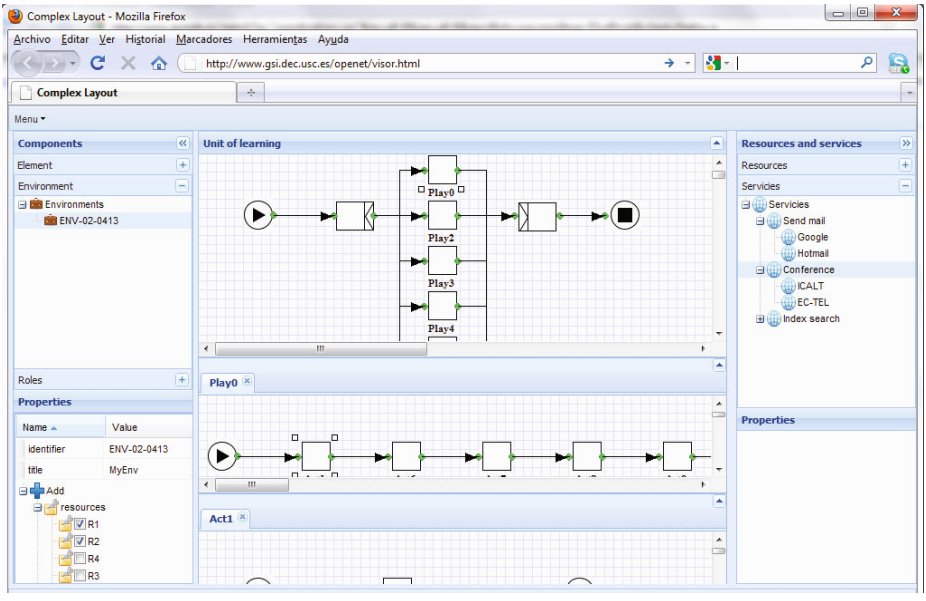


Fig. 7. Screenshot of the GUI for the design/edition of IMS LD units of learning

simple: the selection of play at the top panel will open a tab in the central part with the sequence of acts associated to the selected play; the selection of an act will open all its role-parts as tabs at the bottom part.

The properties panel is located at the bottom of left panel. Through this panel the properties of each of the selected components can be modified. For example, Fig. 7 displays the properties of the environment *MyEnv*. This panel let the designer to change the environment identifier, name, learning objects, services, and other environments it has associated.

Finally, the right side of the GUI is used to create the learning objects and services that may be associated with the environments.

5 Conclusions and Future Work

In this paper we have described a workflow-based approach that simplifies the way to create units of learning specified in IMS LD. The proposed model hides much of the complexity of IMS LD. However, this complexity is only hidden to users, since internally we maintain the structure established by IMS LD by means of a set of mappings to the Petri nets of OPENET4LD, the engine that supports the execution of our models. With this approach we take advantage of the simplified approach for designing UoLs, but maintaining the advantages that OPENET4LD and its Petri nets-based model provides.

As future work we plan to complete (i) our visual model to support the level B of IMS LD, and (ii) the GUI to support the execution of the units of learning in a workflow-oriented way.

References

1. IGL Consortium: IMS Learning Design Information Model, Version 1.0 Final Specification (2003)
2. Griffiths, D., Beauvoir, P., Sharples, P.: Advances in editors for ims ld in the tencompetence project. In: Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), Santander, Spain, pp. 1045–1047. IEEE Computer Society (2008)
3. Gutiérrez Santos, S., Pardo, A., Delgado Kloos, C.: Authoring courses with rich adaptive sequencing for ims learning design. J. UCS 14(17), 2819–2839 (2008)
4. Hernández Leo, D., Villasclaras-Fernández, E.D., Asensio-Pérez, J.I., Dimitriadis, Y.A., Jorrín-Abellán, I.M., Ruiz-Requies, I., Rubia-Avi, B.: Collage: A collaborative learning design editor based on patterns. Educational Technology & Society 9(1), 58–71 (2006)
5. Laforcade, P.: Graphical representation of abstract learning scenarios: the uml4ld experimentation. In: Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), pp. 477–479. IEEE Computer Society, Niigata, Japan (2007)
6. Neumann, S., Oberhuemer, P.: User evaluation of a graphical modeling tool for ims learning design. In: Spaniol, M., Li, Q., Klamma, R., Lau, R.W.H. (eds.) ICWL 2009. LNCS, vol. 5686, pp. 287–296. Springer, Heidelberg (2009)
7. Martínez-Ortiz, I., Moreno-Ger, P., Sierra-Rodríguez, J.L., Fernández-Manjón, B.: A flow-oriented visual language for learning designs. In: Li, F., Zhao, J., Shih, T.K., Lau, R., Li, Q., McLeod, D. (eds.) ICWL 2008. LNCS, vol. 5145, pp. 486–496. Springer, Heidelberg (2008)
8. Dodero, J.M., Martínez del Val, Á., Torres, J.: An extensible approach to visually editing adaptive learning activities and designs based on services. J. Vis. Lang. Comput. 21(6), 332–346 (2010)
9. Karampiperis, P., Sampson, D.: Towards a common graphical language for learning flows: Transforming bpel to IMS learning design level Arepresentations. In: Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), pp. 798–800. IEEE Computer Society, Los Alamitos, CA, USA (2007)
10. Vidal, J.C., Lama, M., Sánchez, E., Bugarín, A.: Application of petri nets on the execution of ims learning design documents. In: Proceedings of the 3rd European Conference on Technology Enhanced Learning (EC-TEL 2008), pp. 95–100. Springer, Maastricht, Holand (2008)
11. van der Aalst, W.M.P.: The application of petri nets to workflow management. The Journal of Circuits, Systems and Computers 8(1), 21–66 (1998)
12. Ellis, C.A., Nutt, G.J.: Office information systems and computer science. ACM Comput. Surv. 12(1), 27–60 (1980)
13. van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M.: Workflow mining: A survey of issues and approaches. Data Knowl. Eng. 47(2), 237–267 (2003)

14. van der Aalst, W.M.P., Hofstede, A.H.M.: Yawl: Yet another workflow language. Information Systems 30(4), 245–275 (2005)
15. Vogten, H., Martens, H.: CopperCore Version 3.3. Open University of Netherlands (2009)
16. Escobedo, J.P., de la Fuente Valentin, L., Gutierrez, S., Pardo, A., Delgado Kloos, C.: Implementation of a learning design run-time environment for the .LRN learning management system. Journal of Interactive Media in Education 1, 1–12 (2007)
17. Paquette, G., Léonard, M.: The educational modeling of a collaborative game using mot+ld. In: Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies (ICALT 2006), pp. 1156–1157. IEEE Computer Society, Kerkrade, The Netherlands (2006)
18. Dodero, J.M., Ghiglione, E.: Rest-based web access to learning design services. IEEE Transactions on Learning Technologies 1(3), 190–195 (2008)

From a Pattern Language to a Pattern Ontology Approach for CSCL Script Design

Jonathan Chacón, Davinia Hernández-Leo, and Josep Blat

ICT Department, Universitat Pompeu Fabra, C/Roc Boronat 138, 08018 Barcelona, Spain
{jonathan.chacon,davinia.hernandez,josep.blat}@upf.edu

Abstract. Collaborative activities, in which students actively interact with each other, have proved to provide significant learning benefits. In Computer-Supported Collaborative Learning (CSCL), these collaborative activities are assisted by technologies. However, the use of computers does not guarantee collaboration, as free collaboration does not necessary lead to fruitful learning. Therefore, practitioners need to design CSCL scripts that structure the collaborative settings so that they promote learning. However, not all teachers have the technical and pedagogical background needed to design such scripts. With the aim of assisting teachers in designing effective CSCL scripts, we propose a model to support the selection of reusable good practices (formulated as patterns) so that they can be used as a starting point for their own designs. This model is based on a pattern ontology that computationally represents the knowledge captured on a pattern language for the design of CSCL scripts. A preliminary evaluation of the proposed approach is provided with two examples based on a set of meaningful interrelated patters computationally represented with the pattern ontology, and a paper prototyping experience carried out with two teaches. The results offer interesting insights towards the implementation of the pattern ontology in software tools.

Keywords: CSCL, Pattern Ontology, Pattern Selection, Collaborative Learning Design, Pattern Language.

1 Introduction

The research presented in this paper is framed in the Technology-Enhanced Learning (TEL) research domain, which focuses on how advanced Information and Communication Technologies can improve teaching and learning. In particular, this work is within the context of Computer-Supported Collaborative Learning, a discipline inside TEL that promotes the use of technology to support collaborative learning activities [14, 18].

Supporting the creation of potentially effective learning designs is a relevant topic in TEL, but it is even more relevant in CSCL because of the inherent complexity of collaborative dynamics [3]. Achieving effective collaboration in learning has a number of implications that go beyond the provision of communication tools that allow students to interact with each other. Furthermore, free collaboration does not necessarily trigger effective learning. A solution proposed in the literature to face this problem is the creation of meaningful CSCL scripts to enhance the effectiveness of

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 547–561, 2011.

collaboration. CSCL scripts are designs that specify how students should form groups, how they should interact and collaborate along a sequence of activities and which resources and tools they should use in each activity in order to solve the presented problems or tasks [2, 3, 10].

However, the creation and design of CSCL scripts is not trivial. It is particularly difficult for educators interested but not experts in Information and Communication Technologies (ICT) and collaborative learning. There are a number of evidences in the literature that show the challenges around teachers designing scripts from scratch [4]. A solution to face this problem, which is being proposed in different literature contributions, is to support the design process through the reuse of existing material to create new ones [7]. According to [8] there are different levels where reutilization could be applied: we can reuse a whole exemplar that is ready-to-run [15]; we can reuse generalizations of successful collaboration scripts formulated as patterns; or we can reuse learning objects / resources as components or building blocks. This paper is focused on the use of design patterns, since they seem to be a highly reusable type of learning design solutions [3, 10].

A design pattern provides a means of organizing information regarding a contextualized common problem and the essence of its broadly accepted solution, so that it can be repetitively applied. These patterns can be provided as templates in authoring tools, if formalized using computer interpretable notation. A collection of interconnected (related) patterns which enable the generation of a coherent whole is called a Pattern Language (PL) [3, 5]. Some visual representation of pattern languages includes: a hierarchical structure of CSCL scripting PL [10], a component and plugging based system [6], or e-learning frameworks, which comprise design patterns and software components [1].

However, supporting reusability is not trivial since reuse-oriented design processes pose a number of challenges that need to be tackled in order to be successful. These challenges include the definition of assembling and refining processes when combining and particularizing the reusable material as well as addressing the searching and selection of the materials themselves [8]. When combining a set of selected materials, it is critical that the result is “harmonic”, i.e., a meaningful coherent design. This paper aims at providing a solution that facilitates educators to select of a set of patterns for the design of a meaningful CSCL script. The result is expected to be potentially effective since the script is based in good practices (patterns) and the educators creating the script do not necessarily need to be experts in the use of ICT and in scripted collaborative learning.

The approach followed in this paper to face this aim exploits the semantic of the patterns to computationally relate the knowledge captured in those patterns. In this paper, we propose to move from pattern languages to pattern ontologies and we present pattern ontology for the design of CSCL scripts. The pattern ontology is illustrated with two sample CSCL scripts. Moreover, the paper also sketches the integration of the ontology in a tool to assist the selection of interrelated patterns for the design of a script reusing those patterns. To identify relevant issues to be considered in that integration, a paper prototyping with two real users has been carried out.

The remainder of the paper is organized as follows. In section 2, we present the state of the art on reusing learning designs that triggered our work. In this section we explain the research gap in which the contributions of this paper are situated, mention the related work and formulate the specific problem tackled. In section 3 we present the pattern ontology and discuss the approach that will be followed to integrate the ontology in a tool that will enable the selection of a set of interrelated patterns. Section 4 is devoted to preliminarily evaluate the proposals of the paper. In particular, it includes two sample CSCL scripts that illustrate the ontology and a paper prototyping experience providing insights regarding the use of the ontology. Finally, section 5 describes the main conclusions of the paper and the future work.

2 Background and Problem Formulation

In this section we describe more deeply the "designing by reusing" approach, when particularized to CSCL, that motivate our research. Furthermore, we discuss two significant examples in CSCL that use ontology as their basis. Finally, we formulate the specific research problem faced in this paper.

2.1 Designing by Reusing

ICT are being introduced in schools progressively, and practitioners are being encouraged to create material and design new practices using these technologies. In this context, the application of CSCL and, particularly CSCL scripts, is an interesting option that would enhance the teaching and learning processes in the class. However, the process of designing a CSCL script is not an easy task for practitioners. Not all of them are both familiar with technology itself and the CSCL approach. In order to facilitate this task the solution proposed in this paper adopts the "designing by reusing" approach, which has already proved to be useful in the literature [8]. This approach proposes to start reusing already existing and proved-to-be-good material instead of starting from scratch. Doing so, practitioners neither familiar with technologies nor with CSCL approach can have a solid help to design their own CSCL scripts.

The TEL literature has applied and studied different levels of reusability [7, 8]. According to these studied levels of reusability, the options to support the creation by reusing of CSCL scripts would be:

- **CSCL Exemplars** as ready-to-run scripts. These scripts may embrace from one activity session to a whole course. Practitioners could select the exemplar and use it directly, or they could use the exemplar as starting point and change it according to their needs.

- **Templates** as generalizations of successful collaboration scripts formulated as design patterns. Design patterns provide a means of organizing information regarding a contextualized common problem and the essence of its broadly accepted solution. But, still, without the particularizations (such as tasks descriptions and resources) that are needed in order to achieve a ready-to-run CSCL script.

Practitioners could use patterns as a starting point and complete or complement them according to the guidelines and information provided by the pattern.

- CSCL script **chunks** as portion of exemplars. These CSCL script chunks could be descriptions of an activity structure containing: activities, environments and resources needed. This kind of material cannot be "playable" by itself. Practitioners could assemble a set of chucks or use them to complete their templates in the creation of a full-fledged CSCL script.
- **Building blocks** or **components** are partially completed chunks. This kind of material could be activities not totally defined, just resources to support activities, etc. Fine-grained design patterns can be provided as refinable building blocks. Practitioners could reuse building blocks or components to build their own chunks and assembled exemplars of CSCL scripts.

From the levels above described we focus on design patterns that can be provided as templates or building blocks, since they have proved to be appropriate reusable components for the creation of potentially effective designs [9]. Besides, the use of patterns establishes the utilization of a common format that it is expected to assist the reutilization along the design process.

2.2 Ontology-Based Tools for CSCL

The contribution presented in this paper is, to the best of our knowledge, the first approach proposing to move from pattern languages to pattern ontologies in TEL. However, there are related works that use ontologies in TEL. For instance, [13] build a framework that represents the learning processes involved in collaborative educational situations, using an ontology adopting concepts from learning theories. In this approach the learning theories provide the concepts to justify and support the development of effective learning scenarios. In order to verify the viability and usefulness of their proposed ontological framework within the context of systematic design, they developed and used and intelligent authoring tool for CSCL design. This system is able to reason on ontologies in order to give suggestions that help users in the creation theory-compliant collaborative learning scenarios.

Another example is [19], which introduces Ontoolcole as an ontology of collaborative learning tools designed with the aim of supporting educations in the search of CSCL tools. This ontology allows educators to integrate external tools, offered as services by software providers, in order to support the realization of collaborative learning situation. Ontolcole has been implemented in Ontoolsearch [20], an application that supports the semantic querying of CSCL tools. Ontoolsearch has been specifically designed so that its interface is easy-to-use by educators, in terms of the query formulation as well of the analysis of the obtained results.

2.3 Problem Formulation

By adopting the designing by reusing approach, teachers are assisted in the creation of CSCL scripts based on existing designs. If these existing designs are patterns (in the

form of templates or building blocks), then the reusable elements represent sound didactic ideas for CSCL scripts. Teachers only need to select the set of patterns that better fit the needs of each educational situation. However, without a guide to select patterns, teachers may end up with a set of isolated patterns that does not fit harmonically with each other. Pattern Languages (PL) were proposed to address this problem, since they are a collection of interconnected (related) patterns which enable the generation of a coherent whole [3]. Though, pattern languages interconnect patterns using natural language. Each pattern, within a pattern language, includes a description indicating which other patterns could be used to complement or complete it. This information cannot be used directly by computers and a computational representation of the interconnections between patterns is needed, if we want to flexibly implement tools supporting the design of CSCL scripts based on patterns.

To achieve the computational representation of pattern languages we propose the use of an ontology-based approach. The semantic characteristics facilitated by ontologies enable the definition of models containing the pattern information along with the relations connecting the patterns. This approach allows to computationally supporting the pattern selection process guiding educators in the design of meaningful CSCL scripts. These scripts are expected to be potentially effective since they will be based on good practices (patterns), even though the educators creating the scripts are not necessarily experts in the use of ICT and in scripted collaborative learning.

3 Pattern Ontology Approach

Moving from pattern languages to pattern ontologies enables the explicit representation of the meaning captured in patterns and their relationships as axioms, obtaining a formal semantic pattern language representation. We use an existing pattern language for the design of CSCL scripts as a starting point [8]. The resulting ontology provides us a solid base, which we plan to extend with more patterns (such as [21]) and linked reusable educational materials. As motivated in the previous section, the final goal is to provide a common framework, which mediated by computers, supports educators in the selection of a set of interrelated patterns to be applied to their own learning situation.

The selection of the ontology instead of other computational approaches such as Databases (DB) is mainly justified because of their flexibility. Although Databases are similar to ontologies regarding both: ontology axioms vs. DB schema, and ontology facts vs. DB data, there are several differences pointed out in **Table 1** [16] that motivate the use of ontologies.

As we can see in **Table 1** ontologies can deal with incomplete information while databases cannot. Furthermore, in ontologies individual elements may have more than one name, what is a critical property when establishing a common vocabulary.

This section introduces the pattern ontology built for the design of CSCL scripts and discusses how this ontology can be implemented in a tool facilitating the selection of patterns.

Table 1. Databases Vs Ontologies

Database	*Ontology*
Closed World Assumption (CWA): missing information treated as false	Open World Assumption (OWA): missing information is treated as unknown
Unique name assumption (UNA): each individual has a single, unique name	No UNA: individuals may have more than one name
Schema behaves as constraints on structure data: define legal database states	Ontology axioms behave like implications (inference rules): entail implicit information

3.1 Pattern Ontology for the Design of CSCL Script

The pattern ontology resulting from the semantic formalization, using the OWL language [12] and the Protégé ontology editor, of the pattern language proposed in [10] can be graphically seen in **Fig 1**. According to the pattern language, the patterns are classified into four categories: *Flow Patterns* (e.g.: JIGSAW), *Activity Patterns* (e.g: INTRODUCTORY ACTIVITY: LEARNING DESIGN AWARENESS), *Resource Patterns* (e.g.: ENRICHING THE LEARNING PROCESS), and *Roles and Common Collaborative Mechanisms Patterns* (e.g.: FACILITATOR or FREE GROUP FORMATION). Extended descriptions of the levels and the patterns can be read at [10].

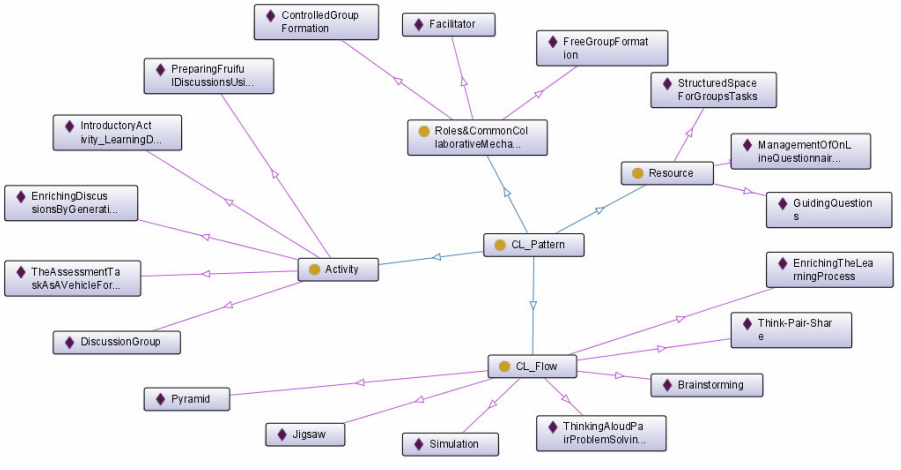


Fig. 1. Tree radial view of our Pattern Ontology (only father to child relationships are showed)

Each pattern represented in our ontology has the following information: category in which the pattern is included, a brief description about the problem it solves, the structure of the solution (especially in flow patterns, which propose sequences of

activities and the group types associated to those activities), and two lists: the first with a set of patterns which *complement* it and the second with a set of patterns which *complete* it. Where *complement* means that the pattern neither modifies nor refines the current one, but together form a larger whole; while *complete* means that the initial pattern is refined with a second pattern that adds further design ideas to those already proposed by the initial pattern. Both lists are a set of recommendations the user should follow once a pattern is selected, in order to complement or complete the pattern towards a coherent set of patterns meaningful for a specific learning situation. **Table 2** shows an example of a pattern belonging to the adopted pattern language and which has been represented in the ontology. Part of this computational representation is shown in **Fig. 2.**

Table 2. Example of a pattern belonging to the adopted pattern language and integrated in the proposed pattern ontology

Category	*Flow pattern*
Name	JIGSAW
Problem description	If groups of students face the resolution of a complex problem/task that can be easily divided into sections or independent sub-problems, an adequate collaborative learning flow may be planned.
Problem Structure	Structure the learning flow so that each student (individual or initial group) in a group ("Jigsaw Group") studies or work around a particular sub-problem. Then, encourage the students of different groups who study the same problem meet in an "Expert Group" for exchanging ideas. These temporary focus groups become experts in the section of the problem given to them. At last, students of each "Jigsaw group" meet to contribute with its "expertise" in order to solve the whole problem.
Complement:	PYRAMID, BRAINSTORMING, TPS, SIMULATION, TAPPS, ENRICHING THE LEARNING PROCESS, INTRODUCTORY ACTIVITY: LEARNING DESIGN AWARENESS.
Complete:	PYRAMID, TPS, BRAINSTORMING, TAPPS, INTRODUCTORY ACTIVITY: LEARNING DESIGN AWARENESS, DISCUSSION GROUP, THE ASSESSMENT TASK AS A VEHICLE FOR LEARNING, FREE GROUP FORMATION, CONTROLLED GROUP FORMATION.

The process suggested to educators when using the ontology in the selection of a set of patterns to be used in the design of a CSCL script is as follows. Initially, educators are proposed to start selecting any flow pattern. Once they pick up their first pattern, the ontology recommends a set of patterns (of different categories) to complement or complete their selection. See the example in **Fig. 3,** which shows how six patterns are suggested to be used in combination with THE ASSESSMENT TASK AS A VEHICLE FOR LEARNING pattern. This process can iteratively followed (users can navigate through the visible patterns and each time they select a pattern a new set of recommended patterns will be shown) until the educators consider they have all the (interrelated) design ideas they need to create their own CSCL script.

```
88    <Declaration>
89        <NamedIndividual IRI="#Jigsaw"/>
90    </Declaration>
...
119   <SubClassOf>
120       <Class IRI="#CL_Flow"/>
121       <Class IRI="#CL_Pattern"/>
122   </SubClassOf>
...
179    <ClassAssertion>
180        <Class IRI="#CL_Flow"/>
181        <NamedIndividual IRI="#Jigsaw"/>
182    </ClassAssertion>
...
219    <SubObjectPropertyOf>
220        <ObjectProperty IRI="#complete"/>
221        <ObjectProperty abbreviatedIRI="owl:topObjectProperty"/>
222    </SubObjectPropertyOf>
...
250     <owl:Class>
251       <owl:unionOf rdf:parseType="complete">
252         <owl:Class rdf:about="#Jygsaw"/>
253         <owl:Class rdf:about="#Pyramid"/>
254       </owl:unionOf>
255     </owl:Class>
```

Fig. 2. Fragment of the pattern ontology showing the definition of the JIGSAW pattern and its relationship with the PYRAMID pattern

Thanks to the flexibility provided by the ontologies, educators can use the name under which they know a specific pattern, though this pattern may be also known under another title (i.e. PYRAMID is also known as SNOWBALL). Also, since ontologies are ruled by inferences, we can add new patterns and they will be properly integrated in the ontology by the reasoner.

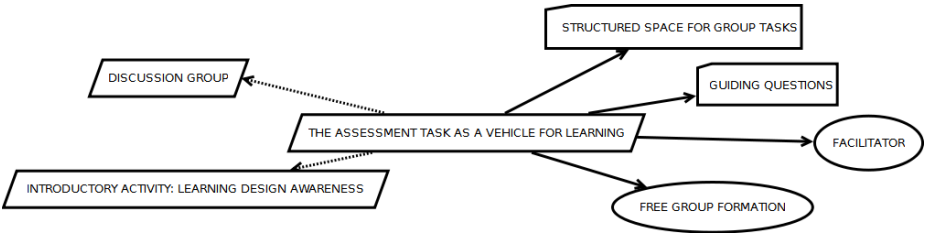


Fig. 3. Pattern selected in the center and patterns recommended in the ontology to meaningfully complement or complete it, shown, at the periphery

3.2 Towards the Implementation of the Pattern Ontology in the LdShake Tool

The pattern ontology, created with the Protégé tool using the OWL language, is, at the end, a valid XML file, see **Fig. 2**. So, to integrate this ontology within any authoring tool or pattern repository two main modules need to be implemented: 1) A parser that reads the ontology and 2) a piece of code that transform this information into objects, so that it can be used by the logic of the tool. Besides, the logic of the tool would need to be extended so that it supports the two main types of users that would interact with the ontology.

The first type of users comprises educators that select a set of interrelated patterns according to the recommendations provided by the ontology and the requisites of their particular learning situations. On the other hand, the second type of users can be practitioners, familiar or experts in CSCL scripts, who may want to use other patterns or consider relationships not represented in the ontology. The tool should be flexible in these cases and may even valorize their design decisions and extend the knowledge captured in the ontology with more patterns or new meaningful relationships between the patterns. In this way, the novice educators would be benefited with the contributed knowledge.

We are currently implementing the pattern ontology in the LdShake tool [11]. LdShake is a Web 2.0 tool for the social sharing and co-edition of learning design solutions, including design patterns. For this particular case we are developing a parser that reads the ontology information and plan to develop a second module that visually shows the patterns recommendations to the users.

4 Preliminary Evaluation

This section provides a preliminary evaluation of the pattern ontology. First, we present two real scenarios designed by practitioners that describe two CSCL scripts based on a set of patterns. In the description of the scenarios we show how the ontology is able to represent the interconnections between the set of selected patterns so that it is meaningful for those scenarios. Second, we explain a paper prototyping experience with two other teachers, familiar with CSCL, who used the ontology for the creation of new scripts. Before creating the new scripts, we proposed the teachers to read the description of the previously mentioned two scenarios without indicating the patterns but only the main ideas that the teachers considered when planning those scenarios (number of students, main goals, how they intended to structure the classes, how many sessions they had, the expected outcomes, etc.). Together with the scenario description, we also gave the teachers a separated sheet with the sets of patterns that "solve" or "state the solution" to both scenarios. We gave the teachers some minutes to read the scenarios and look at the solutions. Then, we asked them to make a description of a similar (imaginary or not) course and to use a paper prototype of the ontology. They must select as many patterns as they think they could use to cover their course needs. Finally, we asked the teachers to complete a questionnaire in order to understand the benefits and limitations of the proposed approach as well as to learn lessons relevant for the implementation of the ontology in pattern selection tools.

4.1 Two Examples Computationally Represented with the Pattern Ontology

The first example belongs to a "Computer Architecture" course, part of the core body of knowledge in the Telecommunications Engineering curriculum in Spanish universities. The whole course is defined as a project that develops along the semester. Its objective is the design and evaluation of a computer system. The teacher defines five fictional clients and assigns one client to the students grouped in dyads. This way, in each laboratory group, different clients are being studied through the course, following the principles of the JIGSAW pattern. The Jigsaw-based structure is

completed with the suggestions of the SIMULATION pattern where teacher plays the role of client. Furthermore, during the course the teacher becomes a FACILITATOR marking the milestones and presenting different assessment task to the students, as indicated by THE ASSESSMENT TASK AS A VEHICLE FOR LEARNING pattern. In each milestone, every laboratory group ("Jigsaw group" phase of the JIGSAW) holds a debate. This debate is arranged as suggested by PREPARING FRUITFUL DISCUSSIONS USSING SURVEYS and complemented with ENRICHING DISCUSSIONS BY GENERATING COGNITIVE CONFLICTS. At the end, a technical report is collaboratively produced among all dyads that have worked with the same client in each laboratory session (forming accordingly a PYRAMID). **Fig. 4** shows the graphical representation of the patterns used in the scenario and their relationships according to the proposed ontological model.

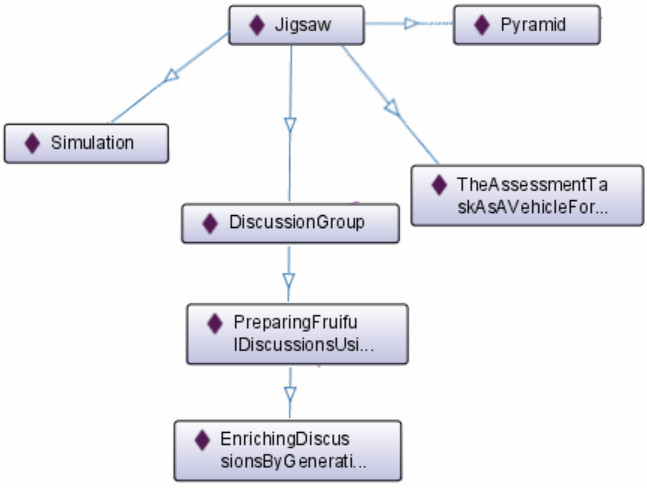


Fig. 4. Patterns and their relationships according to the pattern ontology used in the Computer Architecture CSCL script

The second example is framed in a course on "ICT resources in Education". The global objective of this course is to allow students to create didactic units in collaboration. The course was structured as follow: during the first week they were introduced to the course and the general plan following the indications of the INTRODUCTORY ACTIVITY: EXPLANING THE LEARNING DESIGN pattern. The following weeks were planned according to a two-level PYRAMID: the first level of the Pyramid is, in turn, structured in accordance with a JIGSAW. Taking into account the FREE GROUP FORMATION pattern, the students are assembled in dyads.

In the "Individual phase" of the JIGSAW every dyad studies one of the 3 main topics of the course. Then, students have to summary the main ideas of their topic and elaborate a report for assessment purposes but also as a learning task (ASSESSMENT TASK AS A VEHICLE FOR LEARNING) which pushes them to reflect on a series

of questions that they should answer in the report. These questions are explicitly provided by the teacher, as suggested by GUIDING QUESTIONS.

In the "Expert Group" phase of the JIGSAW, groups of six (or seven) dyads that have worked over the same topic join in a single large group to read and discuss the reports written by their partners. In the "Jigsaw Group" phase, new groups are formed. Every group comprises a pair "expert" on each topic. In this phase, the students read and present the second report (outcome of the "expert group") and elaborate a new common report integrating the three different topics.

Finally, the second (and last) level of the PYRAMID is devoted to ENRICHING DISCUSSIONS BY GENERATING COGNITIVE CONFLICTS in a global debate where all students participate. **Fig. 5** shows the graphical representation of the patterns used in this scenario and their relationships according to the proposed ontological model.

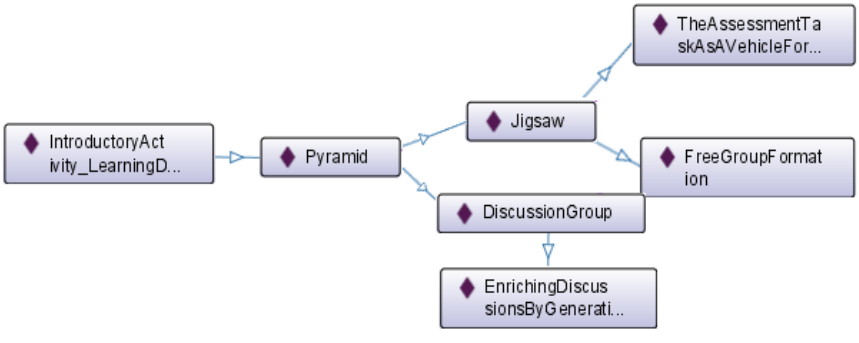


Fig. 5. Patterns and their relationships according to the pattern ontology used in the ICT resources in Education CSCL script

4.2 Paper Prototyping with Teachers: Lessons Learnt for the Pattern Ontology Implementation

It is clear that a preliminary evaluation of the proposed pattern ontology approach is needed in order to provide insight about the proposal and its implementation in tools [22]. So, in order to be able to test the approach and obtain a first feedback we selected to use the paper prototyping method. This approach allows us to present the pattern ontology to users who could perform realistic tasks by interacting with a pattern ontology paper version that is manipulated by ourselves as "computers" [17]. The paper prototyping was carried out with two teachers following the process described above.

For the paper prototyping we draw a total of eighteen charts, the number of considered in the ontology. Examples of these charts are **Fig. 3** and **Fig. 6**. Each chart has a pattern in the center, representing the selected pattern, an all patterns related to it according to the ontology, in the periphery. We distinguish the different pattern category with a different shape: square shape for *Flow Patterns* (i.e.: JIGSAW in **Fig. 6**), square shape with a plane corner for *Resource (i.e.:* THE ASSESSMENT TASK AS A VEHICLE FOR LEARNING in **Fig. 3**), oval shape for *Roles and Common*

Collaborative Mechanisms (i.e.: FREE GROUP FORMATION in **Fig. 6**) and diamond shape for *Activity Pattern* (i.e.: DISCUSSION GROUP in **Fig. 6**). Furthermore, we draw of the diverse types of relationships between patterns with different contour lines. We distinguish three different ones: solid lines for *complete* relations (i.e.: JIGSAW to DISCUSSION GROUP, bottom-right in **Fig. 6**); dotted lines for *complement* relations (i.e.: JIGSAW to SIMULATION, top-left in **Fig. 6**); and dashed-dotted lines to patterns that could be used for both, complement and/or complete (i.e.: JIGSAW to PYRAMID, center-top in **Fig. 6**). We included a legend describing all this information.

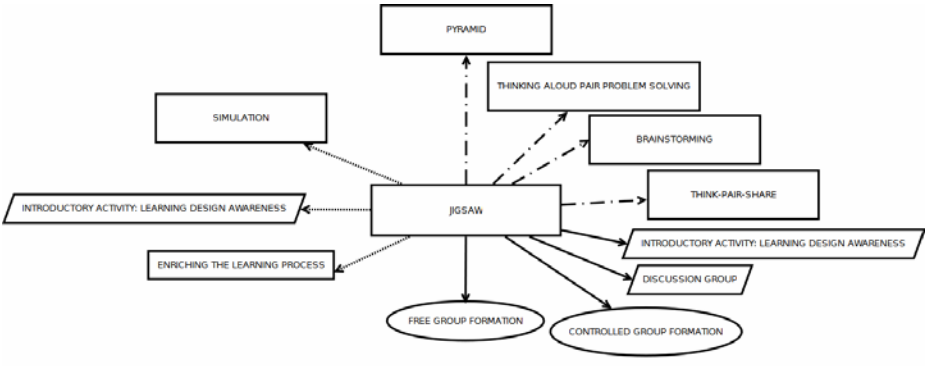


Fig. 6. Paper prototyping example: JIGSAW flow pattern chart

During the sessions, each session take one hour long, we acted as the computer and any time the teachers selected a pattern we marked it as selected and showed teachers another chart expanding the relations of this pattern with other patterns. Together with these charts we also gave teachers tables with the descriptions of the patterns, so that they could consult them if needed. See **Table 2** to see a compacted version of the table for the JIGSAW pattern.

