Lecture Notes on Data Engineering and Communications Technologies 8

Leonard Barolli
Isaac Woungang
Omar Khadeer Hussain *Editors*

Advances in Intelligent Networking and Collaborative Systems

The 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017)

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 8

Series editor

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain
e-mail: fatos@cs.upc.edu

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It publishes latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.
The series has a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

More information about this series at http://www.springer.com/series/15362

Leonard Barolli · Isaac Woungang
Omar Khadeer Hussain
Editors

Advances in Intelligent Networking and Collaborative Systems

The 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017)

Springer

Editors
Leonard Barolli
Faculty of Information Engineering
Fukuoka Institute of Technology
Fukuoka
Japan

Omar Khadeer Hussain
School of Business
UNSW Australia
Canberra, ACT
Australia

Isaac Woungang
Department of Computer Science
Ryerson University
Toronto, ON
Canada

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-3-319-65635-9 ISBN 978-3-319-65636-6 (eBook)
DOI 10.1007/978-3-319-65636-6

Library of Congress Control Number: 2017951404

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Welcome Message from the INCoS-2017 Organizing Committee

Welcome to the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

INCoS is a multidisciplinary