The analysis of the data collected in the questionnaires completed by the two teachers lead to the conclusions shown in **Table 3.** As a positive feedback we obtained that with our approach we foment the creativity of the teachers. According to the teachers' opinion, the ontology suggests relationships between patterns that made them to consider patterns in their designs that otherwise they would not have included. Moreover, they were satisfied with the resulting design and said that the designed script was well structured and the considered strategies seem to have potential to enhance the collaborative learning of their chosen educational situation. Aspects for improvement suggested by the teachers were around the amount of information presented to the users (especially at the beginning). Besides, the teachers needed to read a lot while starting to be aware about the patterns and relate their already known patterns to the patterns suggested by the system. In some cases, it seemed that they were familiar to the design ideas captured in the patterns but the system was using different titles for the patterns. Finally, the teachers also pointed out that a more clear legend would be required to better explain shapes and line counters (described in the above section 4.2).

Table 3. Main positives and negatives aspects reported by the teachers

Positive Aspects	Aspects for improvement
• Variety of patterns and relations foments the creativity. • Relations between patterns lead to the selection of extra design ideas that enrich the collaborative learning activities. • The differences between pattern categories (at flow level, activity and resource) to guide the structure of the designs.	• Too much information at the beginning. • Visualization comprehension: arrows and figures meanings (clearer legend required). • Though the ontology guides the selection of the patterns, it is necessary to read the descriptions of the recommended patterns.

The lessons learnt from the paper prototyping experience with teachers are being currently taken into account in the implementation of the ontology in LdShake. In particular, the amount of information showed in screen when users start selecting patterns will only focus on flow patterns, then the user will be able to continue selecting the patterns that can be meaningfully combined with their selected flow pattern, complementing or completing it. Besides, we are designing an accurate legend that describes the different figures representing the different kind of patterns (learning flow, activities, resources and roles). Finally, summaries or graphical representations of the patterns will be visualized as small tooltips that support users in the understanding of the patterns' solutions when they pass the mouse over the patterns.

5 Conclusions and Future Work

This paper contributes to the design-by-reuse approaches aiming at assisting teachers in the creation of potentially effective CSCL scripts. In particular, it focuses on patterns as the design elements to be reused. Pattern languages collect a number of patterns and document the relationships between them. However, these relationships are expressed in natural language and cannot be interpreted by software tools. In this paper, we propose to move from pattern language to pattern ontology in order to computationally represent the knowledge captured in an existing pattern language for the design of CSCL scripts. This pattern ontology can be implemented in tools so as to enable teachers, typically not experts in CSCL or ICT, to creatively select a coherent set of interrelated patterns that suit the requisites of their particular learning situation. Moreover, if implemented in tools practitioners, expert in CSCL, could also flexibly use the ontology and extend its knowledge with more patterns or new meaningful relationships. The preliminary evaluation presented in this paper shows that the support provided by the proposed pattern ontology is encouraging, but also that its implementation in tools, guiding the selection of patterns, is not trivial. The main challenges deals with the visualization of the interconnected patterns and their descriptions. We are currently working on facing these challenges and implementing the pattern ontology in the LdShake tool, supporting the sharing and co-authoring of

design solutions. Moreover, we are also extending the ontology with the integration of assessment patterns [21] and plan to incorporate more patterns and link other educational materials in the short future.

References

1. Calvo, A., Turani, A.: E-Learning Frameworks = (Design Patterns + Software Components). In: Goodyear, P., Retalis, S. (eds.) E-learning, design patterns and pattern languages, pp. 49–64. Sense Publishers (2010)
2. Dillenbourg, P.: Over-scripting CSCL: the risks of blending collaborative learning with instructional design. In: Kirschner, P. (ed.) Three worlds of CSCL. Can we support CSCL?, pp. 61–91. Open Universiteit Nederland, Heerlen (2002)
3. Goodyear, P., Retalis, S. (eds.): E-learning, design patterns and pattern languages. Sense Publishers, The Netherlands (2010)
4. Griffiths, D., Blat, J.: The role of teachers in editing and authoring units of learning using IMS Learning Design. In: Advanced Technology for Learning, vol. 2 (4), pp. 208–218 (2005)
5. Griffiths, D.: From Primitives to Patterns: a Discussion Paper, UNFOLD project (2005), http://digitalcommons.bolton.ac.uk/cgi/viewcontent.cgi?article=1001&context=iec_journals
6. Harrer, A., Martens, A.: A Heterogeneous Pattern Language for Collaborative Learning Systems and Intelligent Tutoring Systems. In: Goodyear, P., Retalis, S. (eds.) E-learning, design patterns and pattern languages, pp. 49–64. Sense Publishers (2010)
7. Harrer, A.: An Approach to Organize Re-usability of Learning Designs and Collaboration Scripts of Various Granularities. In: Sixth IEEE International Conference on Advanced Learning Technologies ICALT 2006, pp. 164–168 (2006)
8. Hernández-Leo, D., Harrer, A., Dodero, J.M., Asensio-Pérez, J.I., Burgos, D.: A Framework for the Conceptualization of Approaches to "Create-by-Reuse" of Learning Design Solution. Journal of Universal Computer Science 13(7), 991–1001 (2007)
9. Hernández-Leo, D., Jorrín-Abellán, I.M., Villasclaras, E.D., Asensio-Pérez, J.I., Dimitriadis, Y.: A multicase study for the evaluation of a collaborative learning pattern-based visual design approach. Journal of Visual Languages and Computing 21(6), 313–331 (2010)
10. Hernández, D., Asensio, J.I., Dimitriadis, Y., Villasclaras, E.D.: Pattern languages for generating CSCL scripts: from a conceptual model to the design of a real situation. In: Goodyear, P., Retalis, S. (eds.) E-learning, design patterns and pattern languages, pp. 49–64. Sense Publishers (2010)
11. Hernández-Leo, D., Carralero, M.A., Chacón, J., Pérez-Sanagustín, M., Carrió, M., Blat, J.: Sharing and co-editing learning design solutions. In: Proceedings of the European LAMS Conference, Oxford, pp. 20–23 (2010)
12. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A practical guide to building OWL ontologies using the protégé-OWL plugin and CO-ODE tools edition 1.0. Technical Report, University Of Manchester (last visited January 2011), http://www.co-ode.org/resources/
13. Isotani, S., Mizoguchi, R., Inaba, A., Ikeda, M.: The foundations of a theory-aware authoring tool for CSCL design. Computers & Education 54(4), 809–834 (2010)
14. Koschmann, T.D.: CSCL, theory and practice of an emerging paradigm. In: Koschmann, T. (ed.) Lawrence Erlbaum, New Jersey (1996)

15. LN4LD: Example IMS-LD packages. Open University of The Netherlands (2005), http://dspace.learningnetworks.org/handle/1820/257
16. Motik, B., Horrocks, I., Sattler, U.: Bridging the gap between OWL and relational databases. Web Semantics: Science, Services and Agents on the World Wide Web 7(2), 74–89 (2009)
17. Snyder, C.: Paper prototyping: The fast and easy way to design and refine user interfaces. Morgan Kaufmann Pub. (2003)
18. Stahl, G., Koschmann, T., Suthers, D.: Computer-supported collaborative learning: An historical perspective. Cambridge handbook of the learning sciences (2006)
19. Vega-Gorgojo, G., Bote-Lorenzo, M.L., Gómez-Sánchez, E., Asensio-Pérez, J.I., Dimitriadis, Y.A., Jorrín-Abellán, I.M.: Ontoolcole: Supporting educators in the semantic search of CSCL tools. Journal of Universal Computer Science 14(1), 27–58 (2008)
20. Vega-Gorgojo, G., Bote-Lorenzo, M.L., Asensio-Pérez, J.I., Gómez-Sánchez, E., Dimitriadis, Y.A., Jorrín-Abellán, I.M.: Semantic search of tools for collaborative learning with the Ontoolsearch system. Computers & Education 54(4), 835–848 (2010)
21. Villasclaras-Fernández, E.D., Hernández-Leo, D., Asensio-Pérez, J.I., Dimitriadis, Y.: Incorporating assessment in a pattern-based design process for CSCL scripts. Computers in Human Behavior, 25 (l5), 1028—1039 (2009)
22. Zelkowitz, M.V., Wallace, D.: Experimental validation in software engineering. Information and Software Technology, vol 39(11), 735–743 (1997)

Advanced Interface for the Pre-operative Planning and the Simulation of the Abdominal Access in Pediatric Laparoscopy

Lucio Tommaso De Paolis and Giovanni Aloisio

Department of Innovation Engineering,
University of Salento Lecce, Italy
{lucio.depaolis,giovanni.aloisio}@unisalento.it

Abstract. The practice of Minimally Invasive Surgery is becoming more and more widespread and is being adopted as an alternative to the classical procedure. This technique presents some limitations for surgeons. In particular, the lack of depth in perception and the difficulty in estimating the distance of the specific structures in laparoscopic surgery can impose limits to delicate dissection or suturing. The presence of new systems for the pre-operative planning can be very useful to the surgeon. In this paper we present an advanced interface for the pre-operative planning of the surgical procedure and the choice of the abdominal access points in pediatric laparoscopy; using the Augmented Reality technology, these points are overlapped on the real patient's body. Two case studies have been considered for the building of 3D models of the patient's organs from the CT images. The developed application allows the surgeon to gather information about the patient and his pathology, visualizing and interacting with the 3D models of the organs built from the patient's medical images, measuring the dimensions of the organs and deciding the best points to insert the trocars in the patient's body.

Keywords: Augmented Reality, user interface, image processing, surgical pre-operative planning.

1 Introduction

One trend in surgery is the transition from open procedures to minimally invasive laparoscopic operations where the visual feedback to the surgeon is only available through the laparoscope camera and the direct palpation of organs is not possible.

Minimally Invasive Surgery (MIS) has become very important and the research in this field is ever more widely accepted because these techniques provide surgeons with less invasive means of reaching the patient's internal anatomy and allow entire procedures to be performed with only minimal trauma to the patients.

The diseased area is reached by means of small incisions on the body; specific instruments and a camera are inserted through these ports and what happens inside the body is shown in a monitor. The surgeon does not have a direct vision of the organs and so he is guided by the camera images. This surgical approach is very different from the open surgery where the organs can be fully visualized and handled.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 562–571, 2011.

As a promising technique, the practice of MIS is becoming more and more widespread and is being adopted as an alternative to the classical procedure.

The advantages of using this surgical method are evident in the patient because the trauma is reduced, the postoperative recovery is almost always faster and the scarring is reduced.

Despite the improvement in outcomes, these techniques show their limitations for the surgeons. In particular, the lack of the perception of the depth and the difficulty in estimating the distances of the specific organs in laparoscopic surgery can impose some limits on delicate dissection or suturing.

Anyway, the overall risk of complications is of 8.0% in laparoscopy versus 15.2% in laparotomy. Among these, more than 50% of laparoscopic complications occur during the initial entry into the abdomen.

In order to minimize entry-related complications, several techniques and technologies have been introduced in the last years; these include many types of abdominal entries. The choice of a technique more than another depends on the operator experience, the school and the speciality of the surgeon, the laparoscopic upgrading and the work environment. Many surgical techniques are not used yet because some surgeons don't want to change the old method for the one [1].

The laparoscopic access is an alternative to the open entry techniques aiming to prevent visceral and vascular injury due to division of abdominal wall layers. The reasons of a limited use of the open-access method is due to the time needed for performance, the difficulty in maintaining the pneumoperitoneum because of to the gas leakage, and the lack of major evidence for the prevention of intra-abdominal injury by using this method.

The vascular injury during the first laparoscopic access is the first cause of death in laparoscopy, second only to anesthesia and bowel injury, with a reported mortality rate of 15%. Unlike major vascular injuries, where the occurrence and presentation are immediate, many bowel injuries are not recognized at the time of the procedure because of the suboptimal visualization. To overpass the major and minor complications in the laparoscopic access, optically guided trocars are designed to decrease the risk of injury to intra-abdominal structures by allowing the surgeon to visualize abdominal wall layers during the placement [2].

Many research groups, motivated by the benefits that MIS can bring to patients, are now focusing on the development of surgical assistance systems. The progress in technology makes possible the development of systems that can simulate the surgical procedure in a realistic virtual environment [3].

The last developments in medical imaging acquisition and computer systems make possible the reconstruction of 3D models of the organs providing anatomical information barely detectable by CT and MRI slices or ultrasound scan as well as the safe guidance of instruments through the body without the direct sight of the physician.

The emerging Augmented Reality (AR) technology can provide the advantage of a direct visualization in open surgery also in minimally invasive surgery and can increase the physician's view of his/her surroundings with information gathered from patient medical images.

In medicine, Augmented Reality technology makes possible to overlay virtual medical images on the patient, allowing surgeons to have a sort of “X-ray" vision of the body and providing them with a view of the patient’s internal organs. The patient becomes transparent and therefore this virtual transparency makes possible to find tumors or vessels not locating them by touch, but simply visualizing. The virtual information could be directly displayed on the patient’s body or visualized on an AR surgical interface.

In general, AR technology in minimally invasive surgery may be used for training purposes, pre-operative planning and advanced visualization during the real procedure.

Samset et al. [4] present tools based on novel concepts in visualization, robotics and haptics providing tailored solutions for a range of clinical applications. Examples of radio-frequency ablation of liver tumors, laparoscopic liver surgery and minimally invasive cardiac surgery will be presented.

Bichlmeier et al. [5] focus on the problem of misleading perception of depth and spatial layout in medical AR and present a new method for medical in-situ visualization that allows improved perception of 3D medical imaging data and navigated surgical instruments relative to the patient’s anatomy. They describe a method for integrating surgical tools into the medical AR scene in order to improve navigation.

Navab et al. [6] introduce an interaction and 3D visualization paradigm that presents a new solution for using 3D virtual data in many AR medical applications. They introduce the concept of a laparoscopic virtual mirror: a virtual reflection plane within the live laparoscopic video, that allows visualizing a reflected side view of the organ and its interior. A clinical evaluation investigating the perceptive advantage of a virtual mirror integrated into a laparoscopic AR scenario has been carried out.

De Paolis et al. [7] present an Augmented Reality system that can guide the surgeon in the operating phase in order to prevent erroneous disruption of some organs during surgical procedures. Since the simple augmentation of the real scene cannot provide information on the depth, a sliding window is provided in order to allow the occlusion of part of the organs and to obtain the realistic impression that the virtual organs are inside the patient’s body. The distance information is provided to the surgeon and an informative box is shown in the screen in order to visualize the distance between the surgical instrument and the organ concerned.

In this paper we present an advanced platform for the visualization and the interaction with the 3D patient models of the organs built from CT images. The availability of a system for the pre-operative planning can be of great help to the surgeon and this support is even more important in pediatric laparoscopic surgery where a good understanding of the exact conditions and a precise location of the patient’s organs is very important. The pathology can totally change the positions of the organs in the small body of the patient and the standard entry points used in laparoscopic could not be the more appropriate.

The developed application allows the surgeon to choose the points for the insertion of the trocars on the virtual model and to overlap these on the real patient’s body using the Augmented Reality technology.

2 The 3D Models of Patient's Organs

In MIS the use of the registered images of the patient is a prerequisite both for the pre-operative planning and the guidance during the operation. From the medical image of a patient (MRI or CT), an efficient 3D reconstruction of his anatomy can be provided in order to improve the standard slice view; colors associated to the different organs replace the grey levels in the medical images.

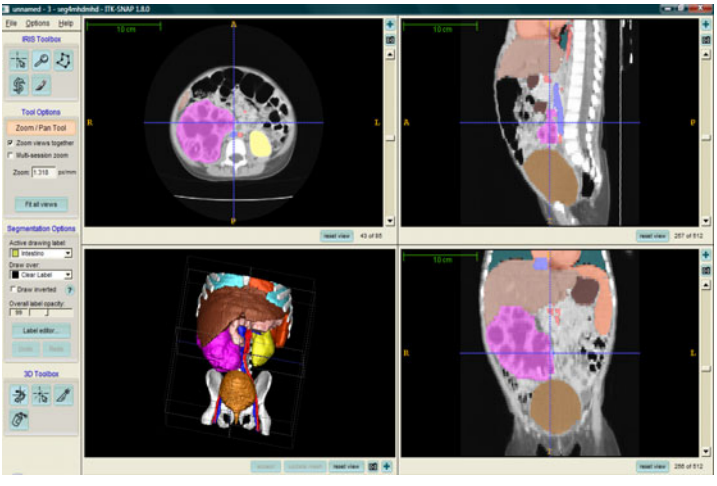


Fig. 1. The result of the image processing using ITK-SNAP

The 3D models of the patient's organs have been reconstructed using segmentation and classification algorithms provided by ITK-SNAP that provides a semi-automatic segmentation using active contour methods, as well as manual delineation and image navigation; it also fills a specific set of biomedical research needs [8].

In our case studies, the slice thickness equal to 3 mm has caused some aliasing effects on the reconstructed 3D models that could lead to inaccuracies. Therefore we have paid special attention to the smoothing of the reconstructed models in order to maintain a good correspondence with the real organs.

Figure 1 shows the result of the image processing using ITK-SNAP.

By means of the developed user interface it is possible to display all the organs of the abdominal region or just some of these using the show/hide functionality; it is also possible to change the transparency of each organ.

We have done the image processing of two different case studies. The first one is a two-year-old child with a benign tumor of the right kidney; the second clinical case is a twelve-year-old child with a tumor of the peripheral nervous system (ganglioneuroma). These two case studies are shown in figure 2 where the skin and the muscles of the abdominal region are displayed in transparency and some organs are hidden in order to better visualize the tumors (in magenta and in orange).

3 The Developed Application

The developed application is supplied with a specific user interface that allows the user to take advantage of the feature offered by the software. The application is provided of 4 sections with the aim to provide support to the surgeons in the different steps of the surgical procedure such as the study of the case, the diagnosis, the pre-operative planning, the choice of the trocar entry points and the simulation of the surgical instruments interaction.

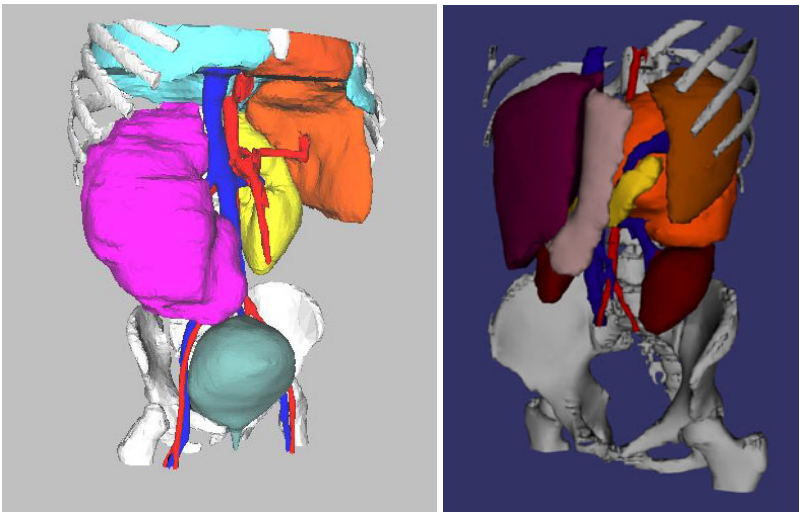


Fig. 2. The 3D models of two case studies

Starting from the models of the patient's organs, the surgeon can notice some data about the patient, collect information about the pathology and the diagnosis, choose the most appropriate positions for the trocar insertion and overlap these points on the patient's body using the Augmented Reality technology.

In this way it is possible to use this platform for the pre-operative surgical planning and during the real surgical procedure too. In addition, it could be used in order to describe the pathology, the surgical procedure and the associated risks to the child's parents, with the aim of obtaining informed consent for the surgical procedure [9].

In the developed application, as shown in figure 3, all the patient's information (personal details, diseases, specific pathologies, diagnosis, medical images, 3D models of the organs, notes of the surgeon, etc.) are structured in a XML file associated to each patient.

A specific section for the pre-operative planning includes the visualization of the virtual organs and the physician can get some measurements of organ or pathology sizes and some distances.

By means of a detailed view of the 3D model, the surgeon can choose the trocar entry points and check if, with this choice, the organs involved in the surgical procedure can be reached and the procedure can be carried out in the best way.

Sometimes, using the standard insertion points for the surgical tools, also a simple surgical procedure can be very difficult due to the specific anatomy of the different patients. The surgeon can have some difficulties to reach the specific organ or the interaction of the surgical tools can be very hard. In this case the surgeon has to choose another insertion point in order to be able to carry out the surgical procedure in the most suitable way. Our aim is to avoid the occurrence of this situation during the real surgical procedure using the visual information provided by means of the 3D models of the patient's anatomy.

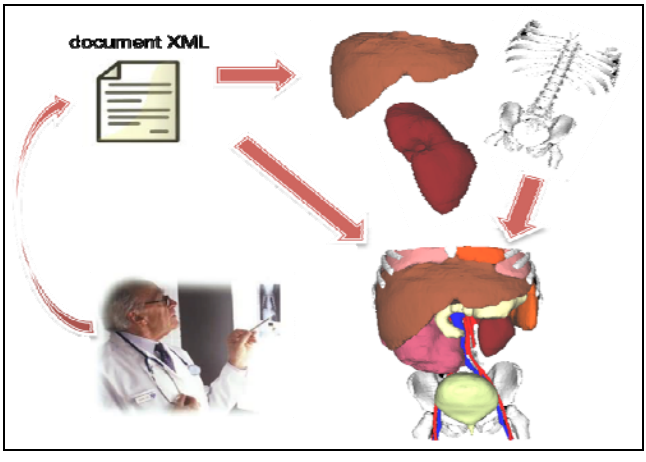


Fig. 3. Patient's data are collected in a XML file

In the developed application, in order to verify if the chosen insertion points allow properly reaching the specific organ interested by the surgical operation and permitting to carry out the procedure in a correct way, it is also possible to simulate the interaction of the surgical instruments; figure 4 shows the specific section of the user interface for the simulation of the surgical tool interaction with the possibility to move the trocar entry points using the arrows.

In the application we also use the AR technology in order to visualize on the patient's body the precise location of the selected points on the virtual model of the patient.

For the augmented visualization, in order to have a correct and accurate overlapping of the virtual organs on the real ones, a registration phase is carried out; this phase is based on fiducial points.

Using the augmented visualization, the chosen entry points for the trocars can be visualized on the patient's body in order to support the physician in the real trocar insertion phase.

Usually an optical tracker is already in the modern operating rooms and provides an important help to enhance the performance during the real surgical procedures. In our application, we have used the Polaris Vicra optical tracker of NDI; the system consists of 2 IR cameras and uses a position sensor to detect retro-reflective markers

affixed to the surgical tools or located on the patient's body; based on the information received from these markers, the sensor is able to determine position and orientation of tools within a specific measurement volume.

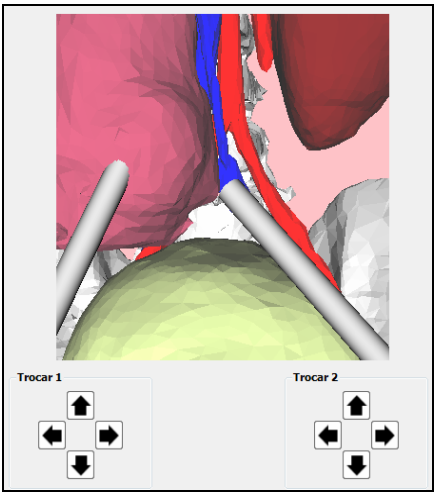


Fig. 4. Simulation of the surgical tools interaction

Taking into account the visual information provided in the AR section of the developed application, it is possible to start the real surgical procedure with an augmented visualization of the chosen trocar entry points on the patient's body.

Figure 5 shows the specific section of the user interface for the choice of the trocar insertion points.

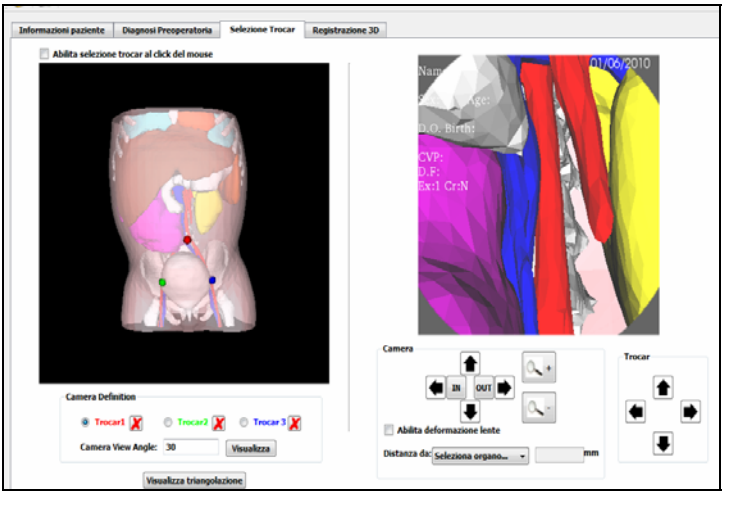


Fig. 5. Choice of the trocar insertion points

4 Usability Tests

In order to evaluate the validity and the usability of the developed application and to receive possible suggestions from the users, some tests have been carried out. The test phase has permitted the users to check all the functionalities of the application.

After a short period (5 minutes) for the training of the application, the users have been tried to carry out the different procedures and, subsequently, they have reported the impressions on a specific questionnaire. 15 subjects have been tested the application for an average time of 7 minutes and 43 seconds.

The obtained results can be considered satisfactory and some annotations to improve the user interface and the usability of the application have been considered. In particular, the users have suggested:

- to improve the session for the choice of trocar entry points by means of a more accurate explication about the use of the arrows in the interface;
- to provide a more simple way to store the measurements of the organs.

Figure 6 shows a graph with the test answers about the usability of the different sessions of the application.

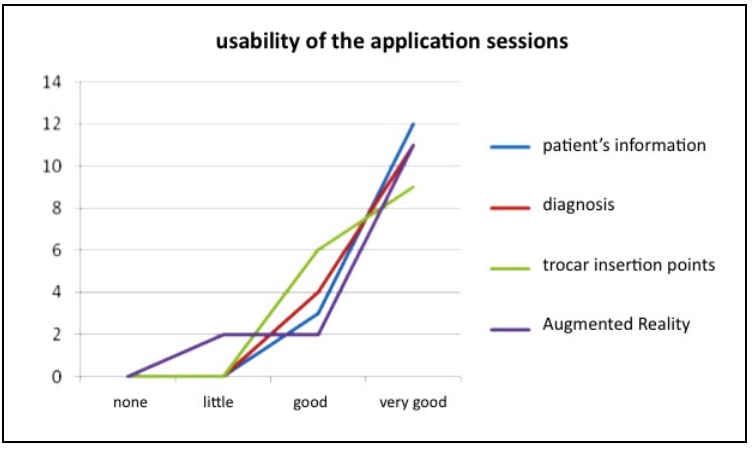


Fig. 6. Test answers about the usability of the application sessions

5 Conclusion and Future Work

The developed application offers a tool to visualize the 3D reconstructions of the patient's organs, obtained by segmentation of a CT slices, and to simulate the placement of the trocars in order to verify the correctness of the insertion sites. A complete user interface allows a simple and efficient utilization of the developed application.

Furthermore the system retains patient and pathology information that the surgeon can insert and includes an Augmented Reality module that supports the placement of

the trocars on the patient's body during the real surgery procedure. An accurate integration of the virtual organs in the real scene is obtained by means of an appropriate registration phase based on fiducial points.

The developed platform can support the physician in the diagnosis steps and in the pre-operative planning when a laparoscopic approach will be followed. This support could also lead to a better communication between physicians and patient's parents in order to obtain their informed consent.

The platform has been tested on study cases already operated by the surgeon; the future work will be the validation of the developed application on a new study case by following all the steps from the diagnosis to the pre-operative planning and to the first phase of the real surgical procedure.

The building of a new Augmented Reality system that could also help the surgeon during the other phases of the surgical procedure has been planned as future work.

The acquisition in real time of a video of the patient and the development of an application able to overlap the virtual organs on the real patient's body is in progress; the changing of the surgeon points of view and the positions of medical instrument will be taken into account.

Acknowledgements

This work is part of the ARPED Project (Augmented Reality Application in Pediatric Minimally Invasive Surgery) funded by the Fondazione Cassa di Risparmio di Puglia. The aim of the ARPED Project is the design and the development of an Augmented Reality system that can support the pediatric surgeon through the visualization of anatomical structures of interest during the pre-operative planning and the laparoscopic surgical procedure.

References

1. Ahmad, G., Duffy, J.M.N., Phillips, K., Watson, A.: Laparoscopic Entry Techniques. Cochrane Database of Systematic Reviews, (2), Art. No.: CD006583 (2008)
2. Tinelli, A., Malvasi, A., Hudelist, G., Istre, O., Keckstein, J.: Abdominal Access in Gynaecological Laparoscopy: a Comparation between Direct Optical and Open Access. Journal of Laparoendosc & Advanced Surgical Techniques 19(4), 529–533 (2009)
3. Soler, L., Nicolau, S., Fasquel, J.-B., Agnus, V., Charnoz, A., Hostettler, A., Moreau, J., Forest, C., Mutter, D., Marescaux, J.: Virtual Reality and Augmented Reality Applied to Laparoscopic and NOTES Procedures. In: IEEE 5th International Symposium on Biomedical Imaging: from Nano to Macro, pp. 1399–1402 (2008)
4. Samset, E., Schmalstieg, D., Vander Sloten, J., Freudenthal, A., Declerck, J., Casciaro, S., Rideng, Ø., Gersak, B.: Augmented Reality in Surgical Proceduresn. SPIE Human Vision and Electronic Imaging XIII (2008)
5. Bichlmeier, C., Wimmer, F., Michael, H.S., Nassir, N.: Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In: Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007), pp. 129–138 (2007)

6. Navab, N., Feuerstein, M., Bichlmeier, C.: Laparoscopic Virtual Mirror - New Interaction Paradigm for Monitor Based Augmented Reality. In: IEEE Virtual Reality Conference 2007 (VR 2007), Charlotte, North Carolina, USA, pp. 10–14 (2007)
7. De Paolis, L.T., Pulimeno, M., Lapresa, M., Perrone, A., Aloisio, G.: Advanced Visualization System Based on Distance Measurement for an Accurate Laparoscopy Surgery. In: Joint Virtual Reality Conference of EGVE - ICAT - EuroVR, Lyon, France(2009)
8. Yushkevich, P.A., Piven, J., Cody, H., Ho, S., Gee, J.C., Gerig, G.: User-Guided Level Set Segmentation of Anatomical Structures with ITK-SNAP. Insight Journal, Special Issue on ISC/NA-MIC/MICCAI Workshop on Open-Source Software (2005)
9. Bollschweiler, E., Apitzsch, J., Obliers, R., Koerfer, A., Mönig, S.P., Metzger, R., Hölscher, A.H.: Improving informed consent of surgical patients using a multimedia-based program? Results of a prospective randomized multicenter study of patients before cholecystectomy. Ann Surg. 248(2), 205–211 (2008)

An Augmented Reality Application for the Radio Frequency Ablation of the Liver Tumors

Lucio T. De Paolis[1], Francesco Ricciardi[2], Aldo F. Dragoni[3], and Giovanni Aloisio[1]

[1] Department of Innovation Engineering, University of Salento Lecce, Italy
[2] Engineering Faculty, Polytechnic University of Marche, Ancona, Italy
[3] Department of Ingegneria Informatica, Gestionale e dell'Automazione, Polytechnic University of Marche Ancona, Italy

Abstract. Hepatic cancer is one of the most common solid cancers in the world. As surgery of hepatic cancer is seldom applicable, different solutions have been found to cure this disease. One of these is Liver Radiofrequency Ablation. This technique consists in a needle insertion inside the liver parenchyma in order to reach the tumor and in an injection of a radiofrequency current to cause tumor cell necrosis for hyperthermia. The needle placement task is really difficult because surgeon uses ultrasound, CT or MRI two-dimensional image to guide the needle. In this paper we present an Augmented Reality system to help the surgeon to place the needle as best as possible; the application can also help the surgeon during the preoperative planning because it offers various visualization modality of 3D models of the patient's organs obtained from the medical images.

Keywords: Liver radiofrequency ablation, Augmented Reality, image-guided surgery.

1 Introduction

Hepatic cancer is one of the most common solid cancers in the world. Hepatocellular carcinoma (HCC) is the most common primary hepatic cancer. Unless primary hepatic cancer is quite rare, liver is second only to lymph nodes as a common site of metastasis from other solid cancers [1]. The liver is often the site of metastatic disease, particularly in patients with colorectal adenocarcinoma.

The use of chemotherapy for malignant form of liver cancer rarely led to good results in long-term survival rate. We have also to consider that chemotherapy produces negative effects in the lifestyle of the patient.

Today surgery is the best approach to avoid the death of the patient and the reversion of hepatic cancer. Unfortunately only from 5 to 15 per cent of HCC or hepatic metastasis patients diagnosed again undergo a potentially curative resection of the liver cancer [2], [3].

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 572–581, 2011.

Patients with confined disease of the liver could not be candidates to resection because of multifocal disease, proximity of tumor to key vascular or biliary structures that preclude a margin-negative resection potentially unfavourable in case of presence of multiple liver metastases. Very often the tumor is also associated to a pre-existent cirrhosis that can reduce resection margins.

Liver transplant is the only radical therapy that eliminates the risk of recurrence but it can't be always used. So, since most of patients with primary or malignancies confined metastatic at the liver are not candidates for surgical resection, we should use different approaches to control and potentially cure liver diseases.

Among these we focused our attention on Liver Radiofrequency Ablation (RFA), a technique used since 1980's consisting in the placement of a needle inside the liver parenchyma to reach the centre of the tumor lesion.

When the lesion is reached, an array of electrodes is extracted from the tip of the needle ant it is that expanded in the tumor tissue. From these electrodes is injected for a certain time a radiofrequency current in the tumor tissue that causes tumor cell necrosis for hyperthermia (the local temperature is higher than 60 °C and cancer cells are more sensitive to heat than normal cells).

One problem in using radiofrequency tumor ablation technique is the correct placement of the needle that should reach the tumor lesion. Today surgeons use ultrasound, CT or MNR acquired during the needle placement in order to correctly direct the needle to the tumor. The use of these two-dimensional images makes the procedure very difficult and requires sometimes more than one insertion.

To reduce the complexity of the needle insertion can be used the Augmented Reality (AR) technology. With the superimposition of the virtual models of the patient's anatomy (liver, cancer, etc) exactly where are the real ones, it is possible to make the needle placement less difficult. In this way the surgery patient's risks and the surgery time should be reduced.

In this paper we present an Augmented Reality application to support surgeons for the needle placement procedure in the treatment of the liver tumors; the application can also provide a support during the pre-operative surgery planning.

2 Previous Works

The first Augmented Reality guidance system for liver RFA is presented by Nicolau et al. in [4]. This system realizes a real-time superimposition of virtual models, reconstructed from the CT, over the patient's abdomen in order to have a three dimensional view of the internal anatomy. The research team has solved the problem to realize a correct augmentation and developed some automatism to create the guidance system.

To realize the scene registration they have used circular radio-opaque markers placed over the patient's skin; in this way the position of this fiducial points is well known. To measure the position of the markers in the patient reference system and the needle position they have used an image analysis based algorithm over the frame acquired by calibrated cameras. An automated algorithm, based on image-extracted information, is used for the registration phase.

To track the needle position in the space they have used a rectangular marker. The system test is carried out on a dummy and the system uncertainty in reaching the target is around 3 mm. Using this system they have noted a drastic reduction of the surgery time.

The same authors have used the developed system in an operating room [5]. The purpose of this study is the determination of the usability of the developed system in operating room. The authors have developed a passive evaluation protocol in order to evaluate the system uncertainty after a traditional insertion made by surgeon. The system guidance accuracy is evaluated by measuring the distance between the real tip position using traditional imaging techniques and the virtual target using the developed system.

Using this approach system uncertainty in operating room is around 9.5 mm. This uncertainty is acceptable for large tumor lesions (wider than 30 mm) but is unacceptable for smaller lesions. The authors have noted that this uncertain value is due to liver motion caused by patient's breathing during surgery. Another reason could be the use of a part of the total marker placed over the patient's abdomen; this because parts of them were covered by surgical instruments and placed not too close the needle entry point for sterility reasons.

To reduce the influence of the liver breathing motion the authors have tried to synchronize the system guidance information and the CT acquisition with the expiratory phases of the patient's breathing cycle [6].

To place more markers in proximity of the entry points they have used ring radio-opaque markers that are removed after the CT acquisition; its centre was marked using indelible ink. The uncertainty reached in an operating room by means of this technique and measured during the expiratory phase is less than 5 mm.

3 Used Technologies

For the 3D model generation task we used 3D Slicer toolkit [7], a multi-platform open-source software that offers many features for the task of visualization and scientific analysis of patient's imaging data. The platform provides functionality for segmentation, registration and three-dimensional visualization of multi-modal image data, as well as advanced image analysis algorithms for diffusion tensor imaging, functional magnetic resonance imaging and image-guided therapy.

Standard image file formats are supported, and the application integrates interface capabilities with biomedical research software and image informatics frameworks.

To obtain the 3D model of the anatomical structures of interest (liver, liver's vein and artery, cancer and thoracic cage) we have used various semi-automatic algorithm. From the results of the segmentation we have created the 3D model using the Model Maker module that uses the Marching Cubes algorithm to produce the model. In Figure 1 is shown the 3D model of some patient's organs; the tumor is visualized in green.

The resulting mesh are exported in VTK format and then converted in MSH using an external script written in C++.

Unlike existing literature approach that uses image analysis techniques to provide registration and tracking tasks, we have used an optical tracking system to measure

the point's position and the surgical tool position and orientation in space. The used tracking system is the Polaris Vicra of NDI [8]; this optical tracker tracks both active and passive markers and provides precise, real-time spatial measurements of the location and orientation of an object or tool within a defined coordinate system.

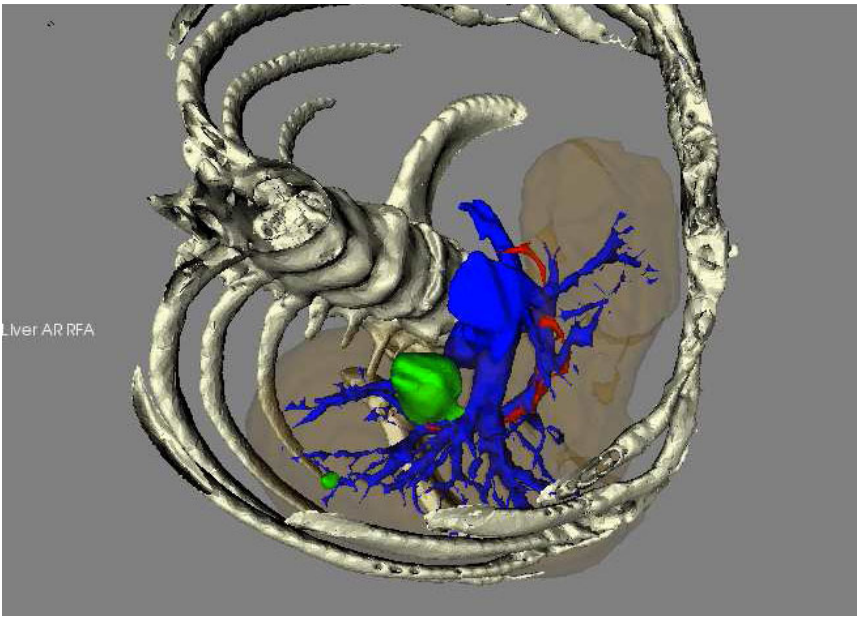


Fig. 1. 3D model of the patient's organs

The system consists of 2 IR cameras and uses a position sensor to detect infrared-emitting or retro-reflective markers affixed to a tool or an object; with the help of the information received from the markers, the sensor is able to determine position and orientation of tools within a specific measurement volume. The tracker can calculate the current position of the tool in the space with an accuracy of 0.2mm and 0.1 tenth of a degree.

For the visualization and image processing we have used the IGSTK library [9], [10]. IGSTK (Image-Guided Surgery Toolkit) is a set of high-level components integrated with low-level open source software libraries and application programming interfaces. IGSTK provides several functionalities as the ability to read and display medical images and the possibility to interface to common tracking hardware.

IGSTK includes ITK (Insight Segmentation and Registration Toolkit), an open source software system for 3D computer graphics, image processing, and visualization and VTK (Visualization Toolkit), an open-source software system that employs leading-edge segmentation and registration algorithms in two, three, and more dimensions.

The graphical interface has been realized using FLTK (Fast Light Toolkit) library [11].

The interaction with the optical system is managed directly by IGSTK and therefore it was not necessary to incorporate an external library to use the Polaris Vicra tracker.

4 The Developed Application

The purpose of our development is to provide an application that can help the surgeon during the needle insertion in liver RFA. This application can also help the surgeon during the surgical pre-operative planning.

To achieve this purpose we have developed the application focusing our attention on the user interface design and software usability.

The use of the Augmented Reality technology can offer to the surgeon an innovative way to approach the needle insertion. Using this technology he has an innovative visualization of the patient's body, as a sort of 3D interactive X-ray view of the patient's organ.

To realize a guidance system that can help the surgeon during the ablation surgery we use a virtual tool that models the real needle and that reaches the tumor in the virtual 3D model. To overlap the virtual tool over the real one we track the position and orientation of the real tool using some reflective spheres detected by the used optical tracker. In this way the position and the orientation of the real tool are measured and used in IGSTK to set the position and the orientation of the virtual tool.

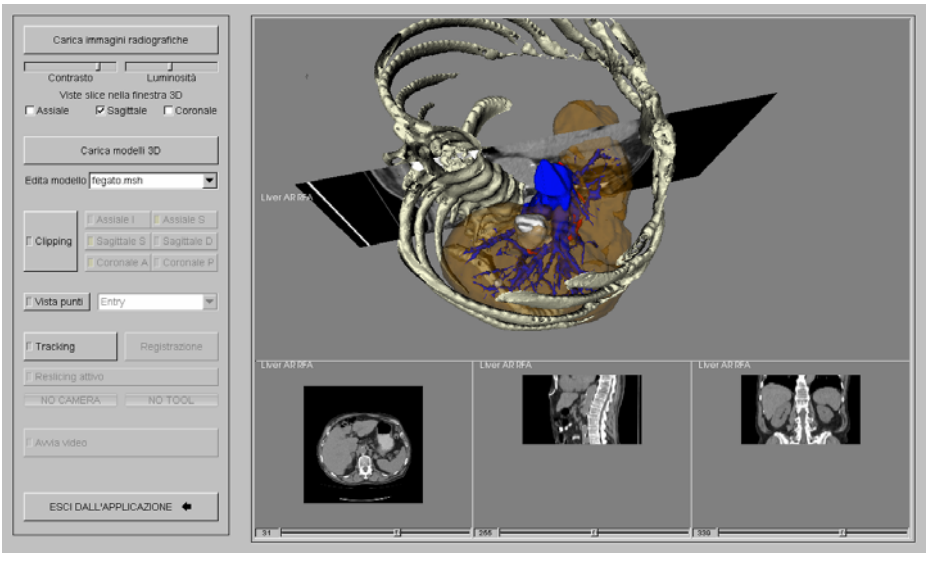


Fig. 2. User interface with transparency effects applied to the liver

To achieve a correct augmentation is necessary to have a perfect correspondence between the virtual organs and the real ones. This task is very difficult in an image-guided surgery application because a very small error in the registration phase can cause a serious consequence for the patient.

The registration process is carried out just before the start of the surgical procedure; the optical tracker by means of a tool placed on the patient detects some possible movements of the patient's body over the operating table.

The applied method is based on the placement of 3 fiducial points on the patient's body before the CT scanning and the detection of these in the 3D model built from the acquired medical images. This detection is achieved placing three virtual points in correspondence of the fiducial points inside the medical image using the application interface.

After the placement of these points the surgeon can measure the position of the corresponding points over the patient using the tracker probe. When all the points coordinate have been acquired, a rigid transformation between the real reference system and the virtual one is determined using the Horn algorithm [12].

4.1 User Interface and Application Features

The user interface is designed to be simple and functional at the same time. In the left side of our interface we have put together the application control. In the right-top window we show the 3D model and the augmented reality scene and in the right-bottom the three smaller windows where we placed the axial, coronal and sagittal view of CT images. We have decided to show always this information to provide a view that is very close to the usual surgeon approach.

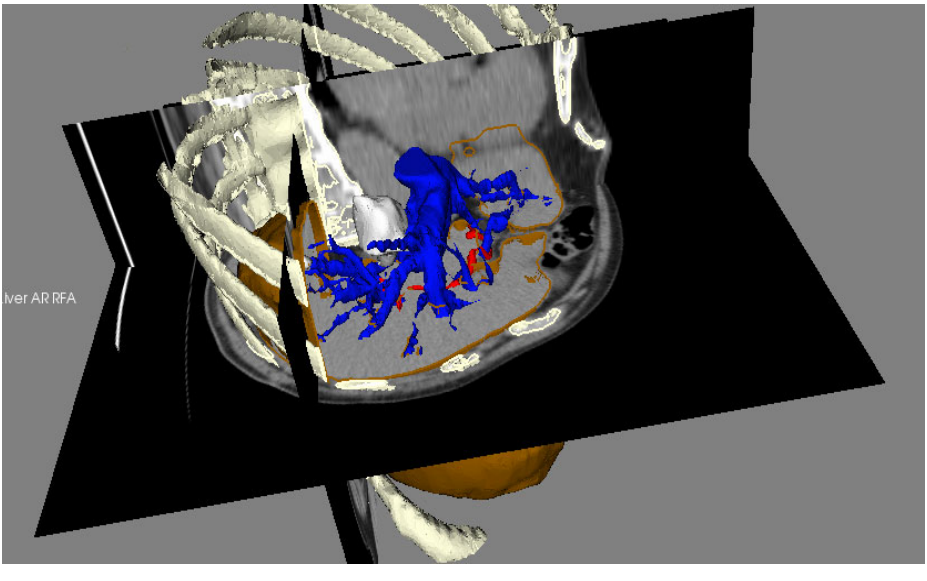


Fig. 3. Clipping visualization modality applied to the liver and thoracic cage

The application controls are organized in a sequence that follows the ideal surgeon workflow. In the top of the stack we have placed the buttons to load and control CT images. The user can load only a sequence at a time and can choose to turn on or off the per-axis slice visualization in the 3D visualization window if this visualization can interfere with the model view.

There are also some control buttons to load the model and a menu list in order to set their properties. In the mesh properties panel the surgeon can modify the colour and the opacity of the mesh in accordance with his preferences. In addition he can set the mesh visibility check and clipping enable check (ON by default). Modifying the mesh opacity the surgeon can have a better and more natural visualization to study the body's internal structure (Figure 2).

The clipping function permits the surgeon to dissect the model and to study its internal structure in an alternative way from mesh opacity modification (Figure 3). The dissection could be made along the three principal axis and six button (two for every axis) can be used to select the view direction.

The user can choose to deactivate the clipping for single mesh in the mesh properties panel. In this way a new visualization modality in which is shown unclipped mesh inside clipped other ones is offered to the users.

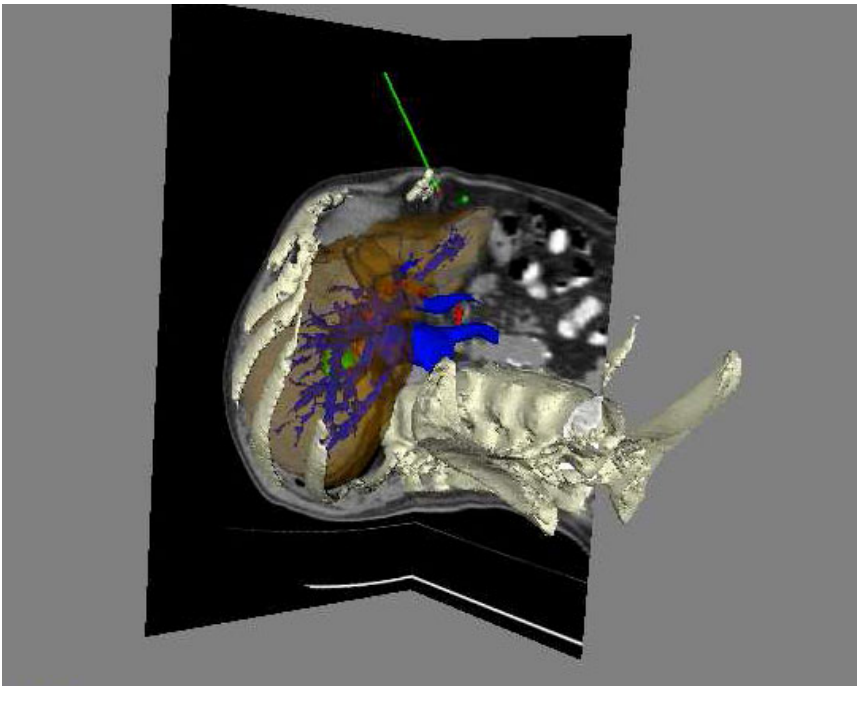


Fig. 4. Example of reslicing

Below the clipping control buttons there are the “Point View” button and the selection menu. These controls are used to select and show some specific points in slice and 3D View. Two of these specific points are the entry and the target points and their positions is chosen by the surgeon during the pre-operative planning.

A line connects these two points in order to simulate the path followed by the needle from the entry point to the target. In this way surgeon can see if the desired path of the needle can touch important vein and can cause bleeding during the needle insertion.

The application features described till now are part of what we consider the pre-operative planning task. During this task the surgeon can use the application to study the pathology in a more simple and natural way than that provided by simple CT slice visualization. Another advantage in using the application is that the surgeon can choose the best way to realize surgery with minimum side effects.

For the navigation and augmentation task the surgeon need to connect the optical tracker and carry out the registration task. When the registration process is complete, the virtual tool is shown in 3D view and is coupled with the real one so it follows every movement that real tool makes.

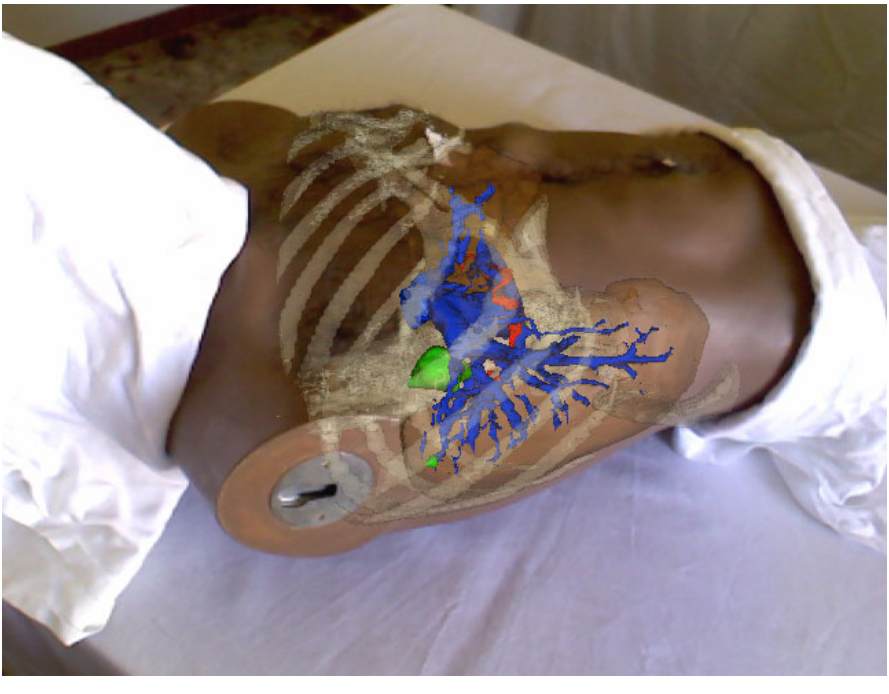


Fig. 5. Augmented Reality visualization

The surgeon can select the “Active Reslice” button to activate the reslicing function. Using this function the CT slices are automatically repositioned along the three principal axes where is positioned the needle tip and, in this way, the surgeon,

can have an accurate visualization of the 3D model during the minimally invasive surgical procedure; in real time the CT slices are visualized exactly next to the actual position of the surgical instrument.

In Figure 4 is shown an example of reslicing where the virtual tool (in green) is detected by means of the optical tracker and is overlapped on the real one.

The last provided feature is the augmentation of reality overlapping the 3D virtual model over the patient body (Figure 5).

5 Conclusions and Future Work

In this paper we present a guidance system for needle placement in liver radiofrequency ablation. Using the developed system surgeon can experiment a more easy way to insert the needle using the Augmented Reality technology.

The software offers to surgeon a set of tool to study the patient's pathology during preoperative planning task. He can set the model transparency to have an internal view of the organ structure. Using the clipping tool surgeon can dissect the organ 3D model and study the internal structure in an alternative way from simple transparency regulation mode.

Till now the application has been tested on a dummy. As future work we are planning to measure the system guidance accuracy and to compare it with the existing results in literature. We want to reduce uncertainty values in vivo practice and, for this reason, we are thinking to develop a model that can take into consideration the liver motion due to the breathing.

Acknowledgments. The authors wish to thank Pierluigi Di Sebastiano and Matteo Scaramuzzi of the Hospital "Casa Sollievo della Sofferenza" (San Giovanni Rotondo, Italy) for the surgical advice in the development of the application.

References

1. Weiss, L., Grundmann, E., Torhorst, J., et al.: ZIB Structure Prediction Pipeline: Hematogenous metastatic patterns incolonic carcinoma: an analysis of 1541 necropsies. Journal of Pathology 150, 195–203 (1986)
2. Liver Cancer Study Group of Japan. Primary liver cancer in Japan: clinicopathologic features and results of surgical treatment. Ann. Surg. 211 (1990)
3. Nagorney, D., Van Heerden, J., Ilstrup, D., et al.: Primary hepatic malignancy: surgical management and determinants of survival. Surgery 106, 740–748 (1989)
4. Nicolau, S., Garcia, A., Pennec, X., Soler, L.: An augmented reality system to guide radio-frequency tumor ablation. Journal of Computer Animation and Virtual World 16(1), 1–10 (2005)
5. Nicolau, S.A., Pennec, X., Soler, L., Ayache, N.: A complete augmented reality guidance system for liver punctures: First clinical evaluation. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 539–547. Springer, Heidelberg (2005)
6. Nicolau, S., Garcia, A., Pennec, X., Soler, L., Buy, X., Gangi, A.: Ayache. N., Marescaux, J.: An augmented reality system for liver thermal ablation: Design and evaluation on clinical cases. Elsevier, Amsterdam (2009)

7. 3D Slicer at, http://www.slicer.org
8. NDI Polaris Vicra at, http://www.ndigital.com
9. Clearya, K., Ibanez, L., Ranjan, S., Blake, B.: IGSTK: a software toolkit for image-guided surgery applications. In: Conference on Computer Aided Radiology and Surgery (CARS 2004), Chicago, USA (2004)
10. Cheng, P., Zhang, H., Kim, H.: IGSTK: Framework and Example Application Using an Open Source Toolkit for Image-Guided Surgery Applications, Washington, DC, USA
11. Fast Light Toolkit at, http://www.fltk.org
12. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America 4, 629 (1987)

Virtual Reality and Hybrid Technology for Neurorehabilitations

Alessandro De Mauro[1], Aitor Ardanza[1], Chao Chen[1], Eduardo Carrasco[1], David Oyarzun[1], Diego Torricelli[2], Shabs Rajasekharan[1], Josè Luis Pons[2], Ángel Gil-Agudo[3], and Julián Flórez Esnal[1]

[1] Visual Interaction Communication Technologies Centre, Vicomtech, San Sebastian, Spain
[2] Bioengineering Group, CSIC, Madrid, Spain
[3] Biomechanics Unit, National Hospital of Paraplegics, Toledo, Spain
{ademauro,aardanza,cchen,doyarzun,srajasekharan,
jflorez}@vicomtech.org,
{torricelli,jlpons}@iai.csic.es,
amgila@sescam.jccm.es

Abstract. Disabilities that follow Cerebrovascular accidents (CVA) and spinal cord injuries (SCI) severely impair motor functions and thereby prevent the affected individuals from full and autonomous participation in daily activities. Several studies have shown that virtual reality (VR) is a technology suitable for rehabilitation therapy due to its inherent ability of simulating real–life tasks while improving patient motivation. In this paper we present our research focuses on the development of a new rehabilitation therapy based on a VR system combined with wearable neurorobotics (NR), motor-neuroprosthetics (MNP) and brain neuro-machine interface (BNMI). This solution, based on hybrid technology aims to overcome the major limitations of the current available therapies. This paper is focused on the Virtual Reality concepts used for the development of the HYPER rehabilitation system.

Keywords: virtual reality, motor-neuroprosthetics, brain neuro-machine interface, neuro-robotics, cerebrovascular accidents, spinal cord injury.

1 Introduction

Cerebrovascular accidents (CVA) and spinal cord injuries (SCI) are the most common causes of paralysis and paresis with reported prevalence of 12,000 cases per million and 800 cases per million, respectively. Disabilities that follow CVA or SCI severely impair motor functions (e.g., standing, walking, reaching and grasping).

Disabilities that follow CVA and SCI are:

1. Tetraplegia refers to the loss of motor and/or sensory function in the cervical segments of the spinal cord (SCI). It results in an impaired function of the arms, trunk, legs and pelvic organs.

2. Paraplegia refers to the loss of motor and/or sensory function in thoracic, lumbar or sacral segments (SCI). Consequently, the arm function is spared, but the trunk, legs and pelvic organs can be affected.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 582–591, 2011.

3. Hemiplegia is paralysis of the side of the body occurring after a CVA. In many cases it comprises weakness of the leg on the affected side, where the drop-foot syndrome often prevents walking.

The main goal of neurorehabilitation is to favor the relearning process of the Central Nervous System (CNS) in the execution of coordinated movements.

The outcome of the neurorehabilitation therapy depends on two main issues [1, 2]:

- the quality and amount of physical activity performed by the patient;
- the active participation of the patient in the rehabilitation process (in other words: motivation).

Physical therapy aims to strengthen the active muscles in several parts of the body. Occupational therapy is specialized in training individuals who have lost muscle strength or coordination to relearn the tasks of daily living, such as eating, dressing, and grooming.

2 Virtual Reality and Rehabilitation

VR environments can provide realistic training for the patient in different scenarios and phases of the rehabilitation. By using VR in conjunction with Human Computer Interfaces (HCI) the training of daily life activities can be much improved in terms of time and quality. This approach permits a realistic and ergonomic training in a safe, interactive and immersive environment.

Examples of interfaces able to interact with VR are mice, joystick, haptic interfaces with force feedback and motion tracking systems.

Repetition is crucial for the re-learning of motor functions and for the training of the cortical activity. This task has to be connected with the sensorial feedback on every single exercise.

Patient motivation is fundamental because active cooperation of the patient is needed to achieve a more functional outcome of the therapy. Motivation can be improved by assigning a serious game format to the therapy. In this way the training activity becomes more attractive and interesting [3, 4]. Moreover, VR shows another advantage: the possibility to be precisely adapted to the patient's therapy and to be specific for each rehabilitation phase. In addition, it represents a precise tool for the assessment of the therapy during each session. The (tracked/saved) data can be used by the rehabilitation specialists for monitoring and managing the therapy [5]. Several researches have shown that, during VR rehabilitation, the movements are very similar to those used in the traditional therapy. Although they appear a bit slower and less accurate, [6, 7] show that they are anyway appropriate for rehabilitation. Finally, [8] have proven good results in executing the movements trained in VR in reality.

Some of the significant studies on the application of robotics for rehabilitation purposes shall be introduced briefly. The Rutgers Arm [9] is one of the first prototypes composed of a PC, a motion tracking system and a low-friction table for the upper extremity rehabilitation. The system has been tested on a chronic stroke

subject and has shown improvements in arm motor control and shoulder range of motion (Fugl-Meyer [10] test scores). The same group has developed the Rutgers Ankle for the lower extremity rehabilitation. It is a haptic/robotic platform, which works with six degrees of freedom, driving the patient's feet movements (Fig. 1, left).

Fig. 1. Successful applied examples of technologies for rehabilitation. Up: Rutgers Ankle (on the left) and Lokomat® (on the right). Down: Armeo®.

In [11], the Rutgers Ankle system has been tested. As a result, the group of patients trained with the robotic device coupled with the VR demonstrated greater changes in velocity and distance than the group trained with the robot alone.

Most of the gait rehabilitation systems currently used for therapy are based both on treadmills and body weight support.

The state-of-art in rehabilitation using virtual reality (VR) and robotics is provided by Lokomat® and Armeo® (from Hocoma) for the lower and the upper extremity, respectively (Fig. 1 center and right). These two systems are validated by the medical community and used in several rehabilitation centers [12].

3 Robotic-Based Neurorehabilitation

Rehabilitation and functional substitution of motor functions is still a very active research area. In the last decade, a number robot-assisted rehabilitation systems have been developed in order to support and improve the therapist's action by delivering intensive physical therapy and providing objective measures of the patient's performance [13, 14, 15].

There is no consensus on what are the most adequate robotic-based intervention for rehabilitation of motor disorders [16, 17]. Nevertheless, some key factors for successful robotic-assisted therapy can be identified:

1) *Active role of the patient.* Brain activity plays a fundamental role on the modulation of the neural mechanisms that generate movement [18, 19]. Passive, repetitive training is very likely to be suboptimal, as it leads to the phenomena of "learned helplessness.

2) *Motivation.* Motivation is one of the most important factors in rehabilitation and it is commonly used as a determinant of rehabilitation outcome [20], since it is strongly correlated with the degree of patient's activity. User's motivation can be achieved by means of various different types of feedback and modes of interaction, so influencing the motor re-learning process at different levels [21].

3) *Assist as needed.* In order to imitate the action of the physical therapist in supporting the movement of the limb, the new-generation of robotic systems have been provided with the so called Assist-as-needed (AAN) paradigm [22]. 4) Challenge. Contrary to the assistive techniques, which help the user to reach the task, the challenge-based robotic strategies aim at opposing to the user's intention of movement, using resistance or error-amplification strategies.

5) *Biofeedback.* Biofeedback is a crucial factor for success of the therapy as it informs about the patient's degree of activity and is a key to maintain and encourage the motivation and increasing the active participation of the patient. Currently, biofeedback rehabilitation relies mainly on a single source of information, i.e. force-based feedback. By combining other forms of feedback beside the pure force feedback, such as brain activity (EEG), muscular activity (EMG) and visual information on limb motion, a more accurate and effective outcome might be achieved [23].

6) *Bioinspiration.* Due to the close cooperation between human and robot, it is necessary to know the properties of the human motor system in order to define the design requirements of a rehabilitation device. With the help of a biological model it is possible to predict the system's behavior and optimize the robotic intervention, in terms of adaptability, functionality and energy consumption [24].

Exoskeletal Robots (ERs) are person-oriented robots, operating alongside human limbs to supplement the function of a limb or to replace it completely [25]. MNPs constitute an approach to restoring function by means of artificially controlling human muscles or muscle nerves with Functional Electrical Stimulation (FES).

The integrated application of ERs and MNPs can give appropriate tools for dealing with the above stated key aspects of robotic-based rehabilitation. The hybrid combination of physical and bio-electrical actions on human body can effectively recover the impaired human motor control mechanisms, in both rehabilitation and functional compensation scenarios. The orchestration of ERs, MNPs and latent motor capabilities involves several issues, principally related to the cognitive aspects of

human-machine interaction. In a successful scenario, the control signals provided by the patient must be interpreted correctly semantically and temporally by the machine in order to provide the mechanical power and electrical stimulation required to carry out the task [26, 27, and 28].

The main challenge of cognitive human-machine interaction is the development of a multimodal system capable of deciphering user's volitional commands in a robust manner and integrating them with the ER-MNP control systems. This cognitive processes needs a deep understanding of the relation between cognition (the process comprising high level functions carried out by the human brain, including perception, comprehension, construction, planning, self-monitoring) and the motor control. Several signals of different typologies must be analyzed to convey meaningful feed-forward and feedback information. These signals are related to muscular activity (EMG), cerebral activity (EEG), visual and auditory perception, tactile and proprioceptive stimuli.

Brain and Neural to Machine Interfaces (BNMIs) has been recently proposed [29] as an effective multimodal interface to the humans' neural system. BNMIs are gaining momentum as a method to command the exoskeleton-based rehabilitation, since it might constitute new means to improve user-centered strategies for robotic-based training. BNMIs have the potential to improve controllers for movement training by demanding neural control within the involved cortical network, by relying on: (1) passive monitoring, which might assess user's motor intention; (2) information derived from the peripheral nervous system, such as reflex actions that might directly trigger muscle activity; and (3) indirect measures of neural activity (such as EMG).

4 Cerebrovascular Accidents and Spinal Cord Injury Rehabilitation Using Virtual Reality and Hybrid Technology

None of the systems described in paragraph III proposes VR in conjunction with a hybrid and wearable MNP-NR system.

The HYPER project collects different researches in neurorobotics (NR) and motor neuroprosthetics (MNP) both for rehabilitation and functional compensation of motor disorders. The project focuses its activities on new wearable NR-MNP systems that will combine biological and artificial structures in order to overcome the major limitations of the current rehabilitation solutions to Cerebrovascular Accident (CVA) and Spinal Cord Injury (SCI).

The main targets of the HYPER project (Hybrid Neuroprosthetic and Neurorobotic devices for Functional Compensation and Rehabilitation of Motor Disorders) are:

1. to speed up the rehabilitation procedures
2. to improve the outcome of the therapy using new paradigm and technology.

Those results shall be achieved by an integrated use of different means of sensing and actuation. A multimodal Brain Neural Machine Interface (BNMI) is applied to enhance the cognitive interaction and to drive a hybrid NR-MNP system.

Using a multi-channel acquisition approach, the user's outputs (EEG, EMG, kinetic and kinematic information) serve as inputs to the controller of the hybrid platform. The controller, as in the natural human control system, (re)presents a feed-forward component based on predetermined motion and biomechanical models, and a

reactive controller that mimics human neuromotor mechanisms and reflexes. In addition to the limb actuation systems (NR/MNP), a virtual reality system generates visual/auditory feedback to increase the user's involvement and immersion, potentiating the cognitive interaction. Both upper and lower parts of the patient body are assisted. The main emphasis is put on restoring daily life activities.

Users' groups have been identified in order to adjust therapy and system components to the various therapy needs. Several scenarios have been developed and elaborated in detail. Each of them includes some or all the components (NP, NR, and VR). The therapy is subdivided in several states from the moment in which the injury happens until the state in which the patient is fully rehabilitated.

In the traditional rehabilitation therapy, different modes of exercises are used:

• Aerobic. Large-muscle activities (eg, walking, treadmill, stationary cycle, combined arm-leg ergometry, arm ergometry, seated stepper).
• Strength. Circuit training, weight machines, free weights, isometric exercise.
• Flexibility. Stretching.

Ranges of movements that are significant in the rehabilitation for CVA or SCI patients have been identified by medical doctors. Any of the daily life functions constitutes a combination of these defined movements.

In the specific, the upper body joints (and related movements) are:

• shoulder (flexion, extension, abduction, adduction, outward medial rotation, inward medial rotation);
• elbow (flexion, extension, pronation, supination);
• wrist (flexion, extension, abduction, adduction)

Similarly, the lower body joints (and related movements) are:

• hip (flexion, extension, abduction, adduction, medial and lateral rotation);
• knee (flexion, extension);
• ankle (plantar flexion, dorsal flexion, inversion and eversion).

For each of them, both degrees and ranges of movement have been specified in order to assess the patient's skills during the rehabilitation process and to parameterize the rehabilitation training based on VR.

During the initial period of our research, patient's movements were tracked by a motion tracking system based on radio frequency. Transmitters were positioned on each joint. This solution offers good tracking performances but it suffers from the use of many cables. Considering the patient's needs it is therefore not an optimal solution. Currently we are using Kinect to provide a marker-less tracking which offer promising preliminary results. The tracking software is based on OpenNI which is a open source library. We are testing the accuracy of such a system comparing with the data provided by the Armeo in the execution of the same arm movements.

The matrices received from the tracker are used for a real-time representation on the screen. The data referring to trajectories are stored in a database for elaboration and therapy assessments.

OpenSceneGraph is used for the 3D rendering since it is an effective open source graphics toolkit. Snapshots of simple VR scenes (reaching, moving and grasping a virtual object) are shown in Fig. 2.

Fig. 2. Up: Snapshots of simple VR scenes: reaching, moving and grasping a virtual object. Down: tracking with Kinect®

The BCI can be intended has a no-muscular communication channel able to send messages and commands from human to an external environment.

An innovative application of this type of devices is to use it in conjunction with a virtual environment for the object manipulation.

Several studies have shown that the use of BCI in the rehabilitation of patient with motor disabilities can provide different potential benefits [30]. Virtual reality can be augmented by an interpretation of the signals coming from different brain area during the execution of motor exercises.

This translation of signals is anyway a very difficult task and it depends on a series of factors (the concentration level is by far one of the most influencing of them).

The conjunction between BCI and realistic virtual reality create a good diagnostic and personalized environment in which it is possible to study the brain signals as answers to external stimuli or to assess the progress of the patient in the rehabilitation therapy.

As first step we have connected a commercial low cost BCI (Epoc® by Emotiv) with an avatar built up on our library for VR environment.

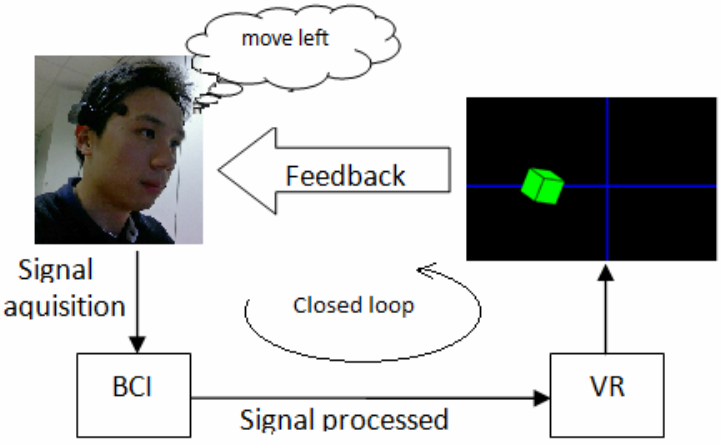


Fig. 3. Closed loop: human-machine interaction trough BCI and VR

The BCI incorporates 14 extensions of electrodes, mostly concentrated around the front of the scalp. The headset is completely wireless and consequently it allows free movements. The headset's electrodes record the resulting brain waves during the concentration, and from them on, the system recognizes that pattern as the specific function. We are using the "Emotiv" cognitive suite to analyze the basic brainwave activities in order to discern the user's conscious intent to perform distinct physical actions on a real or virtual object. It is possible to work with six directional movements (push, pull, left, right, up and down) and six rotations (clockwise, counter-clockwise, left, right, forward and backward). In addition emotional state and facial expressions are augmented the virtual scene. In the near future we will migrate from this basic BCI hardware to a more accurate and professional device ("g.BCIsys" by "G.tec") in order to provide a more precise cognitive analysis.

5 Conclusion

In this paper we have presented the overall architecture and the first development status of an advanced system that combines NR, MNP and VR for rehabilitation and functional compensation. We have focused our attention on the part of the hybrid technology system which concerns about virtual reality enhanced rehabilitation. Preliminary results are promising and we are currently investigating the robustness and accuracy of the tracking system during the execution of the single tasks. Next step will be to provide an advanced and complete scenario of the rehabilitation which will provide a detailed assessment of the patient progresses in rehabilitation.

Acknowledgment

This research is a part of the HYPER project funded by CONSOLIDER-INGENIO 2010, Spanish Ministry for Science and Innovation. The institutions involved are: Spanish National Research Council (CSIC), Center for Electrochemical Technologies (CIDETEC), Visual Communication and Interaction Technologies Centre (VICOMTech), Fatronik-Tecnalia, University of Zaragoza, University of Rey Juan Carlos, University of Carlos III (UC3M), Bioengineering Institute of Cataluña (IBEC) and Hospital for Spinal Cord Injury of Toledo.

References

1. Dobkin, B.: Rehabilitation after stroke. New England Journal of Medicine 352, 1677–1684 (2005)
2. Johansson, B.: Brain plasticity and stroke rehabilitation. The willis lecture, Stroke 31, 223–230 (2000)
3. Weiss, P., Kizony, R., Feintuch, U., Katz, N.: Virtual reality in neurorehabilitation. Textbook of neural repair and neurorehabilitation 2, 182–197 (2006)
4. Fidopiastis, C., et al.: Human experience modeler: Context-driven cognitive retraining to facilitate transfer of learning. CyberPsychology & Behavior 9, 183–187 (2006)
5. Cano de la Cuerda, R., et al.: Telerrehabilitacin y neurologa. Rev. Neurol. 51, 49–56 (2010)
6. Rizzo, A., Kin, G.: A swot analysis of the field of vr rehabilitation and therapy. Presence: Teleoperators and Virtual Environments 14, 119–146 (2005)
7. Viau, A., et al.: Reaching in reality and virtual reality: a comparison of movement kinematics in healthy subjects and in adults with hemiparesis. Journal of neuroengineering and rehabilitation, 1 (December 2004)
8. Subramanian, S., et al.: Virtual reality environments for post-stroke arm rehabilitation. Journal of neuroengineering and rehabilitation 4 (2007)
9. Kuttuva, M., et al.: The rutgers arm: an upper-extremity rehabilitation system in virtual reality. In: 4th International workshop on virtual reality rehabilitation, Catalina Islands, Citeseer (2005)
10. Fugl-Meyer, A., et al.: The post-stroke hemiplegic patient: a method for evaluation of physical performance. Scandinavian journal of rehabilitation medicine,vol 7, 13–31 (1975)
11. Boian, R., et al.: Virtual reality-based system for ankle rehabilitation post stroke. In: Proceedings of the First International Workshop on Virtual Reality Rehabilitation, Citeseer, pp. 77–86 (2002)
12. Koenig, A., et al.: Virtual gait training for children with cerebral palsy using the lokomat gait orthosis. Medicine meets virtual reality 16 (2008)
13. Colombo, G., et al.: Driven gait orthosis to do locomotor training of paraplegic patients. In:Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 4, pp. 3159–3163 (2000), doi: 10.1109/IEMBS.2000.901556
14. Hesse, S.: Locomotor therapy in neurorehabilitation. Neuro Rehabilitation 16, 133–139 (2001)
15. Veneman, J.F., et al.: Design and evaluation of the lopes exoskeleton robot for interactive gait rehabilitation, vol.15, pp. 379–386 (2007)

16. Lam, T., et al.: Lower limb rehabilitation following spinal cord injury. In: Eng, J.J., Teasell, R.W., Miller, W.C., Wolfe, D.L., Townson, A.F., Hsieh, J.T.C., Connolly, S.J., Mehta, S., Sakakibara, B.M. (eds.) Spinal Cord Injury Rehabilitation Evidence, pp. 1–47 (2010)
17. Mehta, S., et al.: Rehabilitation practice and associated outcomes following spinal cord injury. In: Eng, J.J., Teasell, R.W., Miller, W.C., Wolfe, D.L., Townson, A.F., Hsieh, J.T.C., Connolly, S.J., Mehta, S., Sakakibara, B.M. (eds.) Spinal Cord Injury Rehabilitation Evidence (2010)
18. Dietz, V.: Spinal cord pattern generators for locomotion. Clin. Neurophysiol. 114, 1379–1389 (2003)
19. MacKay-Lyons, M.: Central pattern generation of locomotion: a review of the evidence. Phys. Ther. 82, 69–83 (2002)
20. Maclean, N., Pound, P., Wolfe, C., Rudd, A.: Qualitative analysis of stroke patients' motivation. BMJ 321, 1051–1054 (2000)
21. Colombo, R., Pisano, F., Mazzone, A., Delconte, C., Micera, S., Carrozza, M.C., Dario, P., Minuco, G.: "Design strategies to improve patient motivation during robot-aided rehabilitation. Journal of NeuroEngineering and Rehabilitation 4(3) (2007), doi:10.1186/1743-0003-4-3
22. Emken, J.L., Bobrow, J.E., Reinkensmeyer, D.J.: Robotic movement training as an optimization problem: designing a controller that assists only as needed. In: 9th International Conference on Rehabilitation Robotics ICORR 2005, June 1–July, pp. 307–312 (2005)
23. Weiss, A., Suzuki, T., Bean, J., Fielding, R.A.: High intensity strength training improves strength and functional performance after stroke. Am J. Phys. Med. Rehabil. 79, 369–376 (2000)
24. Lünenburger, L., Colombo, G., Riener, R.: Biofeedback for robotic gait rehabilitation. Journal of NeuroEngineering and Rehabilitation 4 (2007), doi:10.1186/1743-0003-4-1
25. Pons, J.L.: Rehabilitation exoskeletal robotics. IEEE Engineering in Medicine and Biology Magazine 29, 57–63 (2010)
26. Pons, J.L.: Wearable Robots: Biomechatronic exoskeletons. John Wiley & Sons, Chichester (2008)
27. Kazerooni, H.: Human–robot interaction via the transfer of power and information signals. IEEE Transactions on Systems, Man, and Cybernetics 20, 450–463 (1990)
28. Gallego, J.A., Rocon, E., Ibanez, J., Dideriksen, J.L., Koutsou, A.D., Paradiso, R., Popovic, M.B., Belda–Lois, J.M., Gianfelici, F., Farina, D., Manto, M., Pons, J.L.: A Soft Wearable Robot for Tremor Assessment and Suppression. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation (2011) (accepted)
29. Leuthardt, E.C., Schalk, G., Roland, J., Rouse, A., Moran, D.W.: Evolution of brain-computer interfaces: going beyond classic motor physiology. Neurosurg. Focus 27 (July 2009)
30. Gert, P., Reinhold, S.: Brain-Computer Interfaces used for Virtual Reality Control. In: ICABB 2010 (2010)

Virtual Angioscopy Based on Implicit Vasculatures

Qingqi Hong[1], Qingde Li[1], and Jie Tian[2]

[1] Department of Computer Science, University of Hull, HU6 7RX, Hull, UK
[2] Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

Abstract. Virtual endoscopy is among the most active areas in medical data visualization, which focuses on the simulated visualizations of specific hollow organs for the purposes of training and diagnosis. In this paper, we present a virtual angioscopy technique based on vasculature geometry reconstructed using skeleton-based implicit splines (SIS). The highly accurate implicit representation of the vasculature not only makes it possible to achieve high visual quality of perspective view inside the vessel structures, but also makes the implementation of an interactive virtual angioscopy a much easier task, as the issue of collision detection of virtual camera with vascular objects can be easily solved when the vasculature is represented in implicit form. Some experiments have been carried out to demonstrate the strengths of our technique.

Keywords: Virtual endoscopy, Virtual angioscopy, Implicit modeling, Interactive navigation.

1 Introduction

Virtual endoscopy is one of the most active areas in medical data visualization. It is a kind of non-invasive diagnosis technique and has no direct deleterious effects on patients [1]. It uses computers to process 3D image datasets to provide simulated visualizations of specific hollow organs, similar or equivalent to those produced by standard endoscopic procedures [2]. This technique has been applied to virtual colonscopy [3,4], bronchoscopy [5], ventriculoscopy [6,7], and so on. Virtual angioscopy [8,9] is a specialized virtual endoscopy technique for exploring the human vascular systems, which generates an interactive environment for the vascular examination from a point of view inside the vessels [10]. Virtual fly-through of vascular structures is a useful technique for educational purposes and some diagnostic tasks, as well as intervention planning and intraoperative navigation [11]. In such a virtual visualization system, it is usually essential to combine the detailed views of the inner structures with an overview of the anatomic structures.

Virtual angioscopy requires a relatively high visual quality of perspective view inside the datasets for the purposes of training and diagnosis. One of the common approaches for the visualization of a virtual angioscopy is surface rendering, yielding images close to a real endoscopy. However, the direct application of

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 592–603, 2011.

surface rendering algorithms (i.e. Marching Cubes [12]) to the segmented vasculatures may suffer from the typical diamond artifacts caused by the trilinear interpolation [1]. Therefore, the smooth and accurate reconstruction of vascular tree is very crucial for virtual angioscopy. In this paper, we use implicit surface to represent the vascular structures, which is reconstructed using a skeleton-based implicit reconstruction technique [13], for the virtual angioscopy system. Our method can achieve high quality perspective views as well as accurate cross sections, which is suitable for training purposes as well as diagnosis tasks. Furthermore, based on the implicit modeling technique, the collision avoidance for the camera navigation of virtual angioscopy, a key problem for implementing interactive navigation of a virtual vascular system, can be easily solved.

2 Related Work

The major issues associated with virtual endoscopy are the perspective rendering techniques and the camera navigation paradigms.

2.1 Perspective Rendering Techniques

Generally, the rendering techniques for virtual endoscopy fall into two categories: surface rendering and volume rendering. Surface based rendering typically extracts surfaces by fitting geometric primitives, such as polygons or patches, to constant-value contour surfaces in volumetric datasets [14]. It has been widely applied for virtual endoscopy because of its high rendering speed. In addition, the extracted surface, containing the space-relation information, is very useful for the navigation of virtual camera. Many current virtual endoscopy systems, such as VESA [15], VirEn [16], 3D Slicer [17] and FreeFlight [18], are based on the surface rendering technique. However, the general visual quality based on surface rendering is comparatively poor. Particularly, the direct isosurface rendering for the segmented dataset may suffer from the diamond artifacts caused by the trilinear interpolation [1]. Therefore, the smooth and accurate reconstruction of segmented results is very crucial for surface rendering-based virtual endoscopy.

Compared to surface rendering, volume rendering (sometimes called direct volume rendering) is a technique for directly displaying 3D sampled datasets in the form of 2D projection, without the intermediate geometric primitive representations used by surface rendering [14]. Generally, the volume rendering techniques, such as ray casting [19] and splatting [20], can generate higher visual quality results for virtual endoscopy. However, the frame rate they can achieve is relatively low, which requires accelerating algorithms for real time rendering. On the other hand, the texture mapping-based volume rendering [21] needs additional support of high-performance graphics hardware. There are several virtual endoscopy systems based on volume rendering, such as VI VoxelView [22] and CRS4 [23]. However, in the case of virtual angioscopy, when the vascular structures are rather small, the direct application of volume rendering techniques usually leads to severe aliasing artifacts [24]. In addition, as stated in [25] the

surface rendering technique was statistically significantly better in visualizing subependymal arteries, cranial nerves, and other lesions.

2.2 Navigation Paradigms

Besides rendering, the camera navigation paradigm is another key problem required to be solved for the development of a virtual endoscopy system. It includes the interaction of the user to control camera movement, as well as the operations of mapping the input device movements to camera parameter modifications. The user should get neither a "lost-in-space" feeling nor a frustration feeling due to a heavily constrained navigation environment [26].

Various virtual endoscopy techniques can be roughly classified into three classes: automatic navigation, manual or free navigation and guided navigation [3]. With automatic navigation, the user defines a certain number of key frames where camera parameters are specified. Smooth camera movements between the key frames are calculated automatically. The drawback of automatic navigation is the lack of interactivity, which means that user interaction is limited and the irrelevant regions cannot be easily skipped [1]. By manual navigation, user can completely control over all parameters of the virtual camera without any constraints. However, the method needs to performance collision-avoidance scheme, which requires costly query operations. In addition, in contrast to automatic navigation, free navigation requires a highly interactive rendering technique, since a significant lag between interaction and rendering will severely disturb the user's interaction [1].

This guided navigation is between the two previous methods, which combines user's guidance with an efficient collision-avoidance scheme [1]. The user controls over the camera parameters but some constraints are added, such as keeping the position of the camera to the optimal path. For example, Vilanova et al. [26] have described a guided navigation system where the location of the camera is fixed to a pre-computed path, while the camera orientation can be selected freely. Another typical system is Hong et al.'s virtual voyage [3], which employs several potential fields and kinematic rules to guide the virtual camera along the human colon. Although this kind of technique can achieve higher frame rate because of the efficient collision-avoidance scheme, the user's interaction is still limited due to the additional restrictions. In addition, the generation of potential fields has the potential risk of changing the morphology of the anatomic structures (we will explain this more detail in section 4). While based on the implicit vascular geometry, the collision avoidance for the camera navigation of virtual angioscopy can be easily solved without additional generation of potential fields, since the constructed model is an connatural implicit volume, which is a favorite kind of geometric object when performing collision detection [27].

3 Virtual Angioscopy

For our virtual angioscopy system, we basically follow the standard processing pipeline [10] (see Fig. 1). Firstly, segmentation technique has been employed to

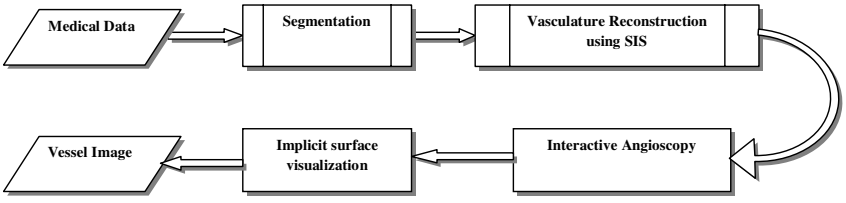


Fig. 1. Pipeline for our virtual angioscopy system

extract the vessel structures from standard 3D medical datasets, such as CT or MRA images. Secondly, in order to achieve high visualization quality, we adopt our skeleton-based implicit splines (SIS) modeling technique [13] to accurately and smoothly reconstruct the vasculatures from the segmented discrete vascular surface points. Based on the implicit vascular geometry, the virtual camera can be automatically or freely navigated along the vascular tree and acquire high-quality perspective view inside the vessel structures.

3.1 Accurate Vasculature Reconstruction Using Skeleton-Based Implicit Splines (SIS)

In this section, we will briefly introduce the technique of accurate vasculature reconstruction using SIS. For more details, please refer to our another paper [13]. This technique is based on an implicit surface modeling method that has been developed to model generalized cylinders. In this implicit generalized cylinders modeling method, the freeform cross-sections are first reconstructed implicitly using the 2D piecewise algebraic splines [27], and then, different cross-section profiles are weighted and summed up along the skeleton using the Partial Shape Preserving (PSP) spline basis functions, the 1D version of 2D piecewise algebraic splines. In addition, the smooth piecewise polynomial blending operations [28] is employed to blend the branches of implicitly constructed generalized cylinders together. This approach can construct a smooth, continuous and analytic surface. The proposed method has been applied to actual 3D medical data for the reconstruction of vasculatures. Direct visual experimental results demonstrate that this method can correctly represent the morphology and topology of vascular structures. In addition, qualitative and quantitative analysis have been performed for validating the accuracy and smoothness of the reconstructed results.

As can be seen from the Fig. 2 and Fig. 3, a visual comparison between our reconstruction results and the segmentation results gives us the first evidence that the method based on SIS can correctly represent the morphology and topology of vascular structures. The isosurface rendering of segmentation results (left column) suffers from strong aliasing artifacts like staircases, which has a strong divergence with realistic vessels and might hamper the visual interpretation of the vessel surface [24]. In contrast to the direct visualization of segmentation

Fig. 2. The reconstruction of MRA cerebral vessels: (left) the isosurface rendering of segmentation result, (right) the reconstruction result using SIS method

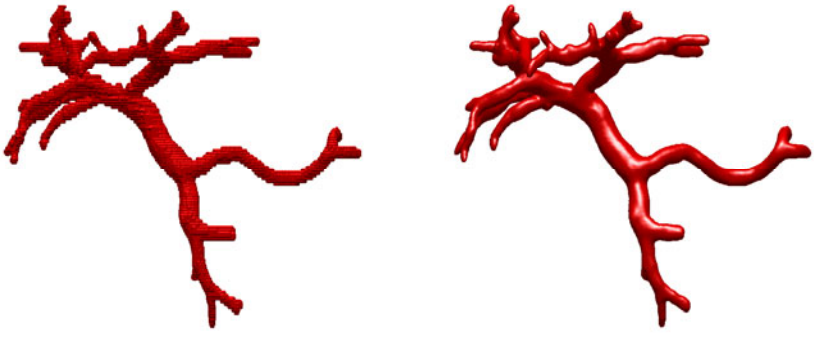


Fig. 3. The reconstruction of liver portal vein: (left) the isosurface rendering of segmentation result, (right) the reconstruction result using SIS method

result, the approach based on SIS can achieve superior visual quality and produce smooth transitions at branchings (right column). Indeed, the SIS method can guarantee the same high geometric continuity, as the reconstruction method that based on the convolution surfaces [29]. Furthermore, the SIS technique can achieve more accurate vessel surfaces since it is constructed without model assumptions. Generally, the SIS method can achieve quite smooth and accurate vessel surfaces. Compared with the direct visualization of segmented result, the vascular structures reconstructed by the SIS method are more faithful to the realistic vessels.

3.2 Interactive Angioscopy

As discuss above, virtual angioscopy requires high visual quality of perspective view inside the dataset. The general surface rendering techniques for segmented datasets are not sufficient to generate high-quality visualization. On the other

hand, by using SIS modeling technique, the segmented vasculatures can be represented as an analytic implicit function. Based on this implicit function, we can render high-quality vessel surfaces with any required continuity and smoothness. Besides the high-quality perspective views, we can also achieve accurate cross sections, since our method is without model assumption.

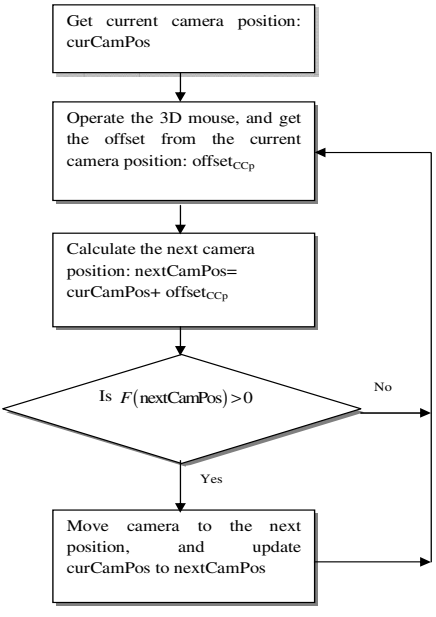


Fig. 4. The flowchart of the manual navigation of virtual camera

For the navigation of camera, we can achieve efficient collision avoidance without the additional generation of potential fields, since the SIS method can guarantee to define an implicit volume by replacing the equality of Eq. 7 in [13] with an inequality:

$$F(x, y, z) = f(X(x, y, z), Y(x, y, z), Z(x, y, z) \geq 0 \tag{1}$$

That is, the vasculatures are represented as a global implicit function $F(x, y, z)$. When $F(x, y, z) = 0$, it represents the vessel surfaces; when $F(x, y, z) > 0$, it represents implicit volume inside the vessel structures; and when $F(x, y, z) < 0$, it represents the implicit volume outside the vessel structures. The implicit volume is a favourite kind of geometric object when performing collision detection [27]. When the vasculatures are modelled as implicit volume, one can tell directly whether a point lies inside or outside the vasculatures and the problem of collision detection can be easily solved [30]. In other words, $F(x, y, z)$ is a kind of

signed distance function (SDF) [31], which guarantees that the closer the point to the vascular axe, the bigger the value of function $F(x, y, z)$.

Fig. 4 demonstrates the flowchart of the manual navigation of virtual camera inside the vasculatures. By operating the 3D Mouse (`http://www.3dconnexion.com/index.php`), the next camera position can be easily calculated. And then, we test the implicit function value: $F(nextCamPos)$. If $F(nextCamPos) > 0$, we suppose the new pre-computed camera position is inside the vasculatures, and we move the camera to the new pre-computed position and update the current camera position. Otherwise, new pre-computed position is outside the vasculatures, which requires the re-operation of 3D mouse.

4 Results

The presented virtual angioscopy system has been applied to the 3D CT angiography (CTA) images of carotid artery, the 3D magnetic resonance angiography (MRA) images of cerebral artery and abdominal aorta supplied by Intelligent Bioinformatics Systems Division, Institute of Automation, the Chinese Academy of Sciences, and the segmented liver portal vein obtained from the public resource (`http://www.ircad.fr/softwares/3Dircadb/3Dircadb1/index.php`)(see Table 1). Based on the reconstruction of vasculatures using SIS, we can easily examine the interior of the vascular systems using the technique of virtual endoscope.

Table 1. Summary of characteristics of the datasets. Voxelsizes are given in millimeter.

Dataset	Resolution	Voxelsize
CTA carotid artery	$512 \times 512 \times 206$	$0.52 \times 0.52 \times 0.63$
MRA cerebral vessels	$352 \times 448 \times 114$	$0.49 \times 0.49 \times 0.80$
MRA abdominal aorta	$512 \times 512 \times 310$	$0.70 \times 0.70 \times 0.63$
Segmented Liver portal vein	$512 \times 512 \times 151$	$0.78 \times 0.78 \times 1.60$

The skeleton of the vessel has been extracted during the process of reconstruction, thus, it is convenient to implement the automatic navigation mode of virtual angisocopy using the pre-extracted skeleton as camera path (see Fig. 5). The camera moves along the skeleton, and the target point is set to a fixed distance ahead on the skeleton so that the camera can smoothly follow each vessel segment.

As discussed in Section 3, compared to the direct application of surface rendering algorithms (i.e. Marching Cubes) to the segmented vasculatures (see Fig. 6 (right)), based on the SIS modeling of vasculatures, the surface rendering methods can achieve much higher quality perspective views as well as accurate cross sections (without model assumption) (see Fig. 6 (middle)), which is suitable for training purposes as well as diagnosis tasks.

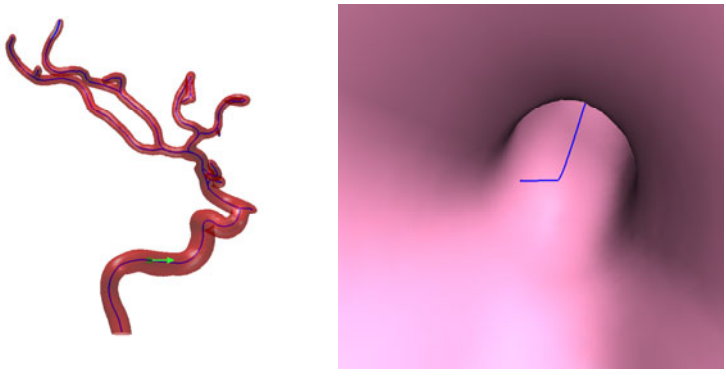


Fig. 5. Virtual Angisocopy using the pre-extracted skeleton as camera path: (left) 3D overview of reconstructed vessel tree with skeleton (green arrow represents the current position and direction of virtual camera), (right) perspective view inside the vasculatures (blue line represents the ongoing camera path)

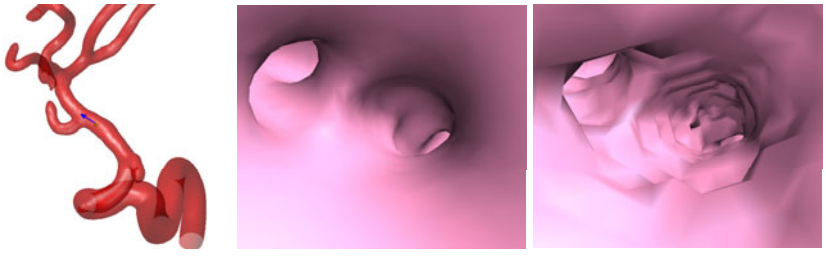


Fig. 6. Virtual Angioscopy: the overview of the vessel structures, and the arrow indicating the current position and orientation of the camera (left), the perspective view inside the vessel based on our implicit modeling vasculatures (middle), and on direct application of Marching Cubes to the segmented vasculatures (right)

For the free or guided navigation mode of virtual angisocopy, the generation of potential fields is required for achieving efficient collision avoidance. However, the generation of potential fields has the potential risk of changing the morphology of the anatomic structures. For instance, we generate the distance field coding the distance to the vascular surface using the fast marching method [32], which has changed the surface morphology of vessel (see Fig. 7 (top right), the vessel surface is not as smooth as that before the generation of distance field). Accordingly, the changes of vessel surface morphology would greatly influence the visual effect of perspective rendering inside the vasculature (see Fig. 7 (bottom right)). On the other hand, as stated in Section 3.2, we can easily achieve efficient collision avoidance based on the constructed implicit function, without the additional generation of potential fields. In addition, by taking the advantages of 3D mouse, the position and orientation of camera can be freely controlled, without the limitations of traditional input devices with limited degrees of freedom. The

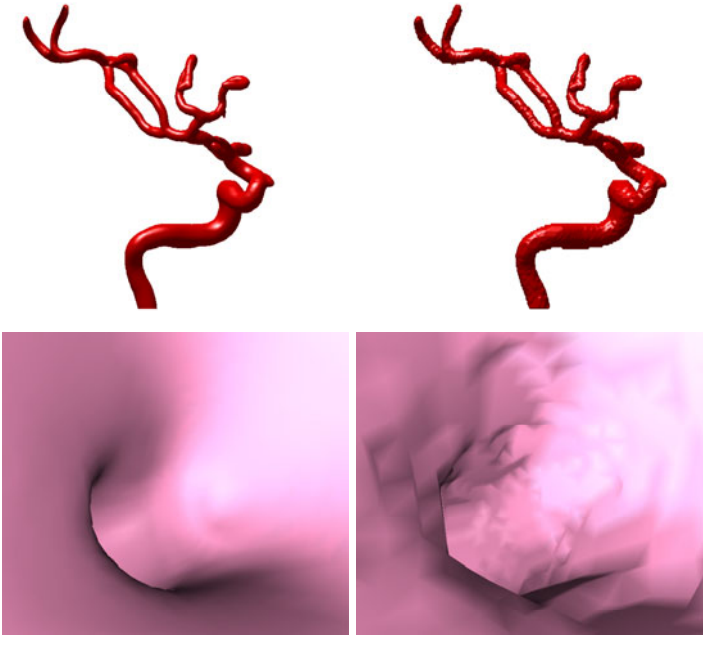


Fig. 7. The side effect of generating distance field: (top left) and (top right) are the surface morphology of vessel before and after generating distance field, and (bottom left) and (bottom right) are perspective views inside the vasculatures based on (top left) and (top right)

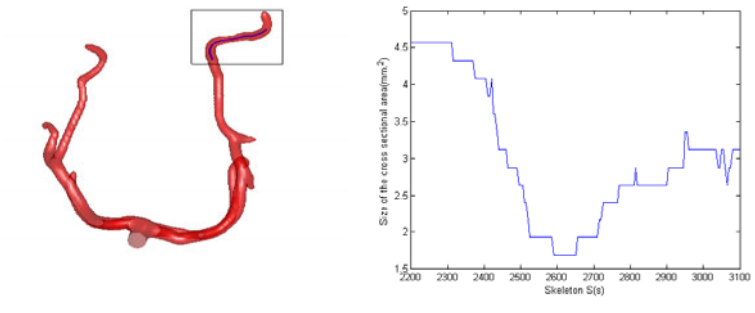


Fig. 8. The diagnosis of vessel stenosis: (left) the overview of the vessel suspected to suffer stenosis, (right) the accurate diagnosis by calculating the distribution curve of the size of the cross sectional area

users can achieve an immersive experience of flexible and effective exploration of the interior of the vascular tree, since they can easily and freely move the camera closer to an object and inspect the region of interest by changing the viewpoint.

For the purpose of diagnosing vessel stenosis, the area quantitative analysis of cross section is necessary, since the visual observation is not adequate for accurate diagnosis. For instance, as shown in Fig. 8 (left), the vessel segment in rectangle is suspected to suffer stenosis from a general overview of the vasculature. Based on our system, the accurate diagnosis can be easily achieved by exploring along the skeleton of the vessel segment to acquire the distribution curve of the size of the cross sectional area (see Fig. 8 (right)). The horizontal axis is the parametric curve $S(s)$ representing the skeleton of the vessel; and the vertical axis represents the size of cross-sectional area for the corresponding skeleton. For normal vessel, according to the surgeon's opinion, the size of the cross-sectional area should in general change gradually along a branch skeleton; while in this case, the size of the cross-sectional area change sharply, which is supposed to be suffering stenosis.

5 Conclusions

Due to the complex nature of vascular structures, it is essential to combine the detailed views of the inner structures with an overview of the anatomic structures. Virtual angioscopy (VA) technique provides us an interactive environment for exploring the human vascular system. The smooth and accurate reconstruction of vascular tree is very crucial for virtual angioscopy, since it requires a relatively high visual quality of perspective view inside the dataset for the purposes of training and diagnosis. In this paper, we present a virtual angioscopy system based on implicitly reconstructed vasculatures using SIS. Compared to the direct application of surface rendering technique to the segmented datasets, our method can achieve much higher visualization quality as well as accurate cross sections. In addition, interactive camera navigation can be easily implemented as it is very simple to perform camera-vascular wall collision detection, since the vascular geometry is represented as implicit surfaces.

The use of virtual angioscopy has several benefits. First and foremost, compared to real endoscopy, it is non-invasive or at least minimally invasive, which can provide insights into some vessel parts that might be difficultly accessible to current medical procedures. It can be easily applied for educational purpose, which provides unusual insights into the vessel of living patients. In addition, it has the potential benefits for non-invasive evaluation of vascular diseases [33]. In our VA system, the free navigation mode allows high-precision manual analysis of the vasculature under various viewing angles and better dynamic localisation of abnormalities. Furthermore, the area quantification of cross section can be easily achieved for the purpose of diagnosing vessel stenosis. However, VA has not yet been extensively evaluated. In the future, clinical evaluation of our system on vascular diseases needs to be performed. And its integration into the actual surgical intervention as a navigation aid is also a challenge for our system.

References

1. Bartz, D.: Virtual endoscopy in research and clinical practice. Computer Graphics Forum 24(1), 111–126 (2005)
2. Wickham, J.: Minimally invasive surgery: Future developments. BMJ 308, 193–196 (1994)
3. Hong, L., Muraki, S., Kaufman, A., Bartz, D., He, T.: Virtual voyage: Interactive navigation in the human colon. In: Proceedings of ACM SIGGRAPH, pp. 27–34 (1997)
4. Bartrolí, A.V.: Visualization Techniques for Virtual Endoscopy. PhD thesis, Technischse Universitä Wien (2001)
5. Ferretti, G.R., Vining, D.J., Knoplioch, J., Coulomb, M.: Tracheobronchial tree: Three-dimensional spiral ct with bronchoscopic perspective. Journal of Computer Assisted Tomography 20(5), 777–781 (1996)
6. Auer, D.P., Auer, L.M.: Virtual endoscopy - a new tool for teaching and training in neuroimaging. International Journal of Neuroradiology 4, 3–14 (1998)
7. Bartz, D., Skalej, M., Welte, D., Straßr, W., Duffner, F.: A virtual endoscopy system for the planning of endoscopic interventions in the ventricle system of the human brain. In: Proc. of BiOS 1999: Biomedical Diagnostics, Guidance and Surgical Assist Systems (1999)
8. Davis, C.P., Ladds, M.E., Romanowski, B.J., Wildermuth, S., Kopflioch, J.F., Debatin, J.F.: Human aorta: Preliminary results with virtual endoscopy based on three-dimensional mr imaging data sets. Radiology 199, 37–40 (1996)
9. Gobbetti, E., Pili, P., Zorcolo, A., Tuveri, M.: Interactive virtual angioscopy. In: Proc. of IEEE Visualization, pp. 435–438 (1998)
10. Bartz, D., Straßr, W., Skalej, M., Welte, D.: Interactive exploration of extra and intracranial blood vessels. In: Proc. of IEEE Visualization, pp. 389–392 (1999)
11. Preim, B., Oeltze, S.: 3d visualization of vasculature: An overview. Visualization in Medicine and Life Science, 39–59 (2007)
12. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Proc. of ACM SIGGRAPH, pp. 163–169 (1987)
13. Hong, Q., Li, Q., Tian, J.: Implicit reconstruction of vasculatures using implicit splines. submitted to IEEE Transactions on Medical Imaging (2011)
14. Elvins, T.: A survey of algorithms for volume visualization. Computer Graphics ACM Siggraph Quarterly 26(3), 194–201 (1992)
15. Lorensen, W., Jolesz, F., Kikinis, R.: The exploration of cross-sectional data with a virtual endoscope. In: Satava, R., Morgan, K. (eds.) Interactive Technology and New Medical Paradigms for Health Care, pp. 221–230 (1995)
16. Nain, D., Haker, S., Kikinis, R., Grimson, W.: An interactive virtual endoscopy tool. In: Proceedings of Workshop on Interactive Medical Image Visualization and Analysis (2001)
17. Bruckner, S.: Efficient volume visualization of large medical datasets. Master's thesis, Computer Science Department, Technical University of Vienna (2003)
18. Vining, D., Stelts, D., Ahn, D., Hemler, P., Ge, Y., Hunt, G., Siege, C., McCorquodale, D., Sarojak, M., Ferretti, G.: Freeflight: A virtual endoscopy system. In: Troccaz, J., Mösges, R., Grimson, W.E.L. (eds.) CVRMed-MRCAS 1997, CVRMed 1997, and MRCAS 1997. LNCS, vol. 1205, pp. 413–416. Springer, Heidelberg (1997)
19. Tuy, H., Tuy, L.: Direct 2-d display of 3-d objects. IEEE Computer Graphics and Applications 4(10), 29–33 (1984)

20. Westover, L.: Footprint evaluation for volume rendering. Computer Graphics 24(4), 367–376 (1990)
21. Cabral, B., Cam, N., Foran, J.: Accelerated volume rendering and tomographic reconstruction using texture mapping hardware. In: 1994 Symposium on Volume Visualization, Conference Proceedings, ACM SIGGRAPH, pp. 91–98 (1994)
22. Serlie, I., Vos, F., Gelder, R.v., Post, F., Nio, Y., Gerritsen, F., Truyen, R., Stoker, J.: Improved visualization in virtual colonoscopy using image-based rendering. In: Data Visualization (Proceedings of Symposium on Visualization), pp. 137–146 (2001)
23. Beier, J., Diebold, T., Vehse, H., Biamino, G., Fleck, E., Felix, R.: Virtual endoscopy in the assessment of implanted aortic stents. Computer Assisted Radiology, 183–188 (1997)
24. Schumann, C., Oeltze, S., Bade, R., Preim, B., Peitgen, H.O.: Model-free surface visualization of vascular trees. In: IEEE/Eurographics Symposium on Visualization 2007, pp. 283–290 (2007)
25. Nakajima, N., Wada, J., Miki, T., Haraoka, J., Hata, N.: Surface rendering-based virtual intraventricular endoscopy: Retrospective feasibility study and comparison to volume rendering-based approach. NeuroImage 37 (suppl. 1), 89–99 (2007)
26. Vilanova, A., Köig, A., Gröler, E.: Viren: A virtual endoscopy system. Journal Machine Graphics and Vision 8(3), 469–487 (1999)
27. Li, Q., Tian, J.: 2d piecewise algebraic splines for implicit modeling. ACM Transactions on Graphics 28(2) (2009)
28. Li, Q.: Smooth piecewise polynomial blending operations for implicit shapes. Computer Graphics forum 26(2), 157–171 (2007)
29. Oeltze, S., Preim, B.: Visualization of vascular structures with convolution surfaces: Method, validation and evaluation. IEEE Transactions on Medical Imaging 25(3) (2005)
30. Lin, M., Gottschalk, S.: Collision detection between geometric models: A survey. In: Proc. of IMA Conference on Mathematics of Surfaces (1998)
31. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Springer, New York (2002)
32. Sethian, J.A.: Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press, Cambridge (1999)
33. Louisa, N., Bruguiereb, E., Kobeiterb, H., Desgrangesa, P., Allairea, E., Kirschc, M., Becquemina, J.: Virtual angioscopy and 3-dimensional navigation findings of the aortic arch after vascular surgery. European Journal of Vascular and Endovascular Surgery 40(3), 340–347 (2010)

Improvement of Security and Feasibility for Chaos-Based Multimedia Cryptosystem*

Jianyong Chen and Junwei Zhou

Department of Computer Science and Technology, Shenzhen University, Shenzhen, China 518060
Cjyok2000@hotmail.com

Abstract. Nonlinear dynamic filter (NDF) has been used in chaos-based multimedia cryptosystem. However, our study shows that the key of randomized arithmetic coding (RAC) based on NDF can be successfully recovered under chosen plaintext attack. Moreover, current ciphertext block can't be decoded unless preceding plaintext is available. In order to enhance the security and feasibility used in multimedia applications, the algorithm is improved by building a new correlation between ciphertext and coefficients. Its security is enhanced that can effectively resist chosen plaintext attack. Its feasibility is also improved that can decode ciphertext block without availability of preceding plaintext with which a user can play multimedia data starting at any place. The analysis and simulations show that the improved algorithm can evidently enhance both security and feasibility

Keywords: Chaos, Cryptography, Security, Arithmetic coding.

1 Introduction

The security and efficiency requirements of data transmission make data compression and encryption become more and more important. In order to improve performance and satisfy requirements of multimedia application such as playing a certain portion of a video or audio, it is worthwhile to joint compression and encryption in a united process [1, 2]. It is reported that the united scheme is more secure and effective than the classical separate compression-encryption schemes [3].

Arithmetic coding is a method for lossless data compression and performs excellence in many respects [4], which is widely used in a variety of multimedia application [5]. Recently, various studies have been taken to integrate cryptography into the arithmetic coding [6-13]. Among them, randomized arithmetic coding (RAC) is proposed with multimedia selective encryption including some randomization in the arithmetic coding procedure to achieve encryption [6]. In order to improve the security of RAC [6], Schemes [7 ,8] proposed two variants of RAC utilizing chaotic

* This work was supported by Shenzhen University Research and Development Fund (SZU R/D) with Grant No. 200903.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 604–618, 2011.

map as the pseudo-random number generator and making the random number bitstream associated with the plaintext. In the scheme [8], a kind of nonlinear dynamic filter (NDF) whose probability density function is distributed uniformly is adopted to expand key space and improve security. The plaintext is mapped into the chaos parameter space and thus the random number bitstream can associate with the plaintext. Unfortunately, after studying the scheme proposed in [8], it is found that the key of NDF is easily recovered under chosen plaintext attack scenario. Furthermore, function requirements, e.g., fast-forward, rewind and playing video or audio from arbitrary position at a user's discretion, are necessary in multimedia player. Because the random number bitstream depends on the preceding plaintexts, the decoder must firstly obtain the all preceding plaintext and then generate the random number bitstream. Finally, the decoder could decrypt the current ciphertext successfully. However, in multimedia application scenario, the starting point is arbitrary at a user's discretion and the decoder couldn't always get all previous plaintext to generate random number bitstream. Therefore, the encryption algorithms [7, 8] are limited in multimedia application since it depended on plaintext sequence.

In order to enhance the security and improve feasibility, a nonlinear correlation model between ciphertext and random number bitstream is established. The ciphertext instead of plaintext is mapped into the chaos parameter spaces based on a nonlinear map, and coefficients of NDF are derived from the ciphertext. Similarly, the random number bitstream depends on both initial value of NDF and the ciphertext, which can employ to flexible design of the coefficients of NDF [14]. Because of the nonlinear correlation model between ciphertext and random number bitstream, the security problem presented in this scheme can be solved. Meanwhile, coefficients of NDF are derived from the ciphertext instead of plaintext, the decoder could decrypt current ciphertext block successfully without any knowledge of preceding plaintext. This makes our approach to meet the needs of multimedia application, such as display of video and audio starts from arbitrary portion as user desire. In proposed algorithm, the randomness of the random number bitstream, which plays the most important role in proposed cryptosystem, is confirmed by the statistical test suite which is recommended by the U.S. National Institute of Standards and Technology (NIST) [15].

The rest of this paper is organized as follows. In next section, RAC and cryptanalysis of RAC based NDF are reviewed. Cryptanalysis of existing scheme is presented in section 3. The improved operation is proposed to enhance the security and the flexibility in Sections 4. Analyses and simulation results can be found in Sections 5. In the last section, conclusions are figured out.

2 Review of the RAC and Its Variants

2.1 RAC

The scheme of RAC [6] is a multimedia cryptography that its order of the symbol intervals is disturbed by secret random number bitstream. Only the decoder who obtains the random number bitstream could decode ciphertext correctly. The scheme

of RAC consists of two parts: One part is pseudo-random number generator for generating random number bitstream and disturbing the order of intervals. In [7], logistic map is used as the pseudo-random generator, while in [8], NDF is used as the generator. The other part uses public traditional arithmetic encoding to achieve compression. In the scheme of RAC and its variants, intervals used in arithmetic coding are swapped according to random number bitstream. Assuming that the probabilities of symbols 0 and 1 are 1/4 and 3/4, the corresponding intervals are [0, 1/4] and (1/4, 1]. In general, if random number bit r_i is equal to 0, corresponding intervals of both symbol 0 and 1 are kept with [0, 1/4] and (1/4, 1]. Otherwise, the interval of symbol is swapped, so the interval mapped to symbol 0 is (3/4, 1] and the interval mapped to symbol 1 is [0, 3/4]. Fig.1 is draft of RAC and its variants. The pseudo-random number generator of variants [7, 8] is correlation with plaintext, which can enhance the security of RAC. It is obvious that the random number bitstream is the most important part of security. Once the opponent obtained the key of the pseudo-random number generator, RAC can be broken easily.

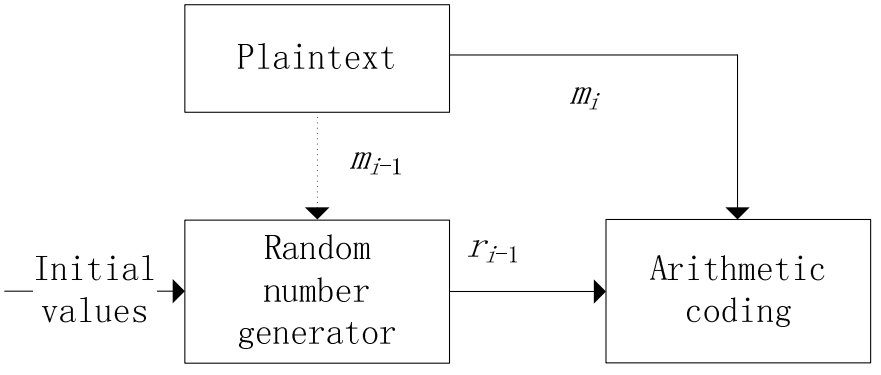


Fig. 1. Structure of the RAC and its variants where RNG is pseudo-random number generator and AC is traditional arithmetic encoding

Fig. 2 (a) shows procedure for encoding sequences 110 used in RAC. The order of intervals mapped to symbol 0 and 1 is determined by a secret random number bitstream $\{r_0, r_1, r_2\}$. Here, the probabilities of symbols 0 and 1 are 1/4 and 3/4, the corresponding intervals are [0, 1/4] and (1/4, 1]. The secret random number bitstream $\{r_0, r_1, r_2\}$ is 010 generated from pseudo-random number generator. Firstly, we encode plaintext block m_1=1. For the corresponding random number bit r_0=0, the order of intervals keeps the same. Now, we get interval [1/4, 1] after we encode plaintext block m_1 as arithmetic encoding. Sequentially, we encode plaintext block m_2=1. For the corresponding random number bit r_1=1, the order of intervals is swapped. After we encode plaintext block m_2=0, we get interval [1/4, 13/16]. Finally, we encrypt the third plaintext m_3, and we get final interval [1/4, 25/64]. For

comparison, the traditional arithmetic encoding is presented in Fig.2 (b). The order of interval keeps the same all the time. After three plaintext blocks are encoded, the final interval [7/16, 37/64] is obtained. Details of arithmetic coding process is not depicted in this figure, more details please refer to [6].

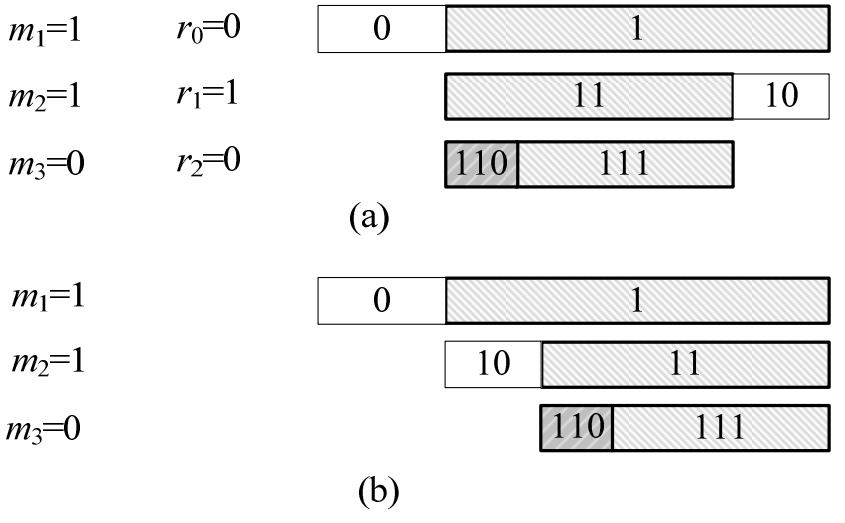


Fig. 2. Illustration of randomized arithmetic coding. (a) is for RAC and (b) is for traditional arithmetic coding

2.2 NDF

Since NDF has uniform distribution and large key space [14, 16], it has been successfully used in chaotic encryption algorithms [8, 17]. To enhance the security of RAC, the plaintext is mapped into the chaos parameter space. The binary random number bitstream produced by NDF is designed as follows.

Without loss of generality, the message is assumed as a sequence of symbols and the count of symbols is l, where $\{m_1, m_2 \ldots m_l\}$ is the ASCII value of original message. $\{r_1, r_2 \ldots r_t\}$ is the random number bitstream generated by NDF. $\{y_{-1}, y_0\}$, p and ϕ are the initial values of NDF.

1. The encoder gets z_{j-2} and z_{j-1} from Eq. (2) after getting ξ_{j-2} and ξ_{j-1} from Eq. (1). Here, m_{j-2} and m_{j-1} are the previous plaintext blocks with ASCII values.

$$\begin{aligned} \xi_{j-2} &= 3 + m_{j-2} / 256 \\ \xi_{j-1} &= 3 + m_{j-1} / 256 \end{aligned} \tag{1}$$

$$\begin{aligned} z_{j-2} &= \xi_{j-2} + \xi_{j-1} \\ z_{j-1} &= -\xi_{j-2} * \xi_{j-1} \end{aligned} \tag{2}$$

2. The encoder obtained variables $\{y_{i-2}, y_{i-1}\}$ from last step. Then, the encoder could get y_i from Eq. (3). The random number bit r_i can be obtained by the Eq. (4).

$$y_i = h \circ \mathrm{mod}(z_{j-2} \times y_{i-1} + z_{j-1} \times y_{i-2} + \phi) \tag{3}$$

$$r_i = \begin{cases} 1, & y_i \geq 0.5, \\ 0, & else \end{cases} \tag{4}$$

3. Repeat Steps 1) and 2), the encoder obtains the whole random number bitstream.

The function mod(.) of Eq. (3) is a modulo map described at Eq. (5), and the function of $h(.)$ represented in Eq. (6) is a piecewise linear map. Because the random number bitstream is key point for security of the algorithm, only the scheme of pseudo-random number generator is presented here. The encoding procedure of traditional arithmetic coding is the same as RAC [8].

$$\mathrm{mod}(v) = v - 2\lfloor (v+1)/2 \rfloor \tag{5}$$

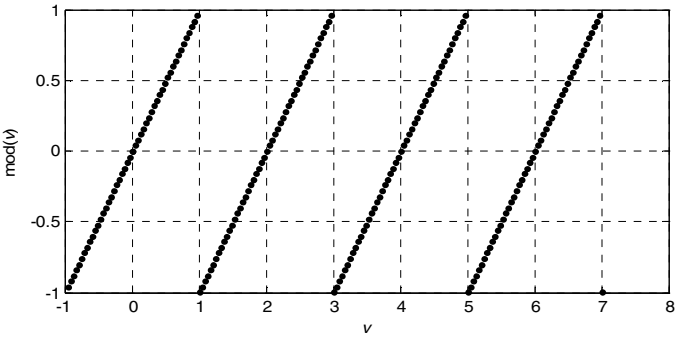
$$h(w) = \begin{cases} w/p, & 0 \leq w < p, \\ (w-p)/(0.5-p), & p \leq w < 0.5 \\ (1-w-p)/(0.5-p), & 0.5 \leq w < 1-p \\ (1-w)/p, & 1-p \leq w < 1 \\ h(-w), & w < 0 \end{cases} \tag{6}$$

3 Cryptanalysis of RAC Based NDF

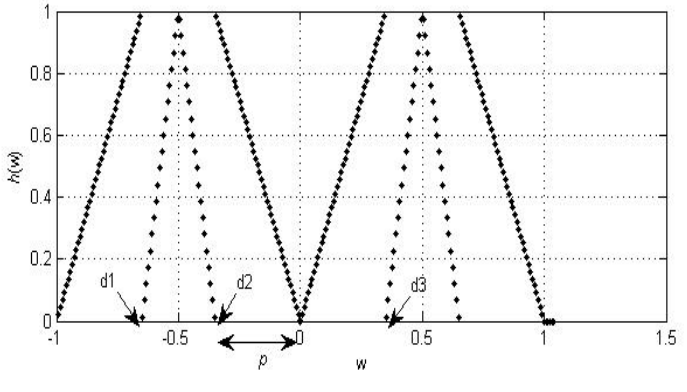
3.1 Chosen Plaintext Attack

A chosen-plaintext attack is an attack model for cryptanalysis which presumes that the attacker has the capability to choose arbitrary plaintext to be encrypted, and obtain the corresponding ciphertext. The goal of the attack is to gain usable information

which can reduce security of the encryption scheme. In the worst case, a chosen-plaintext attack could reveal the secret key. Here, a chosen-plaintext attack is used to show the vulnerability of the NDF. Assume that plaintext and corresponding ciphertext are exactly known to cryptanalyst in chosen plaintext attack scenario. It is easy to obtain the random number bitstream from the plaintext and corresponding ciphertext because the arithmetic coding is reversible operation. Therefore, random number bitstream generated by NDF can be considered as a known variable in chosen plaintext attack scenario.



(a) A plot of Eq. (5)



(b) A plot of Eq. (6)

Fig. 3. Plots of Eq. (5) and Eq. (6)

At the beginning of chosen-plaintext attack, various equations of NDF are examined first. Eq. (5) is a period-2 periodic function. Eq. (7) is an equivalent equation with Eq. (5). Fig.3 (a) shows that Eq. (5) is an identity transform when v is in

the range of [-1, 1]. These equations cannot evidently increase confusion of the algorithm. Here, Eq. (6) is the most important function in NDF which is depicted in Fig.3 (b). The plot of piecewise linear map $h(.)$ is constituted by the discrete points and consists of two distinct lines. Assume that there are a_1 points in the range of $[0, p]$ and a_2 points in the range of $[p, 0.5]$. Because the intervals of two points are equality, the length of $[0, p]$ is represented by a_1 and the length of $[p, 0.5]$ is represented by a_2.Now we know for sure that the Eq. (8) can reflect their relationship.

$$\begin{cases} f(x) = x, & x \in [-1,1], \\ f(x) = f(x+2), & else \end{cases} \tag{7}$$

$$\frac{a_1}{a_2} = \frac{p}{0.5-p} \tag{8}$$

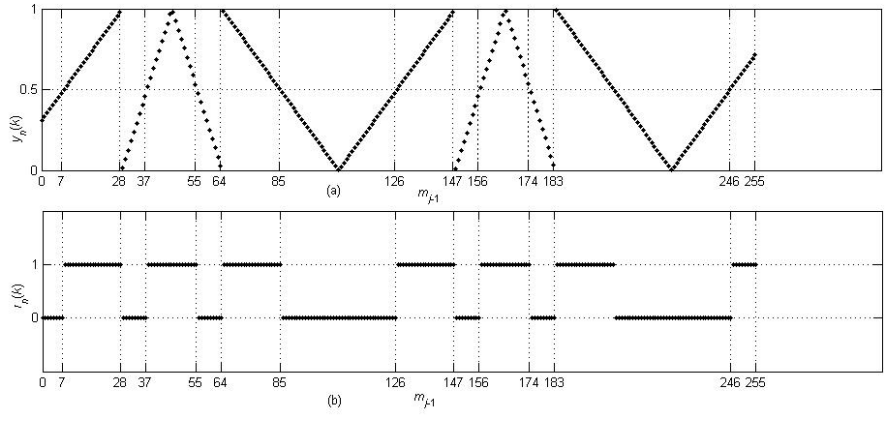


Fig. 4. Plots of plaintext m_{j-1} and the corresponding random number bit $r_n(k)$ and $y_n(k)$. (a) is a plot of m_{j-1} and $y_n(k)$. (b) is a plot of plaintext m_{j-1} and $r_n(k)$

In Eq. (3), z_{j-1}, z_{j-2} and modulo map mod(.) can be got from Eq. (1), Eq. (2) and Eq. (5). The result is presented in Eq. (9) where γ is integer and locates $\left[\frac{y_{n-1}}{256} - \frac{y_{n-2} \times (768 + m_{j-2})}{256^2}\right] * m_{j-1} + \phi_1$ within the range of [-1, 1]. Thus, the modulo map in Eq. (5) is an identity transform, and the operation

$\left[\frac{y_{n-1}}{256}-\frac{y_{n-2}\times(768+m_{j-2})}{256^2}\right]*m_{j-1}+\phi_1$ is a linear transformation of m_{j-1} if m_{j-2} is fixed. Assume that m_{j-1} is input variable, y_n is output and $\{y_{n-2}, y_{n-1}, p, \phi\}$ are constant values. The variable m_{j-1} is integer in the range of [0, 255] and the plaintext block m_{j-2} is set to 1. The corresponding sequence of y_n is $\{y_n(0), y_n(1)...y_n(k)\ldots, y_n(255)\}$ and the corresponding random number bitstream is $\{ r_n(0), r_n(1)...r_n(k)\ldots r_n(255)\}$. The plot of $\{y_n(0), y_n(1)...y_n(k)\ldots y_n(255)\}$ and corresponding m_{j-1} have two distinct parts: one is dense, the other is dilute depicted in Fig.4 (a). Since the cryptanalyst knows the random number bitstream $\{ r_n(0), r_n(1)...r_n(k)\ldots r_n(255)\}$ and the characteristics of distribution, they can easily obtain the value of p from Eq. (8) after getting the count of point a_1 located in dense part and a_2 located in dilute part.

There are still three unknown factors in $\left[\frac{y_{n-1}}{256}-\frac{y_{n-2}\times(768+m_{j-2})}{256^2}\right]\times m_{j-1}+\phi_1$. The cryptanalyst could find numbers of special points just like $\{d_1, d_2, d_3\}$ in Fig.3 (b) and estimate the approximate coordinates of these points. Furthermore, the approximation of $\{y_{n-1}, y_{n-2}, \phi_1\}$ can be got after numbers of $(y_n(k), m_{j-1})$ are substituted in Eq. (9). Furthermore, the cryptanalyst could reduce the error by a large number of statistical computing.

$$\begin{cases} y_n = h\circ \mathrm{mod}(\left[\frac{y_{n-1}}{256}-\frac{y_{n-2}\times(768+m_{j-2})}{256^2}\right]\times m_{j-1}+\phi_1) \\ \phi_1 = y_{n-1}\times\left(6+\frac{m_{j-2}}{256}\right)-y_{n-2}\times 3\times\left(3+\frac{m_{j-2}}{256}\right)+\phi-2\gamma \end{cases} \quad (9)$$

To confirm that algorithm NDF is vulnerable under chosen plaintext attack, a simulation is presented to recover the secret keys of NDF under chosen plaintext attack.

The key of NDF is $\{y_{-1}, y_0, p, \phi\}$ that is set to {0.2, 0.3, 0.6, 0.35}. $\{c_1, c_2\}$ is set to {5.7, 7} as recommendation of the scheme [8]. Fig.4 (a) is a plot of Eq. (9) when m_{j-1} is the input value in the range of [0, 255] and corresponding $y_n(k)$ as output value. The cryptanalyst must recover the keys of NDF only knowing the plaintext $\{m_{j-2}, m_{j-1}\}$ and the random number bit $r_n(k)$ corresponding to $y_n(k)$ in chosen plaintext attack scenario. The proposed chosen plaintext attack is described as follows.

1. In order to locate $\left[\frac{y_{n-1}}{256}-\frac{y_{n-2}\times(768+m_{j-2})}{256^2}\right]\times m_{j-1}+\phi_1$ within the range of [-1, 1], it is better to set y_{i-2} in the range [0.5, 1] and y_{i-1} in the range [0, 0.5), that means r_{i-2} must be equal to 1 and r_{i-1} must be equal to 0. In practical application, this condition is met easily. To simplify the procedure, y_{i-2} is arbitrarily set to 0.8 and y_{i-1} is arbitrarily set to 0.25. Plaintext value m_{j-2} is set to 1. Fig.4 (b) shows the plaintext and the corresponding random number bitstream.
2. As show in Fig.4 (b), we can count the number of points located in the range of [7, 28], and it is equal to 21. Meanwhile, we count the number of points located in the range of [28, 37], and it is equal to 9. Now, we know that a_1 is equal to 21and a_2 is equal to 9. Therefore, the cryptanalyst knows for sure that the approximation value of p is 0.35 from Eq. (8).
3. Put three special points (8, 0.5), (85, 0.5) and (106, 0) into Eq. (9), and then it is easy to obtain the approximation of coefficients $\{y_{i-2}, y_{i-1}, \phi\}$.
4. Repeat steps (1), (2) and (3) with several times, the cryptanalyst could reduce the error rate.

4 Improved Algorithm

4.1 Enhanced NDF

Because the plaintext is mapped into the coefficients of NDF directly, opponent can get these coefficients easily in chosen plaintext scenario. Making further investigation, the opponent could obtain information of other key as showing in section 3. The key problem is that Eq. (1) is linear correlation model and the coefficients of NDF relates to plaintext directly. With the linear correlation model, the opponent can obtain one part of key for NDF in chosen plaintext scenario. Therefore, nonlinear correlation model is necessary to resist the chosen-plaintext attack. In order to make this scheme flexibility in multimedia application, the coefficients must be mapped with the ciphertext. Here, the Eq. (10) is used to instead of Eq. (1). The operation $\oplus$ denotes XOR. The values of a_1 and a_2 are arbitrary positive integer.

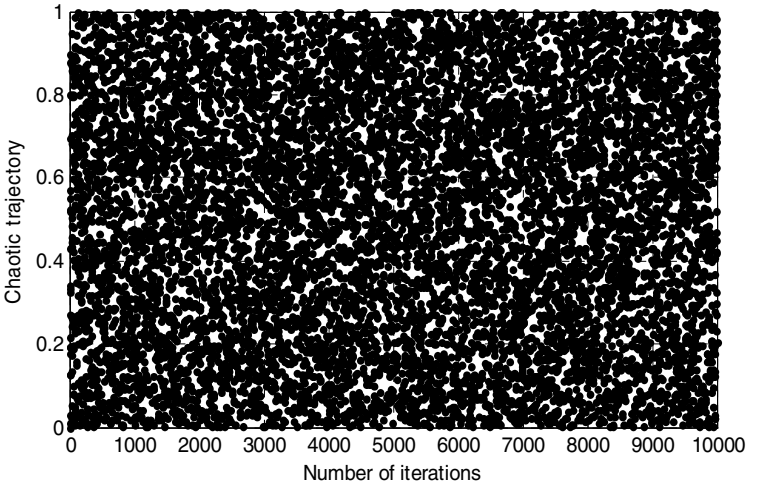
$$\begin{aligned}\xi_{j-2} &= a_0 \oplus C_{j-2} + 3 \\ \xi_{j-1} &= a_1 \oplus C_{j-1} + 3\end{aligned} \tag{10}$$

4.2 Effectiveness of Proposed Algorithm

Kelber [14] has proved that NDF is an ergodic chaotic system with n-D uniform distribution only if the coefficients are set as $z_n \in Z$ and $z_n \neq 0$, the function $h(w)$ is preserved with uniform distribution, and the system is not decomposable. Obviously, in proposed algorithm, z_n is positive integer meeting the first above condition. In a strict sense, the metric entropy is the rate of information generation with respect to the generating partitions of phase space. For 2-D NDF system, the metric entropy should be larger than 2 to obtain 2-bit information at one iteration. Ciphertext is translated to the corresponding ASCII numbers. By means of nonlinear transform, eigenvalues of NDF is obtained as Eq. (10). Then, the metric entropy is showed in Eq. (11), so it satisfies the Kelber condition [14].

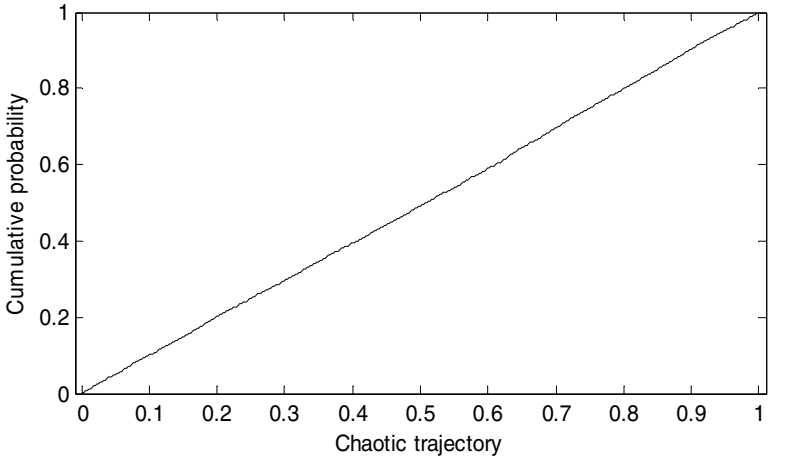
$$H = \sum_{i=1}^{2} \log \mid \xi_i \mid > 1 + 1 = 2 \tag{11}$$

Fig. 5(a) illustrated the trajectory of a second-order NDF map with the conditions that the initial ciphertext is randomly set in range of [0, 255], the initial values are a_1=25, a_2=23, ϕ = 0.6, p=0.35, $y_{-1}(0)$ = 0.2, y_{-2} = 0.3 and the number of iteration is 10,000. From the trajectory of a nonlinear dynamic filter and its distribution function as illustrated in Fig. 5(b), we can see that the chaotic property of the NDF map has the uniform probability density distribution.



(a)

Fig. 5. The uniform property of a variant NDF proposed in our scheme. (a) Trajectory of a variant NDF proposed in our scheme; (b) its distribution function.



(b)

Fig. 5. (*continued*)

5 Simulation Results and Analysis

5.1 Chosen Plaintext Attack

Coefficient of NDF is linearly related to plaintext, and two adjacent chaotic states are closely correlated. From analysis of the correlation between plaintext and random number bitstream in section 3, the coefficient of NDF, which is part of key, is recovered successfully in chosen plaintext attack scenario. In order to solve the problem, the system must guarantee that the plaintext is not closely related to both chaotic trajectory and coefficient of chaotic system. Here, a constructive model is presented in Eq. (10). The XOR operation ensures that the ciphertext is nonlinear related to coefficient. Even in chosen plaintext attack scenario, the opponent can't obtain the coefficient z_{j-2} and z_{j-1} for it doesn't get any knowledge of a_1 and a_2. In a similar way of Fig.4, Fig. 6 is a plot of ciphertext C_{j-1} versus the corresponding random number bit $r_n(k)$ and chaotic trajectory $y_n(k)$. Comparison Fig.4 with Fig.6, it is found that the chaotic trajectory is nonlinear related to ciphertext and the opponent can't get any information about key from this figure.

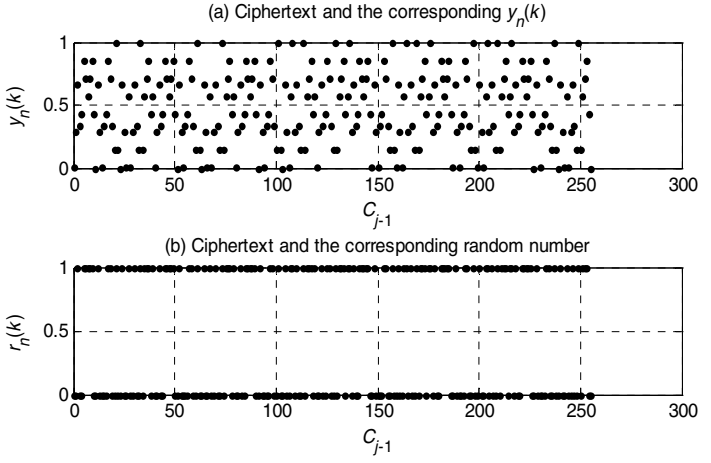


Fig. 6. Plots of ciphertext C_{j-1} and the corresponding chaotic trajectory $y_n(k)$ and random number bits $r_n(k)$ and. (a) is a plot of C_{j-1} versus $y_n(k)$, and (b) is a plot of Ciphertext C_{j-1} versus $r_n(k)$

Table 1. An analysis report for the statistical test suite which produces multiple P-values and proportions including the worst case

STATISTICAL TEST	PROPORTION	P-VALUE	RESULT
Frequency	0. 9880	0. 220159	Success
Block Frequency	0. 9880	0. 299251	Success
Cumulative Sums	0. 9880	0. 494392	Success
Runs	0. 9920	0. 974370	Success
Longest Run	0. 9860	0. 148653	Success
Rank	0. 9940	0. 827279	Success
Non-Overlapping Template	0. 9733	0. 000117	Success
Overlapping Template	0. 9860	0. 016261	Success
Universal	0. 9940	0. 162606	Success
Approximate Entropy	0. 9880	0. 246750	Success
Random Excursions	0. 9844	0. 014741	Success
Random Excursions Variant	0. 9875	0. 082702	Success
Serial	0. 9900	0. 319084	Success
Linear Complexity	0. 9860	0. 236810	Success

5.2 The Random Number Bitstream

The randomness of the random number bitstream is confirmed by the statistical test suite recommended by the U.S. National Institute of Standards and Technology (NIST) [15]. With 300 binary sequences, each of 1,000,000 bits is extracted for testing. They should pass 14 types of statistical tests including approximate entropy, block frequency, linear complexity, frequency, runs, longest run, rank and serial test. All proportion values from the multiple chaotic maps are bigger than the minimum pass rate. Therefore, the sequences are considered as random according to the NIST Special Publication 800–22 [15]. Table 1 is the analysis report from the statistical test suite. The item of proportion is the proportion of sequences passing a test, and the item of distribution of p-value is Chi-square Distribution value. If the value is bigger than 0.0001, it indicates that the sequence is uniformly distributed.

5.3 Multimedia Application Scenario

In multimedia application scenario, the users always move forward through an audio or video at a speed faster than that at which it would usually flow, and play audio or video at arbitrary portion as their wish. For scheme of [7, 8], Eq. (12) is a summarization of generating i^{th} random number bit. It shows that the decoder must obtain all preceding plaintext before getting the random number bit r^i and decoding the i^{th} ciphertext. Unfortunately, in multimedia application scenario, this condition is not always met. In the schemes of [7, 8], they maybe meet these requirements by blocking plaintext and reusing key vector repeatedly. However, in that way, these schemes will be similar to the classical approach to provide compression and encryption separately. And the security is even worse than that scheme for reusing the initial key vector repeatedly.

$$r_i = U(m_{-2}, m_{-1}, \cdots, m_{i-1}, \mathrm{y}_{-2}, \mathrm{y}_{-1}, \phi, p) \tag{12}$$

In proposed model, Eq. (13) is a summarization of generating i^{th} random number bit. For all of available ciphertext, the decoder could get the r^i and decode the i^{th} ciphertext.

$$r_i = U(C_{-2}, C_{-1}, \cdots, C_{i-1}, \mathrm{y}_{-2}, \mathrm{y}_{-1}, \phi, p, a_0, a_1) \tag{13}$$

6 Conclusion

The security of existing approach is analyzed. In the meantime, an enhanced algorithm is proposed with a new nonlinear correlation model between ciphertext and coefficients of NDF. For the nonlinear correlation model, the opponent can't obtain

the keys even in chosen plaintext attack scenario. The randomness of the pseudorandom number bitstream, which reflects security strength of RAC, is confirmed by the statistical test suite which is recommended by the U.S. National Institute of Standards and Technology (NIST). The improved algorithm can also meet users' individual requirements such as fast-forward, rewind or from arbitrary position to play video or audio at a user's discretion. The study in this paper shows that it is better to use ciphertext instead of plaintext as disturbing resource of encryption process for multimedia cryptosystem.

References

1. Wong, K.W., Yuen, C.H.: Embedding compression in chaos-based cryptography. IEEE Transactions on Circuits and Systems II-Express Briefs 55(11), 1193–1197 (2008)
2. Wu, C.P., Kuo, C.C.J.: Design of integrated multimedia compression and encryption systems. IEEE Transactions on Multimedia 7(5), 828–839 (2005)
3. Zhou, J.T., Au, O.C., Wong, P.H.W.: Adaptive chosen-ciphertext attack on secure arithmetic coding. IEEE Transactions on Signal Processing 57(5), 1825–1838 (2009)
4. Witten, I.H., Neal, R.M., Cleary, J.G.: Arithmetic coding for data compression. Communications of the ACM 30(6), 520–540 (1987)
5. Chen, R.C., Pai, P.Y., Chan, Y.K., Chang, C.C.: Lossless image compression based on multiple-tables arithmetic coding. Mathematical Problems in Engineering 2009, Article ID 128317 (2009)
6. Grangetto, M., Magli, E., Olmo, G.: Multimedia selective encryption by means of randomized arithmetic coding. IEEE Transactions on Multimedia 8(5), 905–917 (2006)
7. Mi, B., Liao, X.F., Chen, Y.: A novel chaotic encryption scheme based on arithmetic coding. Chaos Solitons & Fractals 38(5), 1523–1531 (2008)
8. Li, H.J., Zhang, J.S.: A secure and efficient entropy coding based on arithmetic coding. Communications in Nonlinear Science and Numerical Simulation 14(12), 4304–4318 (2009)
9. Wen, J.T., Kim, H., Villasenor, J.D.: Binary arithmetic coding with key-based interval splitting. IEEE Signal Processing Letters 13(2), 69–72 (2006)
10. Kim, H., Wen, J.T., Villasenor, J.D.: Secure arithmetic coding. IEEE Transactions on Signal Processing 55(5), 2263–2272 (2007)
11. Bose, R., Pathak, S.: A novel compression an encryption scheme using variable model arithmetic coding and coupled chaotic system. IEEE Transactions on Circuits and Systems I-Regular Papers 53(4), 848–857 (2006)
12. Li, M.: Generation of teletraffic of generalized cauchy type. Physica Scripta 81(2), 025007, 10 (2010)
13. Zhou, J.T., Au, O.C.: Comments on 'a novel compression and encryption scheme using variable model arithmetic coding and coupled chaotic system'. IEEE Transactions on Circuits and Systems I-Regular Papers 55(10), 3368–3369 (2008)
14. Kelber, K.: N-Dimensional uniform probability distribution in nonlinear autoregressive filter structures. IEEE Transactions on Circuits and Systems I-Fundamental Theory and Applications 47(9), 1413–1417 (2000)

15. Rukhin, A., et al.: A statistical test suite for the validation of random number generators and pseudo-random number generators for cryptographic applications Nat. Inst. Stand. Technol (NIST). Gaithersburg, MD, NIST Special Publication 800-22, May 15 (2001), http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html
16. Kelber, K., Gotz, M., Schwarz, W.: Generation of chaotic signals with n-dimensional uniform probability distribution by digital filter structures. In: Digital Signal Processing Workshop Proceedings, pp. 486–489 (1996)
17. Wang, X.M., Zhang, J.S.: Chaotic secure communication based on nonlinear autoregressive filter with changeable parameters. Physics Letters A 357(4-5), 323–329 (2006)

Recursion Formulas in Determining Isochronicity of a Cubic Reversible System

Zhiwu Liao[1], Shaoxiang Hu[2], and Xianling Hou[3]

[1] School of Computer Science, Sichuan Normal University
liaozhiwu@163.com
[2] School of Automation Engineering, University of Electronic Science and Technology of China
hushaox@126.net
[3] School of Computer Science, Sichuan Normal University
houxianling1985@163.com

Abstract. In this paper, we present a new method to determining the isochronicity of a reversible cubic system based on the recursion formulas and mathematical induction. Unlike the method, which proves the isochronicity of the same system with the integrability, our new method gives the recursion formulas among the period coefficients firstly. Then, based on these recursion formulas and using mathematical induction, the isochronicity is proved.

Keywords: reversible cubic system, recursion formula, weak center, isochronous center, mathematical induction.

1 Introduction

The qualitative theory of differential equations is a very important branch of dynamic system since it can provide qualitative information for the dynamic system.

In 1989, Chicone and Jacobs put forward the concept, weak center, which can answer how many critical periods bifurcate from the center [1]. However, because of the computation complexity, only quadratic systems and some special cubic systems were discussed.

Recently, some efforts for reversible cubic systems were done based on computer algebra [2]-[6]. The main motivation for these methods is that computer algebra can help compute period coefficients of the systems.

Unfortunately, determining isochronicity is a challenge for most of systems since it has to show all period coefficients are equal to zeros [7]-[12]. One feasible alternative for this difficulty is studying the recursion relations among coefficients of the system and proving the isochronicity with the induction algorithm based on the relations.

In order to show the importance of our method, one origin of a specified system proposed by W Zhang etc, whose isochronicity is proved in [6], is considered.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 619–632, 2011.

Unlike the method proposed by W Zhang [6], our new method presents recursion formulas among the period coefficients firstly. Then, based on these recursion formulas and using mathematical induction, the isochronicity is proved.

The remainder of this paper is as follows. Section 2 is devoted to the theory of weak centers. In section 3 some conclusions for weak centers of a reversible cubic system are given. In section 4 we prove the origin of the reversible cubic system is an isochronous center. We also give conclusions finally.

2 The Theory of Weak Centers

Let $V(x, y, \lambda)$ be a family of planar analytic vector fields parameterized by $\lambda \in R^n$ with a *nondegenerate center* at the origin, i. e., the vector field does not have an eigenvalue zero at the origin. $P(r, \lambda)$ denotes the minimum period of the closed orbit passing $(r, 0)$, a point in a sufficiently small open interval $J = (-\alpha, \alpha)$ on x axis.

Definition 1. *Let* $F(r, \lambda_*) = P(r, \lambda_*) - P(0, \lambda_*)$. *The origin is called a* weak center of finite order k *if*

$$F(0, \lambda_*) = F'(0, \lambda_*) = \cdots = F^{(2k+1)}(0, \lambda_*) = 0 \quad \textit{and} \quad F^{(2k+2)}(0, \lambda_*) \neq 0 \quad (1)$$

where the derivatives indicated are taken with respect to the first argument of the function F.

Definition 2. *The origin is called an isochronous center, i. e., all closed orbits surrounding the origin have the same period, that is,* $F^{(k)}(0, \lambda_*) = 0, \forall k \geq 0$.

Definition 3. *Local critical period is a period corresponding to a critical point of the period function which bifurcates from a weak center.*

Lemma 1 (Period Coefficient Lemma). *[1] If* $P(0, \lambda) = 2\pi, \forall \lambda \in R^n$, *then for any given* $\lambda_* \in R^n$,

$$P(r, \lambda) = 2\pi + \sum_{k=2}^{\infty} p_k(\lambda) r^k \quad (2)$$

which is analytic for $|r|$ *and* $|\lambda - \lambda_*|$ *sufficiently small. Moreover,* $p_k \in R[\lambda_1, \cdots, \lambda_n]$, *the noetherian ring of polynomials; for* $k \geq 1, p_{2k+1} \in (p_2, p_4, \cdots, p_{2k})$, *the ideal generated by* $p_{2i}, i = 1, \cdots, k$ *over* $R[\lambda_1, \cdots, \lambda_n]$; *the first* $k > 1$ *such that* $p_k(\lambda) \neq 0$ *is even.*

3 Reversible Cubic System

We consider the differential system

$$\begin{cases} \dot{x} = -y + \phi(x, y, \lambda) \\ \dot{y} = \quad x + \psi(x, y, \lambda) \end{cases} \quad (3)$$

where $\phi(0,0,\lambda)=\psi(0,0,\lambda)=0, \forall\lambda\in R$.

Especially, for C^3, functions $\phi(x,y,\lambda)$ and $\psi(x,y,\lambda)$ are cubic systems

$$\begin{cases}\dot{x}=-y+a_{20}x^2+a_{11}xy+a_{02}y^2+a_{30}x^3+a_{21}x^2y+a_{12}xy^2+a_{03}y^3\\ \dot{y}=x+b_{20}x^2+b_{11}xy+b_{02}y^2+b_{30}x^3+b_{21}x^2y+b_{12}xy^2+b_{03}y^3\end{cases}\tag{4}$$

Definition 4. *A planar vector field is said to be reversible if it is symmetric with respect to a line.*

Consider the cubic system Eq. 4 which is symmetric with respect to y axis. Clearly, such a reversible cubic system is

$$\begin{cases}\dot{x}=-y+a_1x^2+a_2y^2+a_3x^2y+a_4y^3\\ \dot{y}=x+b_1xy+b_2x^3+b_3xy^2\end{cases}\tag{5}$$

with parameter $\lambda=(a_1,a_2,a_3,a_4,b_1,b_2,b_3)\in R^7$. This symmetry ensures that Eq. 5 has a center at the origin.

By Lemma 1 and Definition 1, we have

Theorem 1. *[1] If for a certain $\lambda_*\in R^7$, there is an integer $k\geq 1$ such that*

$$p_2(\lambda_*)=p_3(\lambda_*)=\cdots=p_{2k+1}(\lambda_*)=0\quad \textit{and}\quad p_{2k+2}(\lambda_*)\neq 0\tag{6}$$

The origin is a weak center of order k. Otherwise, the origin is an isochronous center.

Deciding if an origin is an isochronous center using Theorem 1 or Definition 1, is a difficult problem since it needs compute all $F(r,\lambda_*)$s or p_ks. In this paper, we proposed a new method to determine if an origin is an isochronous center by presenting recursion formulas firstly, and then proving the origin is isochronous using these formulas.

Taking polar coordinate

$$x=r\cos\theta,\quad y=r\sin\theta\tag{7}$$

we have

$$\dot{r}=\dot{x}\cos\theta+\dot{y}\sin\theta=r^2G_2(\theta)+r^3G_3(\theta)\tag{8}$$

$$\dot{\theta}=(\dot{y}\cos\theta-\dot{x}\sin\theta)/r=1+rH_1(\theta)+r^2H_2(\theta)\tag{9}$$

where

$$\begin{aligned}G_2(\theta)&=a_1\cos^3\theta+(a_2+b_1)\sin^2\theta\cos\theta\\ G_3(\theta)&=(a_3+b_2)\cos^3\theta\sin\theta+(a_4+b_3)\cos\theta\sin^3\theta\\ H_1(\theta)&=(b_1-a_1)\cos^2\theta\sin\theta-a_2\sin^3\theta\\ H_2(\theta)&=(b_3-a_3)\cos^2\theta\sin^2\theta+b_2\cos^4\theta-a_4\sin^4\theta\end{aligned}\tag{10}$$

Thus

$$\frac{dr}{d\theta}=\frac{r^2G_2(\theta)+r^3G_3(\theta)}{1+rH_1(\theta)+r^2H_2(\theta)}\tag{11}$$

Lemma 2. *[6] The vector field defined by Eq. 11 is analytic and*

$$\frac{dr}{d\theta} = r^2 G_2 + \sum_{k=3}^{\infty} r^k (G_2 A_{k-2} + G_3 A_{k-3}) \tag{12}$$

in a sufficiently small neighborhood of $r = 0$, where

$$A_0 = 1, A_1 = -H_1, A_k = -H_2 A_{k-2} - H_1 A_{k-1}, \forall k \geq 3 \tag{13}$$

Consider the solution of Eq. 11 with $r(0, \lambda) = r_0 > 0$ in the form

$$r(\theta, \lambda) = \sum_{k=1}^{\infty} u_k(\theta, \lambda) r_0^k \tag{14}$$

The initial condition implies

$$u_1(0, \lambda) = 1, u_k(0, \lambda) = 0, \forall k > 1, \lambda \in R^n \tag{15}$$

Replacing r in Eq. 12 with the series Eq. 14 and comparing coefficients of $r_0^k, k = 1, 2, \cdots$, we get the following differential equations

$$\begin{array}{lcl} u_1' & = & 0 \\ u_2' & = & G_2 u_1^2 \\ u_3' & = & (G_2 A_1 + G_3) u_1^3 + 2 u_1 u_2 G_2 \\ \cdots\cdots & & \end{array} \tag{16}$$

where u_k' denotes $\frac{d}{d\theta} u_k(\theta, \lambda)$. Under the initial conditions in Eq. 15, we can obtain their solutions.

$$\begin{array}{lcl} u_1(\theta) & = & 1 \\ u_2(\theta) & = & \int_0^\theta G_2(\xi) d\xi \\ u_3(\theta) & = & \int_0^\theta (G_3 + G_2(2u_2 - H_1)) d\xi \\ \cdots\cdots & & \end{array} \tag{17}$$

Finally, we compute the period $P(r_0, \lambda)$ of the closed orbit $C(r_0)$ through $(r_0, 0)$ from Eq. 9 and

$$\frac{1}{1 + r H_1(\theta) + r^2(\theta)} = \sum_{k=3}^{\infty} r^k A_k \tag{18}$$

then

$$P(r_0, \lambda) = \int_{C(r_0)} dt$$

$$
\begin{aligned}
&= \int_0^{2\pi} \frac{1}{1 + rH_1(\theta) + r^2 H_2(\theta)} d\theta \\
&= \int_0^{2\pi} (1 + \sum_{k=1}^{\infty} r^k A_k) d\theta \\
&= 2\pi + \int_0^{2\pi} \sum_{k=1}^{\infty} r^k A_k d\theta
\end{aligned} \tag{19}
$$

Meanwhile, from Eq. 14, we obtain the following power series expansion

$$
\begin{aligned}
\textstyle\sum_{k=1}^{\infty} r^k A_k &= \textstyle\sum_{k=1}^{\infty} (\sum_{t=1}^{\infty} u_k r_0^t) A_k \\
&= A_1 u_1 r_0 + (A_1 u_2 + A_2 u_1^2) r_0^2 + (A_1 u_3 + 2A_2 u_1 u_2 + A_3 u_1^3) r_0^3 \\
&\quad + (A_1 u_4 + A_2(u_2^2 + 2u_1 u_3) + 3A_3 u_1^2 u_2 + A_4 u_1^4) r_0^4 + \cdots \\
&= p_1' r_0 + p_2' r_0^2 + p_3' r_0^3 + p_4' r_0^4 + \cdots
\end{aligned} \tag{20}
$$

therefore

$$
P(r_0, \lambda) = 2\pi + \sum_{k=1}^{\infty} P_k(\lambda) r_0^k \tag{21}
$$

where

$$
\begin{aligned}
p_1(\lambda) &= \textstyle\int_0^{2\pi} p_1' d\theta = \int_0^{2\pi} A_1 u_1 d\theta = -\int_0^{2\pi} H_1(\theta) d\theta = 0 \\
p_2(\lambda) &= \textstyle\int_0^{2\pi} p_2' d\theta = \int_0^{2\pi} (A_1 u_2 + A_2 u_1^2) d\theta \\
p_3(\lambda) &= \textstyle\int_0^{2\pi} p_3' d\theta = \int_0^{2\pi} (A_1 u_3 + 2A_2 u_1 u_2 + A_3 u_1^3) d\theta \\
p_4(\lambda) &= \textstyle\int_0^{2\pi} p_4' d\theta = \int_0^{2\pi} (A_1 u_4 + A_2(u_2^2 + 2u_1 u_3) + 3A_3 u_1^2 u_2 + A_4 u_1^4) d\theta \\
&\cdots\cdots\cdots\cdots
\end{aligned} \tag{22}
$$

and A_k, u_k, $k = 1, 2, \cdots$ are determined by Eq. 13 and Eq. 17.

3.1 The Algorithm of the Reversible Cubic System

Summary the above procedures in this section, we have

1. Compute $G_2(\theta)$, $G_3(\theta)$, $H_1(\theta)$ and $H_2(\theta)$ using Eq. 10.
2. Compute $A_k, k = 1, 2, \cdots$ using Eq. 13.
3. Compute $\frac{dr}{d\theta}$ using Eq. 11.

4. Compute $u'_k(\theta), k = 1, 2, \cdots$ using Eq. 16 and $u_k(\theta), k = 1, 2, \cdots$ using Eq. 17 one by one.

5. Compute $p_k(\lambda), k = 1, 2, \cdots$ one by one using Eq. 22.

According to above procedures, $p_k(\lambda), k = 1, 2, \cdots$ can be computed one by one. Thus, if all p_ks are equal to zeros, the origin is an isochronous center. That is what we often do for weak focuses.

4 The New Method for Isochronicity

In this section, we analyze the weak center of one reversible cubic system. Consider the system in the following form

$$\begin{cases} \dot{x} = -y - ax^2 + ay^2 + cx^2y \\ \dot{y} = x - 2axy + cxy^2 \end{cases} \tag{23}$$

The isochronicity of the system has been proved by W Zhang in [6] with the integrability. Different from the method of W Zhang, we proposed a new method to prove the isochronicity based on recursion formulas and mathematical induction. In order to the consistency of this paper, the procedures of our method are explained one by one even for the W Zhang's overlapped parts.

4.1 The General Algorithm

Firstly, we compute coefficients about the system using the algorithm presented in Section 3.1.

1. Compute $G_2(\theta)$, $G_3(\theta)$, $H_1(\theta)$ and $H_2(\theta)$ using Eq. 10.

$$\begin{aligned} G_2 &= -a\cos^3\theta + (a - 2a) = -a\cos\theta \\ H_1 &= (-2a + a)\cos^2\theta\sin\theta - a\sin^3\theta = -a\sin\theta \\ H_2 &= 0 \\ G_3 &= (c + 0)\cos^3\theta\sin\theta + (0 + c)\cos\theta\sin^3\theta = c\cos\theta\sin\theta \end{aligned} \tag{24}$$

2. Compute $A_k, k = 1, 2, \cdots$ using Eq. 13.

$$A_0 = 1, A_k = -H_1 A_{k-1} \tag{25}$$

So

$$A_k = a^k \sin^k\theta, \quad k = 0, 1, 2 \cdots \tag{26}$$

3. Compute $\frac{dr}{d\theta}$ using Eq. 11.

$$\frac{dr}{d\theta} = \frac{G_2 A_{k-2} + G_3 A_{k-3}}{G_2 A_{k-3} + G_3 A_{k-4}} = a\sin\theta \tag{27}$$

4. Compute $u_k'(\theta), k = 1, 2, \cdots$ using Eq. 16 and $u_k(\theta), k = 1, 2, \cdots$ using Eq. 17 one by one.

$$\begin{array}{lll} u_1' = 0, & & u_1 = 1 \\ u_2' = G_2 u_1^2 = -a\cos\theta, & & u_2 = -a\sin\theta \\ u_3' = (G_2 A_1 + G_3)u_1^3 + 2u_1 u_2 G_2, & & u_3 = (\frac{1}{2}a^2 + \frac{1}{2}c)\sin^2\theta \\ \quad = (a^2 + c)\sin\theta\cos\theta & & \\ \cdots\cdots\cdots\cdots & & \cdots\cdots\cdots\cdots \end{array} \tag{28}$$

5. Compute $p_k(\lambda), k = 1, 2, \cdots$ using Eq. 22.

$$\begin{aligned} p_1' &= A_1(\theta)u_1(\theta) = a\sin\theta \\ p_1 &= \int_0^{2\pi} A_1 u_1 d\theta = 0 \\ p_2' &= A_1(\theta)u_2(\theta) + A_2(\theta)u_1^2(\theta) = -a^2\sin\theta + a^2\theta = 0 \\ p_2 &= \int_0^{2\pi} (A_1(\theta)u_2(\theta) + A_2(\theta)u_1^2(\theta))d\theta = 0 \\ p_3' &= A_1(\theta)u_3(\theta) + 2A_2(\theta)u_1(\theta)u_2(\theta) + A_3(\theta)u_1^3(\theta) \\ &= a\sin\theta(\frac{1}{2}a^2 + \frac{1}{2}c)\sin^2\theta + 2a^2\sin^2\theta(-a\sin\theta) + a^3\sin^3\theta \\ &= (\frac{c}{2} - \frac{a^2}{2})a\sin^3\theta \\ p_3 &= \int_0^{2\pi} (A_1(\theta)u_3(\theta) + 2A_2(\theta)u_1(\theta)u_2(\theta) + A_3(\theta)u_1^3(\theta))d\theta = 0 \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{29}$$

By Theorem 1, we know that we only need to show the coefficients of $p_{2k}', k = 1, 2, \cdots$, equal to zeros if we want to prove the origin is an isochronous center. Therefore, we only consider the coefficients of p_k, p_k', u_k and $u_k', k = 1, 2, \cdots$. The coefficients of $u_k, p_k, u_k', p_k', k = 1, 2, \cdots$ are written as u_k^*, p_k^*, du_k^*, $dp_k^*, k = 1, 2, \cdots$ respectively in the remainder of this paper.

4.2 Recursion Formulas

In this subsection, we give some preparations for the proof of the isochronous center. Firstly, we give some recursion formulas for the system.

Lemma 3. *The u_k^*s and p_k^*s of the cubic system Eq. 23 have following relation*

$$[(k-2)a^2 + c]u_k^* = \frac{c - a^2}{a}p_k^* - ac(\sum_{i=1}^{k-1} u_i^* u_{i-k}^*) \tag{30}$$

Proof. Because of

$$u'_k = (\frac{p'_k}{a^2\sin^2\theta} - \frac{u_k}{a\sin\theta})G_2 + (\frac{p'_k}{a^3\sin^3\theta} - \frac{u_k}{a^2\sin^2\theta} - \frac{\sum_{i=1}^{k-1} u_i u_{k-i}}{a\sin\theta})G_3$$

we integer both sides of the equation

$$u^*_k = -\frac{1}{k-1}(\frac{dp^*_k}{a} - u^*_k) + \frac{c}{k-1}(\frac{dp^*_k}{a^3} - \frac{u^*_k}{a^2} - \frac{\sum_{i=1}^{k-1} u^*_i u^*_{k-i}}{a}) \tag{31}$$

then

$$(\frac{k-2}{k-1} + \frac{c}{(k-1)a})u^*_k = \frac{dp^*_k}{(k-1)a}(\frac{c}{a^2} - 1) - \frac{c}{(k-1)a}\sum_{i=1}^{k-1} u^*_i u^*_{k-i}$$

times $(k-1)a^2$ both sides

$$[(k-2)a^2 + c]u^*_k = \frac{c-a^2}{a}dp^*_k - ac(\sum_{i=1}^{k-1} u^*_i u^*_{i-k})$$

Lemma 4. *p'_n can be computed using*

$$p'_n = (p'_{n-1}u_1 + p'_{n-2}u_2 + \cdots + p'_1 u_{n-1} + u_n)A_1 \tag{32}$$

Proof. Since

$$\begin{aligned}&(u_1 + u_2 + \cdots + u_m)^{k+1} = (u_1 + u_2 + \cdots + u_m)^k(u_1 + u_2 + \cdots + u_m)\\ &= (u_1 + u_2 + \cdots + u_m)^k u_1 + (u_1 + u_2 + \cdots + u_m)^k u_2 + \cdots\\ &+(u_1 + u_2 + \cdots + u_m)^k u_m\end{aligned} \tag{33}$$

combing with Eq. 20, we have

$$p'_n = (p'_{n-1}u_1 + p'_{n-2}u_2 + \cdots + p'_1 u_{n-1} + u_n)A_1$$

Corollary 1. *The dp^*_k and $u^*_k, k = 1, 2, \cdots$ of the cubic system Eq. 23 can be represented as:*

$$dp^*_k = \frac{k-2}{k-1}a\sum_{i=1}^{k-1} dp^*_i u^*_{k-i} + \frac{c}{(k-1)a}(\sum_{i=1}^{k-1} dp^*_i u^*_{k-i} - a\sum_{i=1}^{k-1} u^*_i u^*_{k-i}) \tag{34}$$

$$u^*_k = \frac{-1}{k-1}\sum_{i=1}^{k-1} dp^*_i u^*_{k-i} + \frac{c}{(k-1)a^2}(\sum_{i=1}^{k-1} dp^*_i u^*_{k-i} - a\sum_{i=1}^{k-1} u^*_i u^*_{k-i}) \tag{35}$$

$$(k-1)a^2 u^*_k = (c-a^2)\sum_{i=1}^{k-1} dp^*_i u^*_{k-i} - ac\sum_{i=1}^{k-1} u^*_i u^*_{k-i} \tag{36}$$

Proof. According to Eq. 31

$$u_k^* = -\frac{1}{k-1}(\frac{dp_k^*}{a} - u_k^*) + \frac{c}{k-1}(\frac{dp_k^*}{a^3} - \frac{u_k^*}{a^2} - \frac{\sum_{i=1}^{k-1} u_i^* u_{k-i}^*}{a})$$

combining with Eq. 32, we have

$$u_k^* = \frac{-1}{k-1}\sum_{i=1}^{k-1} dp_i^* u_{k-i}^* + \frac{c}{(k-1)a^2}(\sum_{i=1}^{k-1} dp_i^* u_{k-i}^* - a\sum_{i=1}^{k-1} u_i^* u_{k-i}^*)$$

so

$$(k-1)a^2 u_k^* = (c-a^2)\sum_{i=1}^{k-1} dp_i^* u_{k-i}^* - ac\sum_{i=1}^{k-1} u_i^* u_{k-i}^*$$

From Eq. 32

$$p_n' = (p_{n-1}' u_1 + p_{n-2}' u_2 + \cdots + p_1' u_{n-1} + u_n)A_1$$

we have

$$dp_n^* = (dp_{n-1}^* u_1^* + dp_{n-2}^* u_2^* + \cdots + dp_1^* u_{n-1}^* + u_n^*)a \tag{37}$$

Replacing u_k^* in Eq. 37 with the right side of Eq. 35, the dp_k^* becomes

$$dp_k^* = \frac{k-2}{k-1}a\sum_{i=1}^{k-1} dp_i^* u_{k-i}^* + \frac{c}{(k-1)a}(\sum_{i=1}^{k-1} dp_i^* u_{k-i}^* - a\sum_{i=1}^{k-1} u_i^* u_{k-i}^*)$$

4.3 The New Method for Determining the Isochronicity

In this subsection, we will give a new method for determining the isochronicity of the same system proposed by W Zhang [6]. Although W Zhang has proved the isochronicity of the system in [6], in order to preserve the consistence of our statement, the isochronicity still is presented as the most important theorem in this subsection.

Lemma 5. *For the system defined by Eq. 23, if*

$$dp_{2i}^* = 0, i = 1, 2, \cdots, k \tag{38}$$

$$u_{2i}^* = -ac^{i-1}, i = 1, 2, \cdots, k \tag{39}$$

then

$$2ku_{2k+1}^* = [(2k-1)c - (2k-3)a^2]u_{2k-1}^* - ac\sum_{i=1}^{2k-2} u_i^* u_{2k-1-i}^*. \tag{40}$$

Proof. For $m = 1$,

$$2u_3^* = (c + a^2).$$

If for all $m \leq 2k$, Eq. 40 hold, we will prove the Equation will hold on $m = 2k+1$. According to Eq. 36, we have

$$2ka^2u^*_{2k+1} = (c-a^2)\sum_{i=1}^{2k-2} dp^*_i u^*_{2k-1-i} - ac\sum_{i=1}^{2k-2} u^*_i u^*_{2k-1-i}.$$

Since $dp^*_{2i} = 0, i = 1, 2, \cdots k$ and $u^*_{2j} \times u^*_{2i} = a^2c^{i+j-2}$ so we only consider the terms of u^*_{2m+1} and dp^*_{2m+1}, where m is an integer.

$$\begin{aligned}
2ka^2u^*_{2k+1} &= (c-a^2)\sum_{i=1,\ \text{i is an odd number}}^{2k} dp^*_i u^*_{2k+1-i} \\
&\quad -ac\sum_{i=1,\ \text{i is an odd}}^{2k} u^*_i u^*_{2k+1-i} \\
&= (c-a^2)\sum_{i=1,\ \text{i is an odd}}^{2k-2} dp^*_i u^*_{2k-1-i}c + dp^*_{2k-1}u^*_2(c-a^2) \\
&\quad -ac\sum_{i=1,\ \text{i is an odd}}^{2k-2} u^*_i u^*_{2k-1-i} - 2acu^*_{2k-1}u^*_2 \\
&= (2k-2)u^*_{2k-1}ca^2 + (c-a^2)dp^*_{2k-1}u^*_2 - 2acu^*_{2k-1}u^*_2
\end{aligned}$$

Dividing a^2 on both sides, above equation becomes

$$\begin{aligned}
2ku^*_{2k+1} &= (2k-2)u^*_{2k-1}c - [(2k-3)a^2 + c]u^*_{2k-1} \\
&\quad -ac\sum_{i=1,\ \text{i is an odd}}^{2k-2} u^*_i u^*_{2k-1-i} \\
&= [(2k-1)c - (2k-3)a^2]u^*_{2k-1} \\
&\quad -ac\sum_{i=1,\ \text{i is an odd}}^{2k-2} u^*_i u^*_{2k-1-i}
\end{aligned}$$

Corollary 2. *For the system defined by Eq. 23, if*

$$\begin{aligned}
dp^*_{2i} &= 0, i = 1, 2, \cdots, k \\
u^*_{2i} &= -ac^{i-1}, i = 1, 2, \cdots, k
\end{aligned}$$

then

$$2ku^*_{2k+1} = [(4k-3)c - (2k-3)a^2]u^*_{2k-1} + [-(2k-3)c + (2k-3)a^2]cu^*_{2k-3}. \quad (41)$$

Proof. Since

$$\sum_{i=1,\ \text{i is an odd}}^{2k-2} u^*_i u^*_{2k-1-i} = c\sum_{i=1,\ \text{i is an odd}}^{2k-4} u^*_i u^*_{2k-3-i} + 2u^*_{2k-3}$$

by Eq. 40, we have

$$(2k-2)u^*_{2k-1} - [(2k-3)c - (2k-5)a^2]u^*_{2k-3} = -ac \sum_{i=1,\ \text{i is an odd}}^{2k-4} u^*_i u^*_{2k-3-i} \tag{42}$$

Because of

$$\begin{aligned} 2ku^*_{2k+1} &= [(2k-1)c - (2k-3)a^2]u^*_{2k-1} \\ &\quad -ac^2 \sum_{i=1,\ \text{i is an odd}}^{2k-4} u^*_i u^*_{2k-3-i} - 2ca^2u^*_{2k-3} \\ &= [(2k-1)c - (2k-3)a^2]u^*_{2k-1} + (2k-2)cu^*_{2k-1} \\ &\quad -[(2k-3)c - (2k-5)a^2]cu^*_{2k-3} - 2ca^2u^*_{2k-3} \\ &= [(4k-3)c - (2k-3)a^2]u^*_{2k-1} + [-(2k-3)c + (2k-3)a^2]cu^*_{2k-3} \end{aligned}$$

Theorem 2. *For the system defined by Eq. 23, if*

$$\begin{aligned} dp^*_{2i} &= 0, i = 1, 2, \cdots, k \\ u^*_{2j} &= -ac^{j-1}, i = 1, 2, \cdots, k \end{aligned}$$

for $k \geq 2$

$$\sum_{i=1,\ i\ is\ an\ odd}^{2k-1} u^*_i u^*_{2k-i} = [(k-1)a^2 + c]c^{k-2}. \tag{43}$$

Proof. For $m = 2$, we have

$$2u^*_3 = (a^2 + c)$$

We assume that Eq. 43 are correct for all $2 \leq m \leq 2k$, thus we will prove the equation is correct for $m = 2k + 1$. According to Eq. 43, we have

$$2u^*_3 = (c + a^2) \tag{44}$$

$$4u^*_5 = (5c - a^2)u^*_3 + (-c + a^2)c \tag{45}$$

$$\cdots\cdots\cdots\cdots$$

$$2ku^*_{2k+1} = [(4k-3)c - (2k-3)a^2]u^*_{2k-1} + [-(2k-3)c + (2k-3)a^2]cu^*_{2k-3} \tag{46}$$

then we times Eq. 44, Eq. 45, Eq. 46 with u^*_{2k-1}, u^*_{2k-3} and u^*_1 respectively and sum these equations both sides, we have

$$k \sum_{i=1,\ \text{i is an odd number}}^{2k+1} u^*_i u^*_{2k+2-i}$$

$$= [(2k-1)c-(k-2)a^2] \sum_{i=1,\ \text{i is an odd}}^{2k-1} u_i^* u_{2k-i}^*$$
$$+[-(k-1)c+(k-1)a^2]c \sum_{i=1,\ \text{i is an odd}}^{2k-3} u_i^* u_{2k-2-i}^*$$
$$= [(2k-1)c-(k-2)a^2][(k-1)a^2+c]c^{k-2}+c[-(k-1)c+(k-1)a^2][(k-2)a^2+c]c^{k-3}$$
$$= k(ka^2+c)c^{k-1}$$

so

$$k \sum_{i=1,\ \text{i is an odd}}^{2k+1} u_i^* u_{2k+2-i}^* = k(ka^2+c)c^{k-1}$$

That is

$$\sum_{i=1,\ \text{i is an odd}}^{2k+1} u_i^* u_{2k+2-i}^* = (ka^2+c)c^{k-1}$$

Theorem 3. *For the system defined by Eq. 23, if*

$$dp_{2i}^* = 0, i = 1, 2, \cdots, k$$
$$u_{2i}^* = -ac^{i-1}, i = 1, 2, \cdots, k$$

we have

$$\sum_{i=1}^{2k-1} dp_i^* u_{2k-i}^* = ac^{k-1}. \tag{47}$$

Proof. For $m = 1$, we have

$$dp_1^* u_1^* = a = ac^{1-1}$$

We assume that Eq. 47 are correct for all $m \leq 2k+1$, thus we will prove the equation is correct for $m = 2k+2$. Since

$$dp_{2s}^* = 0, s = 1, 2, \cdots, k,$$

for all $m \leq 2k+1$, we have

$$\sum_{i=1,\ \text{i is an odd}}^{2k-1} dp_i^* u_{2k-i}^* = ac^{k-1}$$

According to Eq. 47, we have

$$dp_1^* = a \tag{48}$$
$$dp_3^* = au_3^* + adp_1^* u_2^* \tag{49}$$
$$\cdots\cdots\cdots\cdots$$
$$dp_{2k+1}^* = au_{2k+1}^* + adp_1^* u_{2k}^* + \cdots + adp_{2k-1}^* u_2^* \tag{50}$$

then we times Eq. 48, Eq. 49, Eq. 50 with u^*_{2k+1}, u^*_{2k-1} and u^*_1 respectively and sum these equations both sides, we have

$$\begin{aligned}
&\sum_{i=1,\ \text{i is an odd}}^{2k+1} dp^*_i u^*_{2k-i} \\
&= a\sum_{i=1}^{2k+1} u^*_i u^*_{2k+2-i} \\
&\quad -a^2[\sum_{i=1}^{2k-1} dp^*_{2k-i}u^*_i + c\sum_{i=1}^{2k-3} dp^*_{2k-2-i}u^*_i + \cdots] \\
&= a(ka^2+c)c^{k-1} - ka^3c^{k-1} \\
&= ac^k
\end{aligned}$$

Theorem 4. *For the system defined by Eq. 23, we have*

$$\begin{aligned}
dp^*_{2i} &= 0, i = 1, 2, \cdots \\
u^*_{2i} &= -ac^{i-1}, i = 1, 2, \cdots.
\end{aligned}$$

Proof. For m=1, we have

$$\begin{aligned}
dp^*_2 &= 0 \\
u^*_2 &= -a = -ac^{1-1}
\end{aligned}$$

We assume that Eq. 38 and Eq. 39 are correct for all $m \le 2k$, thus we will prove the equation is correct for $m = 2k+2$. According to corollary 1, dp^*_{2k+2} can be represented by

$$\begin{aligned}
(2k+1)adp^*_{2k+2} &= 2ka^2\sum_{i=1}^{2k+1} dp^*_i u^*_{2k+2-i} + c(\sum_{i=1}^{2k+1} dp^*_i u^*_{2k+2-i} - a\sum_{i=1}^{2k+1} u^*_i u^*_{2k+2-i}) \\
&= (2ka^2+c)\sum_{i=1}^{2k+1} dp^*_i u^*_{2k+2-i} - ac\sum_{i=1}^{2k+1} u^*_i u^*_{2k+2-i}
\end{aligned}$$

By theorem 2 and 3, we have

$$\begin{aligned}
(2k+1)adp^*_{2k+2} &= (2ka^2+c)\sum_{i=1}^{2k+1} dp^*_i u^*_{2k+2-i} - ac\sum_{i=1}^{2k+1} u^*_i u^*_{2k+2-i} \\
&= (2ka^2+c)ac^k - ac(2ka^2+c)c^{k-1} \\
&= 0
\end{aligned}$$

Thus, for all $i \ge 1$, $dp^*_{2i} = 0$.

Theorem 5. *The center of the system defined by Eq. 23 is an isochronous center.*

By theorem 4 and 1, the proof of theorem 5 is very clear.

5 Conclusions

In this paper, we have analyzed the coefficients of the reversible cubic system defined by Eq. 23 and give some recursion formulas for its coefficients. Based on these formulas, the origin of this system is an isochronous center is proved. This proof gives a new method to determining the isochronicity of the same system proposed by W Zhang [6].

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Nos 60873102 and 60873264), National Key Basic Research Program Project of China (No 2010CB732501) and China Postdoctoral Science Foundation of China (No 20070410843).

References

1. Chicone, C., Jacobs, M.: Bifurcation of critical periods for plane vector fields, Amer. Math. Soc. Transactions 312, 433–486 (1989)
2. Chen, X.W., Zhang, W.N.: Weak Centers of a Cubic Isochrone Perturbed by Fourth Degree Homogeneous Polynomial. Journal Of Sichuan University (Natural Science Edition) 02, 251–255 (2004)
3. Chen, X.W., Zhang, W.N.: Decomposition of algebraic sets and applications to weak centers of cubic systems. Journal of Computational and Applied Mathematics 232, 565–581 (2009)
4. San, B., Zhu, S.M.: Isochronous center of a cubic reversible system. Journal of Systems Science and Mathematical, 129–135 (2008)
5. San, B., Li, Y.H., Zhu, S.M.: Indirect Methods for Deriving Center Conditions. Acta Mathematicae Applicatae Sinica 30, 138–145 (2007)
6. Zhang, W., Hou, X., Zeng, Z.: Weak centers and bifurcation of critical periods in reversible cubic systems. Computers and Mathematics with Applications 40(6-7), 771–782 (2000)
7. Chavarriga, J., Garca, I.: Isochronous centers of cubic reversible systems. In: Dynamical Systems, Plasmas and Gravitation. Lecture Notes in Physics, vol. 515, pp. 255–268 (1999)
8. Gin, J., Grau, M.: Characterization of isochronous foci for planar analytic differential systems. Proc. Roy. Soc. Edinburgh Sect. A 135(5), 985–998 (2005)
9. Chavarriga, J., Gin, J., Garca, I.: Isochronous centers of cubic systems with degenerate infinity. Differential Equations and Dynamical Systems 7(1), 49–66 (1999)
10. Mardesic, P., Rousseau, C., Toni, B.: Linearization of Isochronous Centers. Journal of Differential Equations 121(1), 67–108 (1995)
11. Sabatini, M.: Charactering isochronous centres by Lie brackets. Differential Equations Dynam. Systems 5(1), 91–99 (1997)
12. Christopher, C., Devlin, J.: Isochronous Centres in Planar Polynomial Systems. SIAM Jour. Math. Anal. 28, 162–177 (1999)

Skeletonization of Low-Quality Characters Based on Point Cloud Model

X.L. Hou[1], Z.W. Liao[1,*], and S.X. Hu[2]

[1] School of Computer Science, Sichuan Normal University, Chengdu, Sichuan, China
houxianling1985@163.com,
liaozhiwu@163.com
[2] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, china
hushaox@uestc.edu.cn

Abstract. Skeletonization of low-quality Characters (LCs) is a very difficult problem. Since only detected contours (DCs) are known, existing methods focus on how to extract skeletons only from well located real contours (RCs), named real contour model (RCM), perform very badly. A new model, named point cloud model (PCM) is proposed to replace RCM in extracting skeletons for LCs. PCM can preserve more information for LCs and can obtain satisfied skeletons for LCs based on principal curves. The experimental results also show that our method proposed in this paper can obtain satisfied skeletons for LCs, especially in preserving topology and being consistent with the human perception even in serious quality reduction.

Keywords: Low-quality Character (LC), Ideal Character (IC), Principal Curve (PC), Skeletonization, Point Cloud Model (PCM).

1 Introduction

Skeletons extraction/skeletonization of characters is an important area in pattern recognition [1]-[4]. However, skeletonization for characters, especially for Low-quality characters (LCs) is a challenge.

Skeleton is a set of lines or curves, which passes through the data centers [1]-[2]. It is a very important feature with the virtue of maintaining the shape of the topological and geometrical properties. There are lots of efforts for skeletonization of characters and most of them assume that the RCs are detected well.

Two well known types of skeletonization are Symmetry axis analysis (SAA) and thinning both proposed by Blum and then followed up by some researchers [5]-[9]. SAA extracts skeletons by looking for symmetry axis of DCs while thinning get symmetry axis from point's growth [5]. However, both methods still have disadvantages such as, disconnections of resulting skeletons and generated unwanted artifacts (see Fig. 1(b)-(c)).

* Corresponding author.

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 633–643, 2011.

Some researchers suggest that shape decomposition (SD) can be used to get connected skeletons and to suppress small artificial branches [10]-[13]. SD can get satisfied skeletons for complex shapes (see Fig. 1(d)). However, it is very complex both in theory and applications. In addition, for some distortions which damage RCs greatly, SD will fail to obtain satisfied skeletons.

One important work proposed by some statist in [14] extracts skeletons of "one-stroke" characters using principal curves (PCs), which is a well known method for data dimensional reduction. But its applied field is very limited because of the assumption of "one-stroke".

(a) A character. (b) SAA. (c) Thinning (d) SD

Fig. 1. A character and its skeletons extracted by different methods. The artifacts in (c) and (d) are indicated by red circles. The SD has the best performance in these three methods for its unbroken strokes and the least artifacts. The thinning makes many artifacts in its skeleton while the strokes of the SAA are composed by many disordered discrete points.

In summary, most of existing methods can not obtain satisfied skeletons in some distortions which will damage the RCs, named quality reductions (QRs). Since only DCs are located, these methods imply that the RCs of the characters are the same as their DCs. Thus, the satisfied skeletons can be extracted only from DCs. The characters, whose DCs are the same as their RCs, are called ideal characters (ICs). However, in QRs, the RCs are damaged which leads to unsatisfied skeletons.

In this paper, we propose a new method for skeletonization of LCs based on PCM because points can preserve more information in quality reduction (QR) and provide a unified model for LCs and ICs.

Based on PCs, the skeletons are extracted by two steps: firstly, primary skeletons (PSs) are extracted by principal curves. After that, the final skeletons can be obtained by connecting the PS segments.

Following above discussing, our main contributions in this paper are

1) Propose a new model to skeletonization of LCs;
2) Develop a method for extraction primary skeletons of LCs based on PCs.

The remainder of this paper is organized as follows: section 2 introduces backgrounds, how to extract primary skeletons is proposed in section 3; connecting primary skeletons is discussed in section 4; experiments are in section 5; finally, the conclusion and acknowledgements.

2 Backgrounds

Skeleton is an important feature for character recognition and one popular definition of skeleton is [1]-[2]:

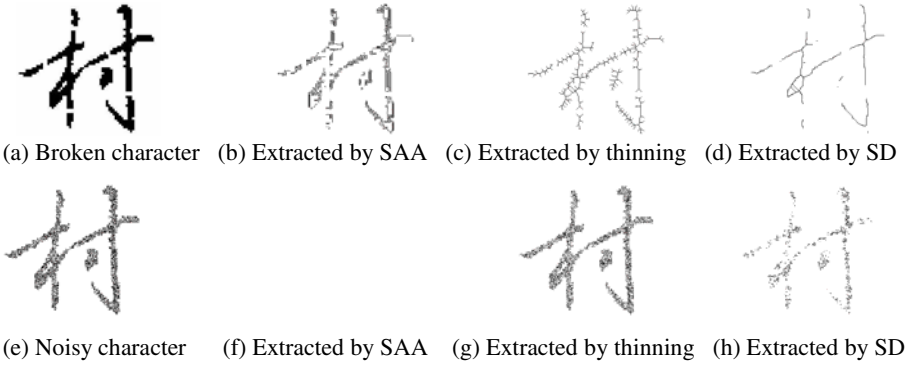
Definition 1. SKELETON of a character is a collection of curves passing through the centers of the strokes, whose connectivity and topology are consistent with the original character.

In section 1, we argue that skeletonization for LCs is still a challenging problem and the main reason for this situation is that the underlying assumption, which the contours of the corresponding ICs for LCs can be located well, is wrong. In order to start our discussion in more formal style, some related terms are introduced firstly.

Definition 2. QUALITY REDUCTION (QR) is defined as transformations for a character which reduce the quality of the character. A character affected by one or more than one type of QR is called a LOW-QUALITY CHARACTER (LC).

In fact, the most serious influence for QRs is the damage for RCs which causes PCM to fail in skeletonization. In order to distinguish RCs and DCs for LCs more clearly, we give their formal names.

Definition 3. DETECTED CONTOUR (DC) for a LC is the contour for the LC while The REAL CONTOUR (RC) for a LC is the contour for the corresponding IC.



(a) Broken character (b) Extracted by SAA (c) Extracted by thinning (d) Extracted by SD

(e) Noisy character (f) Extracted by SAA (g) Extracted by thinning (h) Extracted by SD

Fig. 2. Skeletons extracted by existing methods for LCs

It should be indicated that the RC for an ICC is the same as its DC while these two contours are different for a LC. One example is show in Fig. 2, which is the skeletons for a noisy character in (a) and a broken character in (e).

In Fig. 2(e), the stroke points of the noisy character are very sparse. Skeletonization by SAA for the sparse character has no skeleton points since SAA finds the symmetry centers only from isolated stroke points and patches with a constant width parameter (see Fig. 2(f)). Sparse stroke points also make the skeletonization using thinning and SD performs badly (see Fig. 2(g) and see

Fig. 2(h)). For the broken character, skeletons extracted by existing methods are also broken and have many artifacts.

Until now, we have not found any explicit models for skeletonization of characters. But the existing skeletonization algorithms are based on an underlying model, named real contour model (RCM), assuming that the DCs of the characters are the same as their RCs and the skeletons can be extracted only from their DCs.

Definition 4. The DC is the same as its RC and the skeleton can be extracted only from the DC. This skeletonization model is named as REAL CONTOUR MODEL (RCM).

However, based on RCM, SAA and thinning can not obtain ``good" skeletons even for ICs (see Fig. 1). For LCs, they and SD have to meet the difficulties for lost and falsely located RCs either.

Just as above discussion, since skeletonization based on DCs is unsuitable for the LCs, a new substitute should be proposed to replace the DCs. This substitute should be a lower-level feature to preserve more information in QRs. Considering the complex cases for the QRs, the lowest feature-the stroke points of the LCCs can be used for skeletonization.

Definition 5. POINT CLOUD is a point collection in which is stroke points of a LC.

Definition 6. The model that the whole points of a LC are considered as a point cloud and skeletonization from this point cloud is named as POINT CLOUD MODEL (PCM).

3 Extracting Primary Skeleton (PS)

In this section, we focus on extraction one-pixel width PSs to meet the symmetry for skeletonization.

3.1 Principal Curves and Skeletons of LCs

PCs are smooth self-consistent curves which pass through the "middle" of a d-dimension probability distribution or data cloud [14]-[15], whereas the medial axis is a set of smooth curves or lines that have minimum projection sum distance for stroke points. Therefore, we can use the PCs of the data set to approximate the medial axis of the LCs.

A k -segments algorithm for finding principal curves was proposed by Verbeek in 2000 [16], which by means of local principal component analysis (PCA) finds k-segments, and then connects the principal curves using Hamiltonain path. But its applied field is very limited because of the assumption that the character is "one-stroke". However, we use the modified k-segments algorithm for extraction the medial axis of LCs. The medial axis extracted by k-segments algorithm is named as primary skeleton (PS) and the PS is composed by some curves which are named as primary skeleton segments (PSSs).

Let $X = (x_1, x_2, ..., x_m)$, where x_i is a stroke point, m is the number of the stroke points. $S = \{s_1, s_2, ..., s_k\}$ is the set of PSSs. V ($V = \{v_1, v_2, ..., v_k\}$) is the stroke region of the LC.

$$v_i = \{x \in X \mid i = \arg\min_j d(x, s_j)\} \tag{1}$$

where, v_i contains the all data points for which the ith segment is the closest.

The objective for skeletonization is to find S that minimizes the total squared distance of stroke points:

$$\sum_{i=1}^{k} \sum_{x \in V_i} d(x, s_i)^2 \tag{2}$$

3.2 Primary Skeleton Extraction

The basic idea for skeletonization using PCs is from $k = 1$ incrementally increasing the value of k until it meets a predefined condition. The new line is added by the key point, which is selected according to the fastest decrease of the objective function. After adding the key point x_q, a new region V_q is formed by the nearest points of x_q. The s_q is the principal component segment in region V_q, updating segments S and regions V, through iteration, form the PS for the LC.

The block diagram for extraction the PS of a LC using the modified k -segments algorithm shown in Fig. 4(a). The Algorithm is composed of three steps: initialization, adding a new key point and computing s_q for the qth regions, update S and V .The steps are as follows:

1) **Initialization:** Input data (the object points of the LC) $X = (x_1, x_2, ..., x_m)$, compute the first PC line s_1, and the cut off the line at $3\sigma / 2$ on each side of its center of gravity, where σ^2 is the variance of the data when projected orthogonally to the first principal component of the data.
   - Set $k = 1$, the initial line segment is $S = \{s_1\}$, corresponding region is $V_1 = X$ and $k_{\max} = 3$.
2) **Adding a new key point:** Calculate a new key point x_q. This new key point should satisfy:

$$x_q = \inf\left\{x_q \mid \sum_{i=1}^{m} \min\left(d\left(x_i, x_q\right), nearest\_dt\_line\left(x_i\right)\right)\right\} \tag{3}$$

where $d\left(x_i, x_q\right) = \| x_i - x_q \|^2, \left(i = 1...m\right)$, $nearest\_dt\_line\left(x_i\right) = \min_{j=1...k} d\left(x_i, s_j\right)$.

The points labeled by x_q form the region V_q

$$V_q = \left\{x \in X \mid \left\|x - x_q\right\| \le \min d\left(x, s_j\right), j = 1, 2, ...k\right\} \tag{4}$$

If $\left|V_q\right|$ (the number of elements in V_q) $\le k$ max ,stop and output PSSs; otherwise, compute the s_q for V_q, let $k = k + 1$, go step 3.

3) Updating segments S and regions V

- Label k regions $V_l, l = 1, 2, ..., k$, whose clustering centers are $S = \{s_1, s_2, ..., s_k\}$, according to

$$V_l = \{x_i \mid d(x_i, s_l) \le \min(d(x_i, s_l), l = 1, 2, ..., k)\} \tag{5}$$

- For $i = 1:k$, compute s_i for V_i. Then $S = \{s_1, s_2, ..., s_k\}$.

4) **Carry on step 2-3 iteratively until $|V_q| \le k$ max and output the PSs.**

4 Connecting Primary Skeleton Segments

Since the PSSs are not connected, we will discuss how to connect PSSs conforming to human conception and preserving topology. In order to get satisfied skeletons according to our knowledge, imposing priors is a feasible way. In this section, the priors are imposed using a cost function followed by the work of [18]. And then, the procedures for connecting PSSs are presented.

4.1 The Cost Function [18]

The main objective for optimal theory is to find optimal solutions in some senses for illness problems. It starts from defining a cost function accroding to some priors. That is, low costs are given to desired patterns and the undesired patterns are penalized by high costs. Then the main target is convert to finding an optimal resolution with minimum cost. Genenrally, it is defined as

$$f^* = \arg\min_f C(f) \tag{6}$$

In this paper, PS is a set of PSSs. In order to obtain skeletons maintaining topology of the original characters and being consistent with human perceptions, we must study the connection method for PSSs.

We first need to determine the set of the sites, that is, the set includes all objects need to be considered. In this paper, the set of sites is the PS, where each element is a PS segment, that is, a site. Assume that there are k segments in the PS, the set of sites is defined as $S = \{1, ..., k\}$.

The sites in S are related to one another via a neighborhood system. A neighborhood system for S is defined as $N = \{N_i \mid \forall i \in S\}$ where N_i is the set of sites neighboring i. Neighborhood system in this paper is a global neighborhood, in which each segment is adjacent to all the rest of segments. A label is given to each of site. Let L be a set of labels. In the discrete case, a label assumes a discrete value in a set of M labels $L = \{1, ..., M\}$.

The labeling problem is to assign a label from the label set L to each of the sites in S. The set $f = \{f_1, ..., f_k\}$ is called a labeling of the sites in S with the labels in L.

In our framework, the definition of costs must follow the rule of setting wanted pattern with small energy while the unwanted with large energy. In this issue, the two

segments with small cost should be connected while two segments with large cost will not be connected. Thus, by looking for pair-wise segments with small costs and connect them, we can obtain skeletons for LCs, which can meet the requirements of "good" skeletons. Followed [18], the cost defined between a pair of segments is

$$C(f) = \sum_{\{i,i'\}\in N} [\alpha_1 d_1(i,i') + \alpha_2 d_2(i,i') + \alpha_3 d_3(i,i') + \alpha_4 d_4(i,i')] \tag{7}$$

where $\alpha_i, i = 1,...,4$ are super parameters and are chosen to make the CE comparable.

$d_1(i,i')$ is the distance between the segment i and i', which is defined as

$d_1(i,i') = \inf\{d(x,y) \mid \forall x \in i, \forall y \in i'; \{i,i'\} \in C_2\}$

$d_2(i,i')$ is the angle between segment i and i'.

$d_3(i,i')$ is difference in orientation of the segments i and i' with the current course. The current course is the trend of the strokes direction of connecting i and i' defined as [18]

$$\phi_j = \tan^{-1}\left(\frac{L_i \sin\theta_i + \alpha L_{i'} \sin\phi_{i'}}{L_i \cos\theta_i + \alpha L_{i'} \cos\phi_{i'}}\right)$$

where L_i measures the length of segment i, α puts some stiffness in the stroke of connecting i and i' as it is traced.

$d_4(i,i')$ is the length of the connected line segment.

4.2 Algorithm of Connection Primary Skeletons

The basic idea for connecting PSs is to update the set of sites and their labels using iterative greedy algorithm until all segments have been connected. The results are the final skeletons. The block diagram of algorithm is shown in Fig. 4(b), the basic steps are:

Assume there are k segments for PS, the set of sites is $S = \{1,...,k\}$.

1) Initialization: Set $S = \{1,...,k\}$, $f = \{f_1 = 1,..., f_k = k\}$ and the threshold of cost T.
2) $\forall$ segments i
   For $i' = 1:k$ and $j \neq i$
   - Compute the $C(f)$ between i and j using Eq. 7.
   - If $C(f) < T$, connect i and j. Update the set of sites S and f.
3) Carry on step 2 iteratively until the costs of any pair wise segments are large than T.
4) Output the final skeletons.

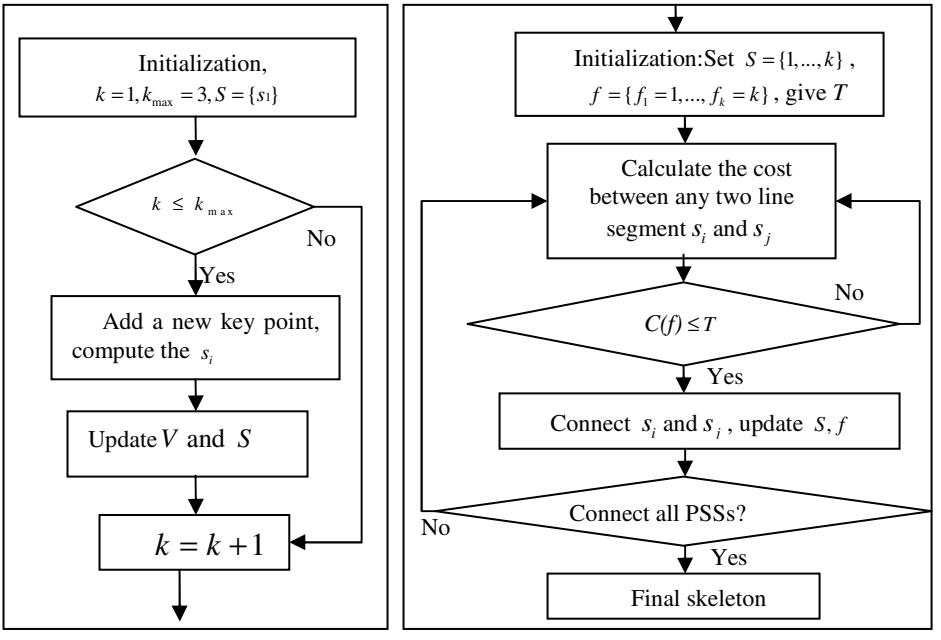


Fig. 3. The flow diagrams for extraction primary skeletons (see (a)) and connection primary skeletons (see (b))

5 Experiments

We compare our method with the thinning using mathematical morphology since thinning is a robust method for LCs while SD and SAA will fail for most of LCs (see Fig. 2). However, thinging can not obtain satisfied skeletons for LCs shown on the second column of Fig. 5. The undesired patterns for the skeletons extracted by thinning include artifacts, brokens etc (see the second column of Fig. 5).

In order to evaluate the skeletons more objectively, the first row of Fig. 5 shows original LCs. That is, the skeletons for the LCs should be indentical to the skeletons for the corresponding ICs which means that the skeletons for LCs should fit corresponding ICs well (see the fourth column of Fig. 5).

The second column is the skeletons extracted by thinning and the last column is the skeletons extracted by our method. Note that there are so many unwanted artifacts and the brokens for skeletons extracted thinning (see the second column of Fig. 5) which coincides the above discussion. Skeletons obtained by our method are one pixel width without the brokens and artifacts (see the third column of Fig. 5) which means the proposed method can obtain satisfied skeletons even in serious quality reduction.

We give the primary skeletons on the third column of Fig. 5 where the original LCs are represented by black points and the primary skeletons are represented by the purple thinning line segments. It should be indicated that there are differences among

the third and the fourth columns. That is, the PSs represented by purple thinning line segments in the third column are not connected especially in some complex regions and broken regions. It is the reason for connection step in our framework.

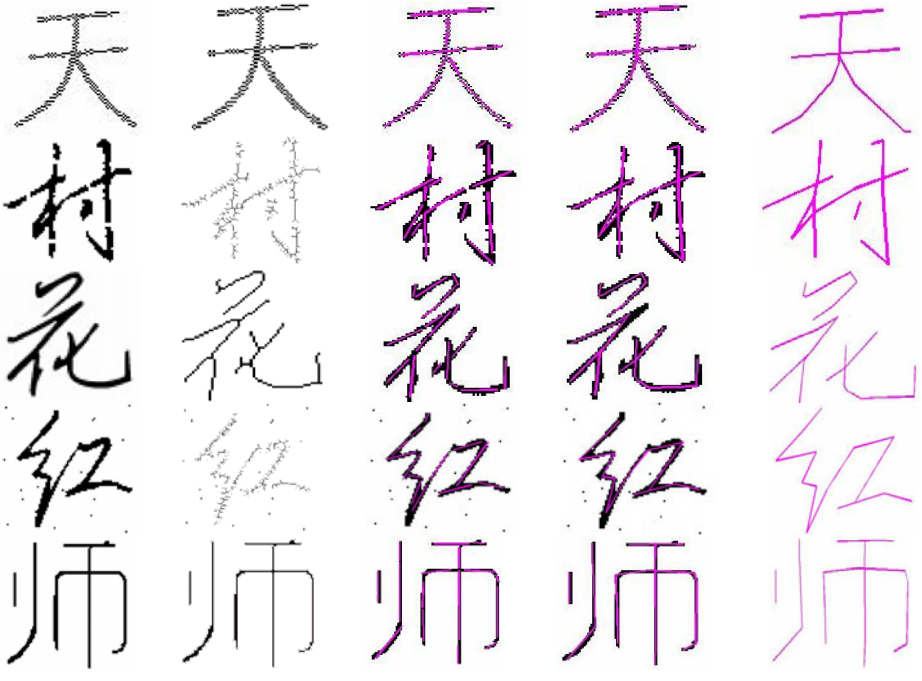


Fig. 4. Skeletons for LCs

The fourth column overlays the final skeletons on the original LCs, where the original LCs are represented by black points and the final skeletons are represented by the purple thinning line segments. The skeletons fit the original LCs well, which means that the skeletons preserve topology of LCs well and pass through the centers of the original LCs. Moreover, no artifacts and the brokens in the final skeletons for proposed method indicates that they conforms with human conception.

In summary, the skeletons extracted by proposed method can obtain one pixel width, preserving topology of LCs well, passing through the centers of the original LCs and conforming with human conception skeletons even in serious QRs. It demonstrates the promising power for our framework in skeletonization of LCs.

6 Conclusion

In this paper, we developed a new framework for skeletonization of LCs based on PCM. PCM can preserve more information in quality reduction (QR) and provide a unified model for LCs and ICs. Thus, the skeletons extracted by proposed method can obtain one pixel width skeletons with good natures including preserving topology of

LCs well, passing through the centers of the original LCs and conforming with human conception even in serious QRs. These good natures show us the promising power for proposed method in skeletonization of LCs.

Ackonwlegements

This paper is supported by the National Natural Science Foundation of China (Nos 60873102, 60573125, 60973157 and 60873264), National Key Basic Research Program Project of China (No 2010CB732501), Open Foundation of Visual Computing and Virtual Reality Key Laboratory of Sichuan Province (No J2010N03), and Innovation Foudation for Scientific Research to Post-gradutate Students of Sichuan Normal University.

References

1. Lam, L., Lee, S.W., Suen, C.Y.: Thinning methodologies-A comprehensive survey. IEEE Trans. Pattern Anal. Machine Intell. 14, 869–885 (1992)
2. Liao, Z.W.: A survey for 2-D skeletonization, Acta of Sichuan Normal University, vol. 32, No. 5676-688 (2009)
3. Klette, G.: A comparative discussion of distance transforms and simple deformations in digital image processing (a survey). Machine Graphics and Vision International Journal 12(2), 235–256 (2003)
4. Aslan, C.: Disconnected skeletons for shape recognition. MPh thesis for Middle East Technical University (May 2005)
5. Blum, H.: Biological shape and visual science (Part 1). J. Theo. Biol. 38, 205–287 (1973)
6. Liao, Z.W., Wang, Z.S., Hu, S.X.: Skeletonize multi width ribbon-like shapes based on difference images and Frenet Frame. In: 2008 IEEE International Conference on Systems, Man, and Cybernetics, Singapore, October 12 - 15 (2008)
7. Tang, Y.Y., You, X.G.: Skeletonization of ribbon-like shapes based on a new wavelet function. IEEE Trans. Pattern Anal. Mach. Intell. 25(9), 1118–1133 (2003)
8. You, X., Chen, Q., Fang, B., Tang, Y.Y.: Thinning character using modulus minima of wavelet transform. International Journal of Pattern Recognition and Artificial Intelligence 20(3), 361–375 (2006)
9. You, X., Tang, Y.Y.: Wavelet-Based Approach to Character Skeleton. IEEE Transactions On Image Processing 16(5), 1220–1231 (2007)
10. Simon, J.C.: A Complemental Approach to Feature Detection, in From Pixels to Features. In: Simon, J. C. (ed.), pp. 229-236. Elsevier Science, Amsterdam(1989)
11. Zou, J.J., Yan, H.: Skeletonization of ribbon-like shapes based on regularity and singularity analyses. IEEE Trans. Syst. Man. Cybern.B, Cybern. 31(3), 401–407 (2001)
12. Morrison, P., Zou, J.J.: Triangle Refinement in a Constrained Delaunay Triangulation Skeleton. Pattern Recognition 40(10), 2754–2765 (2007)
13. Zou, J.J.: Efficient Skeletonization Based on Generalized Discrete Local Symmetries. Optical Engineering 45(7), 1–7 (2006)

14. Kégl, B., Krzyzak, A., Linder, T., Zeger, K.: Learning and Design of Principal Curves. IEEE Trans. Pattern Analysis and Machine Intelligence 22(3), 281–297 (2000)
15. Kégl, B., Krzyzak, A.: Piecewise linear Skeletonization using principal curves, IEEE Trans. Pattern Anal. IEEE Trans. Pattern Anal. Mach. Intell. 24(1), 59–74 (2002)
16. Verbeek, J.J., Vlassis, N.: A -segments algorithm for finding principal curves. Pattern Recognition Letters 23, 1009–1017 (2002)
17. Li, S.Z.: Markov random field modeling in computer vision, 3rd edn. Springer, Heidelberg (2009)
18. Whichello, A.P., Yan, H.: Reconstruction of character skeletons using Gabor filter features. Electronics Letters 37(22), 1911–1912 (1995)

Family of Curves Based on Riemann Zeta Function

Carlo Cattani

dipMat, University of Salerno,
Via Ponte Don Melillo, 84084 Fisciano (SA) Italy
ccattani@unisa.it

Abstract. In this paper some orthogonal functions defined by the Riemann zeta function are studied. In particular, it is shown that they generalize the harmonic functions and are related to the harmonic wavelets. Through their plots it is seen that they are bounded, self crossing with some typical symmetries.

AMS Classification:42C40, 42C10, 42C15, 65T60, 28A80, 28A10.

Keywords: Riemann zeta function, self-intersecting curves, harmonic wavelet.

1 Introduction

In recent papers [1,2,12,15,16,17,18,24] it has been shown that wavelets and, in particular, harmonic wavelets [5,6,7,8,20] play an important role in the analysis of scale depending functions. The scale invariance of fractals [14,19] can be easily understood through the linearly dependent wavelet coefficients, and also no-where differentiable functions such as the Riemann-Weierstrass [23] function, can be projected into a wavelet basis (under suitable hypotheses). The concept of scale is a characteristics feature in physics, and usually mathematical models, which try to describe complex phenomena, must be adapted to different scales and specifically defined, for each scale, in a different way. On the other hand, the concept of scale has a direct meaning in the interpretation of complexity. Many complex phenomena, and chaotic dynamical systems show some scale dependence [1,2,12,14,15,24].

Wavelets are some special scale-depending functions [9,11,18,24] localized in space (or frequency), with finite energy and a (slow/fast) decay to zero. The localization property is a key factor that make these functions a powerful tool to analyse $L_2(\mathbb{R})$ functions. In fact, any finite energy function can be represented by an infinite series of wavelets whose coefficients describe the projection into a given scale by "local" oscillation amplitude.

In the following these two concepts of scale and oscillation, will be investigated with respect to a family of functions derived from the Riemann zeta function. Some similarities between the harmonic wavelets and these family of functions will be commented, with the aim to define a wavelet basis on Riemann zeta

B. Murgante et al. (Eds.): ICCSA 2011, Part IV, LNCS 6785, pp. 644–657, 2011.

function. This would help us to better understand the properties of the zeta function [4,21,22].

2 Bounded Self-intersecting Functions

In this section we will consider some examples of oscillating, self-crossing, bounded functions with real or complex values.

A trivial example of oscillating functions with bounded parametric plots are the harmonic functions

$$\sin x = i\frac{e^{-ix} - e^{ix}}{2} \quad , \quad \cos x = \frac{e^{-ix} + e^{ix}}{2}$$

They are orthogonal and respectively odd and even functions.

2.1 Rieman-Weierstrass Function

As first example let us consider the generalized Riemann-Weierstrass function (Fig. 1)

$$f(x) \stackrel{\text{def}}{=} \frac{1}{\pi}\sum_{k=1}^{\infty} k^{-a} e^{i\pi k^a x} \qquad (a > 1) \ , \tag{1}$$

which is, in trigonometric form,

$$f(x) \stackrel{\text{def}}{=} \frac{1}{\pi}\sum_{k=1}^{\infty} \left[k^{-a}\left(\cos \pi k^a x + i \sin \pi k^a x\right)\right] \ , \qquad (a > 1) \ .$$

In practical computation the upper limit of the series runs up to a finite value $n_{\max} < \infty$. It should be also noticed that this function is self-similar, oscillating, no-where differentiable and bounded in $\mathbb{R}^2$, as can be seen through its parametric representation (Fig. 1).

From the real part of Eq. (1), we can derive the Weierstrass function [23]

$$\begin{gathered} f(x) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} a^k \cos 2\pi b^k x \quad , \\ \left(0 < a < 1 \ , b = 2h + 1 > 1 \ , ab > 1 + \frac{3}{2}\pi \ , h \in \mathbb{N}\right) \ , \end{gathered} \tag{2}$$

and the Mandelbrot-Weierstrass function [3]

$$f(x) = \sum_{k=-\infty}^{\infty} a^{-kH}\left(1 - \cos a^k x\right) \ , \qquad (a > 1 \ , 0 < H < 1) \ .$$

The imaginary part of (1) is the so-called Cellérier (or Weierstrass-Cellérier) function [10]

$$f(x) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} a^{-k} \sin 2\pi a^k x \ , \qquad a > 1 \ . \tag{3}$$

All the above functions show some high frequency oscillations, and their parametric plots are bounded.

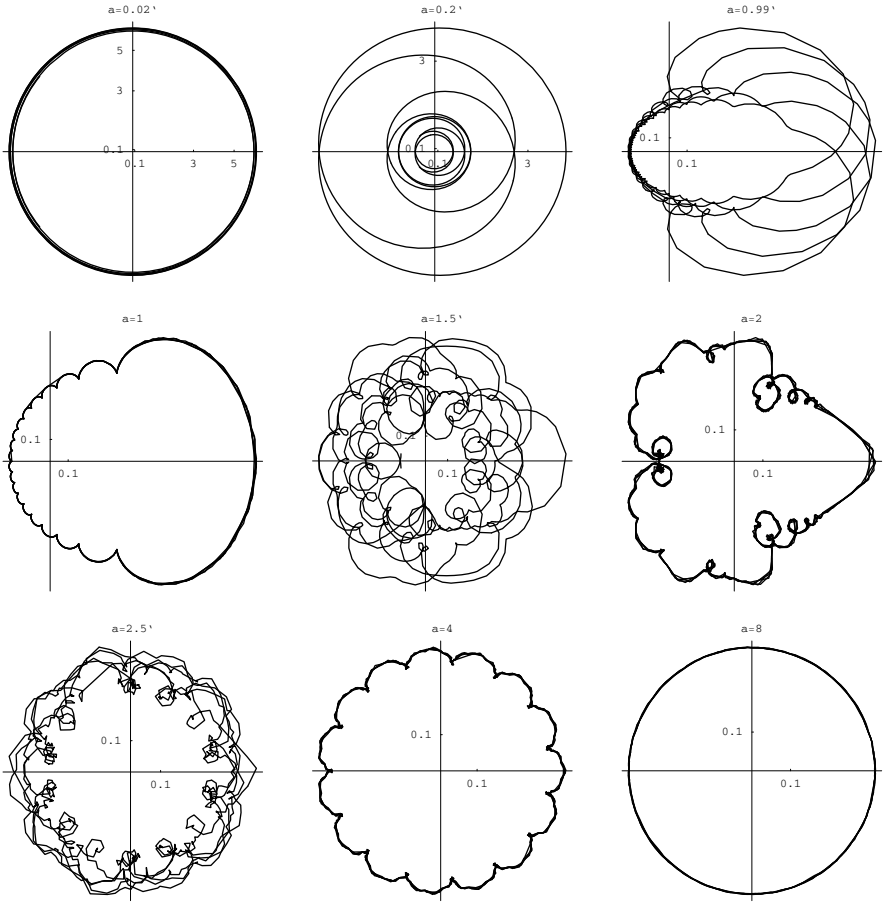


Fig. 1. Parametric plot $(\Re[f(x)],\ \Im[f(x)])$ of the generalized Riemann-Weierstrass function (11) for different values of the parameter a, with $n_{\max} = 20$, $x \in [-5\ , 5]$

2.2 Harmonic Wavelet Basis

A second example of localized, oscillating and complex value functions are the harmonic wavelets. The dilated and translated instances of the scaling $\varphi(x) = \varphi_0^0(x)$ and harmonic wavelet $\psi(x) = \psi_0^0(x)$ functions are [5,6,7,8,9,20]

$$\begin{cases} \varphi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2} \dfrac{e^{2\pi i(2^n x - k)} - 1}{2\pi i\,(2^n x - k)}\ , \\ \\ \psi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2} \dfrac{e^{4\pi i(2^n x - k)} - e^{2\pi i(2^n x - k)}}{2\pi i\,(2^n x - k)}\ , \end{cases} \tag{4}$$

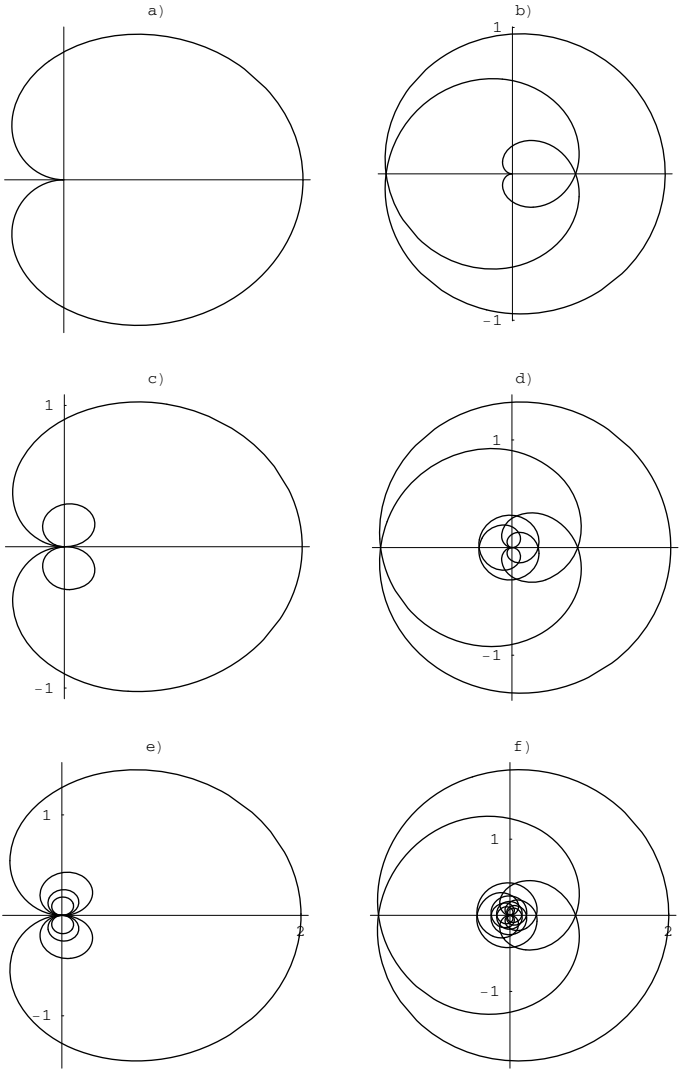


Fig. 2. Parametric plot of the scaling (left column) and harmonic wavelets: a) $\Re[\varphi_0^0(x)]$, $\Im[\varphi_0^0(x)]$; b) $\Re[\psi_0^0(x)]$, $\Im[\psi_0^0(x)]$; c) $\Re[\varphi_0^1(x)]$, $\Im[\varphi_0^1(x)]$; d) $\Re[\psi_0^1(x)]$, $\Im[\psi_0^1(x)]$; e) $\Re[\varphi_0^2(x)]$, $\Im[\varphi_0^2(x)]$; f) $\Re[\psi_0^2(x)]$, $\Im[\psi_0^2(x)]$

with

$$\varphi_k^n(x) : \mathbb{R} \to \mathbb{C} \quad , \quad \psi_k^n(x) : \mathbb{R} \to \mathbb{C} \qquad (n \in \mathbb{N}\,, k \in \mathbb{Z})\,.$$

It should be noticed that, in $\mathbb{R}^2$ these functions are bounded, self-intersecting (Fig. 2) with basins of attraction.

The space of the $L_2(\mathbb{R},\mathbb{C})$ functions is an Hilbert space. We can define the inner product, which fulfills the Parseval equality, as follows (the bar stands for complex conjugation):

$$\langle f,g\rangle \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x)\,\overline{g(x)}dx\ , \tag{5}$$

for any $f(x)\in L_2(\mathbb{R},\mathbb{C})$ and $g(x)\in L_2(\mathbb{R},\mathbb{C})$.

It can be shown [5,6,7,8,20] that harmonic wavelets are orthonormal functions, in the sense that

$$\langle \psi_k^n(x),\psi_h^m(x)\rangle = \delta^{nm}\delta_{hk}\ , \tag{6}$$

where δ^{nm} (δ_{hk}) is the Kronecker symbol.

Analogously it is

$$\begin{cases} \langle \varphi_k^n(x),\varphi_h^m(x)\rangle = \delta^{nm}\delta_{kh},\ \langle \overline{\varphi}_k^n(x),\overline{\varphi}_h^m(x)\rangle = \delta^{nm}\delta_{kh}, \\ \\ \langle \varphi_k^n(x),\overline{\varphi}_h^m(x)\rangle = 0\ , \\ \\ \left\langle \overline{\psi}_k^n(x),\overline{\psi}_h^m(x)\right\rangle = \delta^{nm}\delta_{kh},\ \left\langle \psi_k^n(x),\overline{\psi}_h^m(x)\right\rangle = 0\ , \\ \\ \left\langle \varphi_k^n(x),\overline{\psi}_h^m(x)\right\rangle = 0,\ \langle \overline{\varphi}_k^n(x),\psi_h^m(x)\rangle = 0\ \ . \end{cases} \tag{7}$$

Some simple technical computations also show that the harmonic scaling function and the harmonic wavelets fulfill the multiresolution conditions

$$\int_{-\infty}^{\infty}\varphi(x)\mathrm{d}x = 1\quad ,\int_{-\infty}^{\infty}\psi_k^n(x)\mathrm{d}x = 0\ .$$

2.3 The Riemann Zeta Function

The Riemann Zeta function is defined as

$$\zeta_R(s)\stackrel{\text{def}}{=}\sum_{n=1}^{\infty}\frac{1}{n^s}\quad ,\quad (n\in\mathbb{N}\ ,s\in\mathbb{C})\ .$$

In the following we will consider a slightly modified version of the zeta function by taking into account only the imaginary part $s=ix$, and dropping the first term (n=1) so that

$$\zeta(s)=\sum_{n=2}^{\infty}\frac{1}{n^{ix}}\quad ,\quad (n\in\mathbb{N}\ ,\ x\in\mathbb{R})\ . \tag{8}$$

The absolute value $|\zeta(ix)|$ is a decaying function only in a finite interval. In particular, by fixing the upper bound, the absolute value of the function

$$\zeta^N(x)\stackrel{\text{def}}{=}\sum_{n=2}^{N}\frac{1}{n^{ix}}\quad ,\quad (n\in\mathbb{N},\ x\in\mathbb{R}) \tag{9}$$

decays in the interval $\left[-\frac{N}{2},\ \frac{N}{2}\right]$ and oscillates, with bounded amplitude, elsewhere.

Around $x = 0$ the Taylor series expansion, of the zeta function (9) is

$$\zeta^N(x) = \sum_{k=0}^{\infty} \frac{i^k}{k!} \left(\sum_{n=2}^{N} \log^k n \right) x^k \tag{10}$$

so that the real and imaginary part are

$$\begin{cases} \Re[\zeta^N(x)] = \displaystyle\sum_{k=0}^{\infty} \frac{i^{2k}}{(2k)!} \left(\sum_{n=2}^{N} \log^{2k} n \right) x^{2k} \\ \Im[\zeta^N(x)] = \displaystyle\sum_{k=0}^{\infty} \frac{i^{2k+1}}{(2k+1)!} \left(\sum_{n=2}^{N} \log^{2k+1} n \right) x^{2k+1} \end{cases} \tag{11}$$

By a direct computation it can be seen that

$$\int_{-x}^{x} \Re(\zeta^N(x))\mathrm{d}x = \sum_{n=2}^{N} 2\frac{\sin(x \log n)}{\log n} = \sum_{n=2}^{N} \frac{i}{\log n}(n^{-ix} - n^{ix}) \quad , \quad (N \geq 2) \tag{12}$$

and

$$\int_{-ai}^{ai} \Im(\zeta^N(x))\mathrm{d}x = 0 \quad , \quad (N \geq 1,\ a \in \mathbb{R}) \ .$$

It is also

$$\int_{-a}^{a} \Re(\zeta^N(x))\Im(\zeta^N(x))\mathrm{d}x = 0 \quad , \quad (N \geq 1,\ a \in \mathbb{R})$$

so that the real and imaginary part of the function (9) are (somehow) orthogonal in the interval $[-a,\ a]$. If we change the integration bound

$$\begin{aligned} \int_{-x}^{x} \Re(\zeta^N(t))\mathrm{d}t &= \sum_{n=2}^{N} 2\frac{\sin(x \log n)}{\log n} = \sum_{n=2}^{N} \frac{i}{\log n}(n^{-ix} - n^{ix}) \\ &= \sum_{n=2}^{N} \frac{i}{\log n} \left(\frac{1}{n^{ix}} - \frac{1}{n^{-ix}} \right) \quad , \quad (N \geq 2) \end{aligned}$$

These properties of the zeta function suggest us the possibility to define a set of functions:

$$\begin{cases} \psi(x) \stackrel{\text{def}}{=} \displaystyle\sum_{n=2}^{\infty} \frac{i}{\log n}(n^{-ix} - n^{ix}) \\ \varphi(x) \stackrel{\text{def}}{=} \displaystyle\sum_{n=2}^{\infty} \frac{1}{\log n}(n^{-ix} + n^{ix}) \end{cases} \tag{13}$$

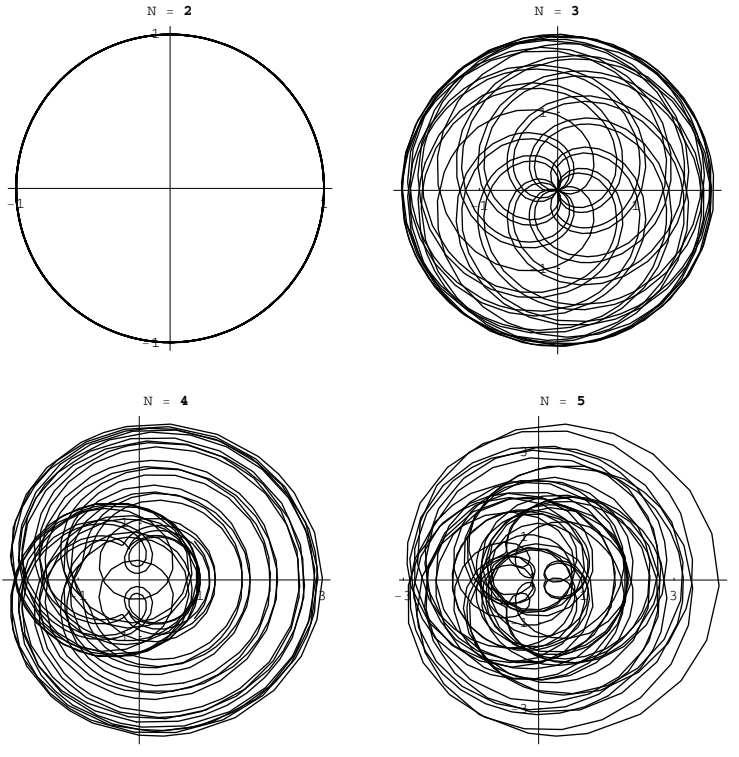


Fig. 3. *Parametric plot of* $\Re[\zeta^N(x)]$ *and* $\Im[\zeta^N(x)]$ *when* $(N = 2, \ldots, 5)$

which are orthogonal since it can be easily shown, by a direct computation, that,

$$\int_{-\infty}^{\infty} \varphi(x)\psi(x)dx = 0 \ .$$

Functions $\varphi(x)$, $\psi(x)$ are real functions but they are no-where differentiable functions, like the Weierstrass functions (1) in the sense that

$$\lim_{N\to\infty} \frac{d}{dx}\varphi^N(x) \leq \infty \quad , \quad \lim_{N\to\infty} \frac{d}{dx}\psi^N(x) \leq \infty \ . \tag{14}$$

Up to the approximation N we have

$$\begin{cases} \psi^N(x) \stackrel{\text{def}}{=} \displaystyle\sum_{n=2}^{N} \frac{i}{\log n}(n^{-ix} - n^{ix}) \\ \\ \varphi^N(x) \stackrel{\text{def}}{=} \displaystyle\sum_{n=2}^{N} \frac{1}{\log n}(n^{-ix} + n^{ix}) \end{cases} \tag{15}$$

Comment: It should be noticed that when $N = 2$ and $n \to e$ these functions coincide, apart a costant factor, with the harmonic functions $(\sin x, \cos x)$.

The Taylor series expansion, around $x = 0$, of (15) are

$$\begin{cases} \psi^N(x) \stackrel{\text{def}}{=} \displaystyle\sum_{k=0}^{\infty}(-1)^k \frac{2}{(2k+1)!}\left(\sum_{n=2}^{N}\log^{2k} n\right) x^{2k+1} \\ \varphi^N(x) = \displaystyle\sum_{k=0}^{\infty}(-1)^k \frac{2}{(2k)!}\left(\sum_{n=2}^{N}\log^{2k-1} n\right) x^{2k} \ . \end{cases} \tag{16}$$

Although the Fourier transform of $\varphi(x)$, $\psi(x)$ can't be analytically computed we can easily get the integrals of $\varphi^N(x)$, $\psi^N(x)$

$$\Phi^N(\omega, x) \stackrel{\text{def}}{=} \int_{-x}^{x} \varphi^N(t) e^{i\omega t} dt \qquad , \qquad \Psi^N(\omega, x) \stackrel{\text{def}}{=} \int_{-x}^{x} \varphi(t) e^{i\omega t} dt$$

being

$$\Phi^N(\omega, x) = -\sum_{n=2}^{N} \frac{i e^{-i\,\omega\,x}\, n^{-i\,x}}{\log n}\left[\frac{e^{2\,i\,\omega\,x} - n^{2\,i\,x}}{\omega - \log n} + \frac{n^{2\,i\,x}\,\left(e^{2\,i\,\omega\,x} - n^{-2\,i\,x}\right)}{\omega + \log n}\right]$$

and

$$\Psi^N(\omega, x) = \sum_{n=2}^{N} \frac{e^{-i\,\omega\,x}\, n^{-i\,x}}{\log n}\left[\frac{e^{2\,i\,\omega\,x} - n^{2\,i\,x}}{\omega - \log n} - \frac{n^{2\,i\,x}\,\left(e^{2\,i\,\omega\,x} - n^{-2\,i\,x}\right)}{\omega + \log n}\right]$$

It is interesting to notice that also these functions based on Riemann zeta function show some symmetries (Fig.4).

3 Family of Curves Based on Zeta Function

In the following we will consider the set of parametric curves defined as

$$\begin{cases} \xi^{i_0,i_1,i_2,i_3}(x) \ = \xi[\Re[\zeta^{i_0}(x)], \Im[\zeta^{i_1}(x)], \varphi^{i_2}(x), \psi^{i_3}(x)] \\ \eta^{j_0,j_1,j_2,j_3}(x) = \eta[\Re[\zeta^{j_0}(x)], \Im[\zeta^{j_1}(x)], \varphi^{j_2}(x), \psi^{j_3}(x)] \ , \end{cases} \tag{17}$$

by assuming that when an index is zero the corresponding function is missing, like e.g.

$$\xi^{2,0,0,3}(x) = \xi[\Re[\zeta^{2}(x)], \psi^{3}(x)] \ .$$

According to (14) these curves are no-where differentiable, however their behavior depends on the upper indexes in the sense that different values of N might give us some curves with different symmetries.

It can be seen that for values higher than 3 there is an increasing loss of symmetry. In Fig. 5 are drawn the following curves:

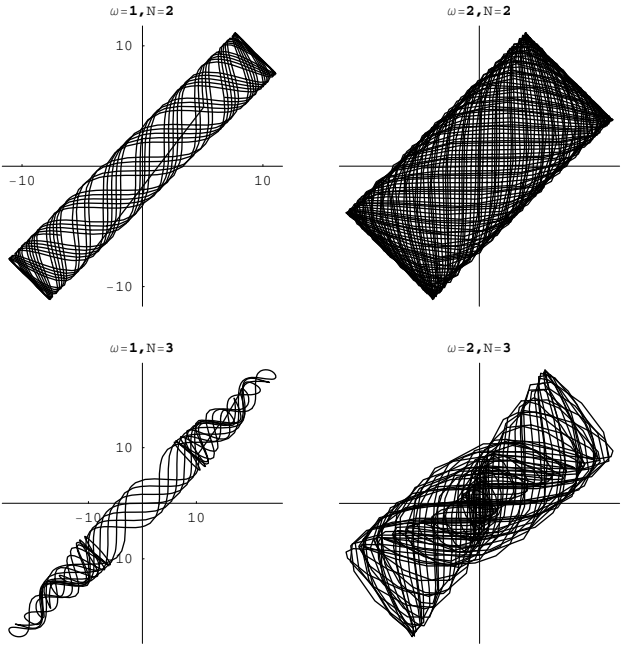


Fig. 4. *Parametric plot of* $\Re[\Phi^N(\omega,x)]$ *and* $\Im[\Psi^N(\omega,x)]$ *for some values of* N *and* ω

$$\begin{cases} \xi^{i_0,0,0,0} = \Re[\zeta^{i_0}(x)] \\ \eta^{0,0,j_2,0} = \varphi^{j_2}(x)\ , \end{cases} , \quad \begin{cases} \xi^{0,i_1,0,0} = \Im[\zeta^{i_1}(x)] \\ \eta^{0,0,j_2,0} = \varphi^{j_2}(x)\ . \end{cases} \tag{18}$$

As a second set of curves, based on zeta function, we will consider the following (see Fig. 6)

$$\begin{cases} \xi^{i_0,0,0,0} = \Re[\zeta^{i_0}(x)] \\ \eta^{0,0,0,j_3} = \psi^{j_3}(x)\ , \end{cases} , \quad \begin{cases} \xi^{0,i_1,0,0} = \Im[\zeta^{i_1}(x)] \\ \eta^{0,0,0,j_3} = \psi^{j_3}(x)\ . \end{cases} \tag{19}$$

Let us now consider the family of curves, based on the functions (15) (see Fig. 7)

$$\begin{cases} \xi^{0,0,i_2,0} = \varphi^{i_2}(x) \\ \eta^{0,0,0,j_3} = \psi^{j_3}(x)\ , \end{cases} \tag{20}$$

Comment There is a particular similarity between $\xi^{0,0,3,0}, \eta^{0,0,0,3}$ and the Michelangelo humbelical shape of Rome capital hill square (see Fig. 8)

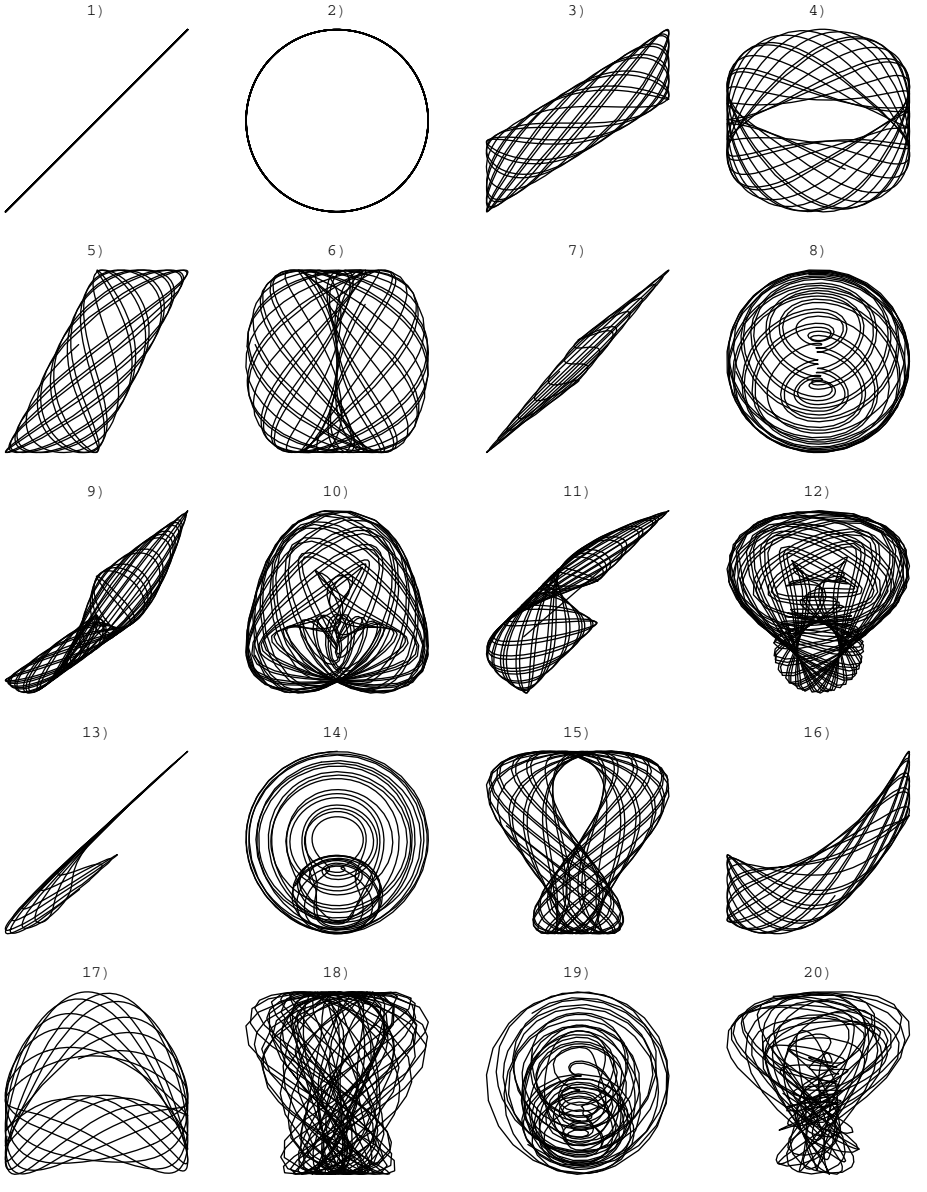


Fig. 5. *Parametric plot of* (18)*:* 1) $\xi^{2,0,0,0}, \eta^{0,0,2,0}$; 2) $\xi^{0,2,0,0}, \eta^{0,0,2,0}$; 3) $\xi^{2,0,0,0}, \eta^{0,0,3,0}$; 4) $\xi^{0,2,0,0}, \eta^{0,0,3,0}$; 5) $\xi^{3,0,0,0}, \eta^{0,0,2,0}$; 6) $\xi^{0,3,0,0}, \eta^{0,0,2,0}$; 7) $\xi^{3,0,0,0}, \eta^{0,0,3,0}$; 8) $\xi^{0,3,0,0}, \eta^{0,0,3,0}$; 9) $\xi^{3,0,0,0}, \eta^{0,0,4,0}$; 10) $\xi^{0,3,0,0}, \eta^{0,0,4,0}$; 11) $\xi^{4,0,0,0}, \eta^{0,0,3,0}$; 12) $\xi^{0,4,0,0}, \eta^{0,0,3,0}$; 13) $\xi^{4,0,0,0}, \eta^{0,0,4,0}$; 14) $\xi^{0,4,0,0}, \eta^{0,0,4,0}$; 15) $\xi^{0,4,0,0}, \eta^{0,0,2,0}$; 16) $\xi^{2,0,0,0}, \eta^{0,0,4,0}$; 17) $\xi^{0,2,0,0}, \eta^{0,0,4,0}$; 18) $\xi^{0,5,0,0}, \eta^{0,0,2,0}$; 19) $\xi^{0,5,0,0}, \eta^{0,0,5,0}$; 20) $\xi^{0,6,0,0}, \eta^{0,0,3,0}$

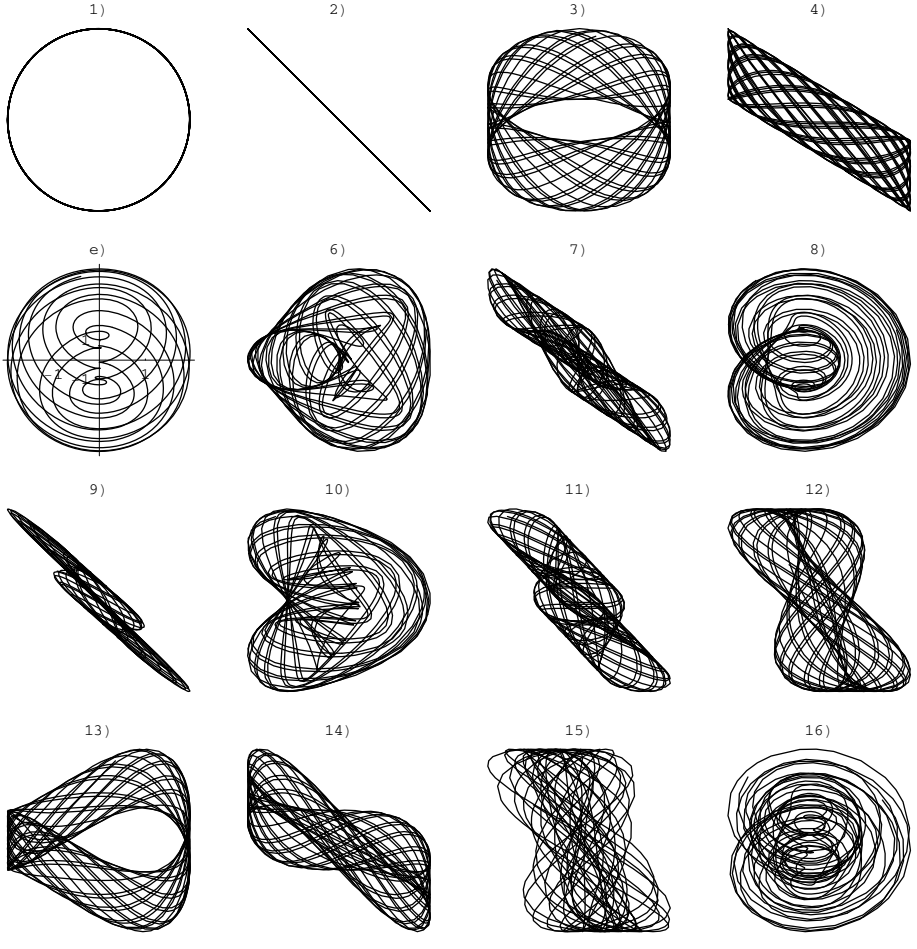


Fig. 6. *Parametric plot of* (19)*:* 1) $\xi^{2,0,0,0}, \eta^{0,0,0,2}$*;* 2) $\xi^{0,2,0,0}, \eta^{0,0,0,2}$*;* 3) $\xi^{2,0,0,0}, \eta^{0,0,0,3}$*;* 4) $\xi^{0,2,0,0}, \eta^{0,0,0,3}$*;* 5) $\xi^{3,0,0,0}, \eta^{0,0,0,3}$*;* 6) $\xi^{3,0,0,0}, \eta^{0,0,0,4}$*;* 7) $\xi^{0,3,0,0}, \eta^{0,0,0,4}$*;* 8) $\xi^{4,0,0,0}, \eta^{0,0,0,4}$*;* 9) $\xi^{0,4,0,0}, \eta^{0,0,0,4}$*;* 10) $\xi^{4,0,0,0}, \eta^{0,0,0,3}$*;* 11) $\xi^{0,4,0,0}, \eta^{0,0,0,3}$*;* 12) $\xi^{0,4,0,0}, \eta^{0,0,0,2}$*;* 13) $\xi^{2,0,0,0}, \eta^{0,0,0,4}$*;* 14) $\xi^{0,2,0,0}, \eta^{0,0,0,4}$*;* 15) $\xi^{0,5,0,0}, \eta^{0,0,0,2}$*;* 16) $\xi^{5,0,0,0}, \eta^{0,0,0,5}$

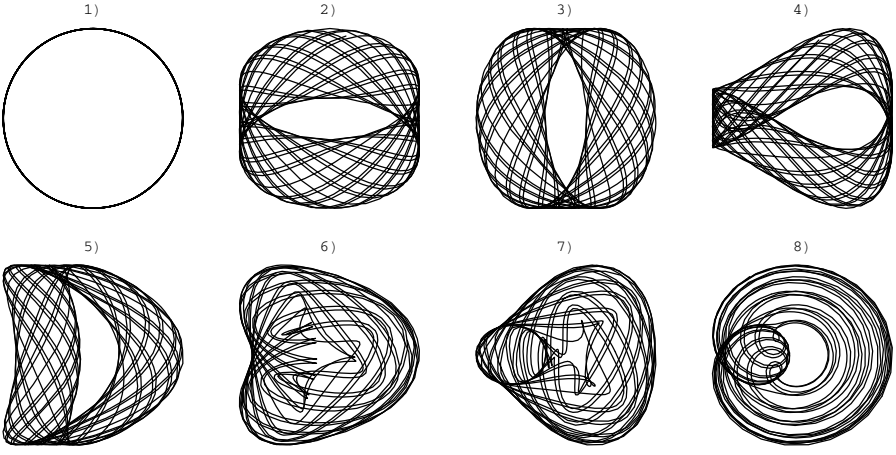


Fig. 7. *Parametric plot of* (20)*:* 1) $\xi^{0,0,2,0}, \eta^{0,0,0,2}$; 2) $\xi^{0,0,2,0}, \eta^{0,0,0,3}$; 3) $\xi^{0,0,3,0}, \eta^{0,0,0,2}$; 4) $\xi^{0,0,2,0}, \eta^{0,0,0,4}$; 5) $\xi^{0,0,4,0}, \eta^{0,0,0,2}$; 6) $\xi^{0,0,4,0}, \eta^{0,0,0,3}$; 7) $\xi^{0,0,3,0}, \eta^{0,0,0,4}$; 8) $\xi^{0,0,4,0}, \eta^{0,0,0,4}$

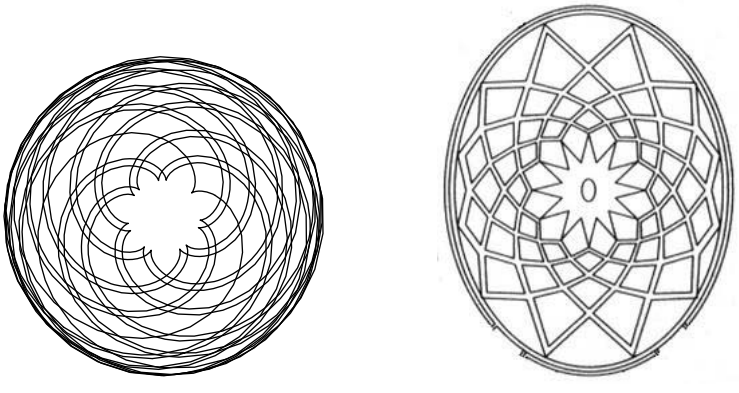


Fig. 8. *Parametric plot of* (20) $\xi^{0,0,3,0}, \eta^{0,0,0,3}$ *(left) and Michelangelo's square.*

The nonlinear combination of the functions (15) gives rise to some interesting classes of curves (Fig. 9)

$$\begin{cases} \xi^{0,0,i_2,0} &= \varphi^{i_2}(x) \\ \eta^{0,0,j_2,j_3} &= \varphi^{j_2}(x)\psi^{j_3}(x) \ , \end{cases} \quad , \quad \begin{cases} \xi^{0,0,0,i_3} &= \psi^{i_3}(x) \\ \eta^{0,0,0,j_3} &= [\psi^{j_3}(x)]^{\alpha} \ , \end{cases} \tag{21}$$

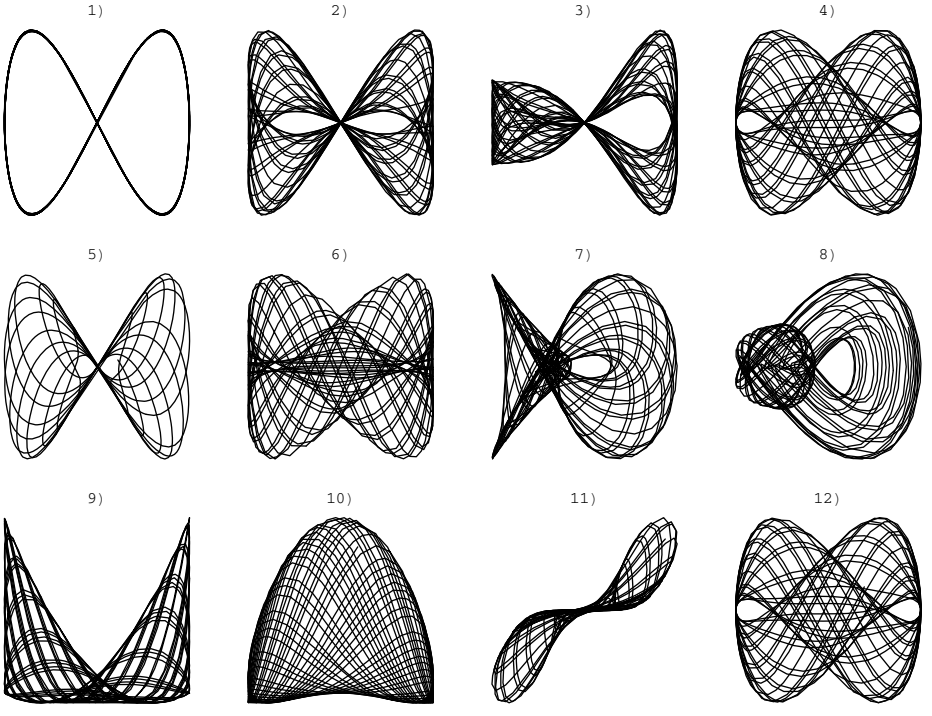


Fig. 9. *Parametric plot of* (21)*:* 1) $\xi^{0,0,2,0}, \eta^{0,0,2,2}$*;* 2) $\xi^{0,0,2,0}, \eta^{0,0,2,3}$*;* 3) $\xi^{0,0,2,0}, \eta^{0,0,2,4}$*;* 4) $\xi^{0,0,2,0}, \eta^{0,0,3,2}$*;* 5) $\xi^{0,0,3,0}, \eta^{0,0,3,2}$*;* 6) $\xi^{0,0,2,0}, \eta^{0,0,3,3}$*;* 7) $\xi^{0,0,4,0}, \eta^{0,0,3,2}$*;* 8) $\xi^{0,0,4,0}, \eta^{0,0,2,4}$

4 Conclusion

In this paper some properties and symmetries of a family of functions, based on Riemann zeta function, have been investigated.

References

1. Abry, P., Goncalves, P., Lévy-Véhel, J.: Lois d'éschelle, Fractales et ondelettes, Paris, Hermes (2002)
2. Arnéodo, A., Grasseau, G., Holschneider, M.: Wavelet Transform of Multifractals. Phys. Review Letters 61(20), 2281–2284 (1988)
3. Borgnat, P., Flandrin, P.: On the chirp decomposition of Weierstrass-Mandelbrot functions, and their time-frequency interpretation. Applied and Computational Harmonic Analysis 15, 134–146 (2003)
4. Burnol, J.-F.: On Fourier and zeta(s). Forum Math. 16(6), 789–840 (2004)
5. Cattani, C.: Harmonic Wavelets towards Solution of Nonlinear PDE. Computers and Mathematics with Applications 50(8-9), 1191–1210 (2005)

6. Cattani, C.: Wavelet Based Approach to Fractals and Fractal Signal Denoising. In: Gavrilova, M.L., Tan, C.J.K. (eds.) Transactions on Computational Science VI. LNCS, vol. 5730, pp. 143–162. Springer, Heidelberg (2009)
7. Cattani, C.: Shannon Wavelets Theory. Mathematical Problems in Engineering 2008, 24, Article ID 164808 (2008)
8. Cattani, C.: Harmonic Wavelet Approximation of Random, Fractal and High Frequency Signals. Telecommunication Systems, 207–217 (2009)
9. Cattani, C., Rushchitsky, J.J.: Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure, Series on Advances in Mathematics for Applied Sciences, 74. World Scientific, Singapore (2007)
10. Cellérier, M.C.: Note sur les principes fondamentaux de l'analyse. Darboux Bull. 14, 142–160 (1890); SIAM J. Numer. Anal., 30, 507–537 (1993)
11. Daubechies, I.: Ten Lectures on wavelets. SIAM, Philadelphia (1992)
12. Dutkay, D.E., Jorgensen, P.E.T.: Wavelets on Fractals. Rev. Mat. Iberoamericana, 22(1), 131–180 (2006)
13. Edwards, H.M.: Riemanns zeta function. In: Pure and Applied Mathematics, vol. 58. Academic Press, New York (1974)
14. Falconer, K.: Fractal Geometry. John Wiley, New York (1977)
15. Jorgensen P.E.T., Analysis and Probability, Wavelets, Signals, Fractals, Graduate Texts in Mathematics, pp. 234. Springer, Heidelberg (2006)
16. Hardin, D.P., Kessler, B., Massopust, P.: Multiresolution Analysis based on Fractal Functions. J. Approx. Theory 71, 104–120 (1992)
17. Hardin, D.P., Massopust, P.: The capacity for a class of Fractal Functions. Comm. Math. Phys. 105, 455–460 (1986)
18. Mallat S.: A Wavelet tour of signal processing. Academic Press, New York (1998)
19. Mandelbrot, B.B.: Fractals, Form, Chance and Dimension. W. H. Freeman, San Francisco (1977)
20. Newland, D.E.: Harmonic wavelet analysis. Proc. R. Soc. Lond. A 443, 203–222 (1993)
21. Srivastava, H.M., Choi, J.: Series associated with the zeta and related functions. Kluwer, Dodrecht (2001)
22. Titchmarsh, E.C.: The theory of the Riemann zeta-function, 2nd edn. The Clarendon Press, Oxford University Press, New York (1986)
23. Weierstrass, K.: Über continuirliche Functionen eines reelles Arguments, die für keinen Werth des letzteren einen Bestimmten Differentialquotienten besitzen, König. Akad. der Wissenschaften, Berlin, July 18 (1872); Reprinted in: Weierstrass, K. : Mathematische Werke II, Johnson, New York, 71–74 (1967)
24. Wornell, G.: Signal Processing with Fractals: A Wavelet-Based Approach. Prentice Hall, Englewood Cliffs (1996)

Author Index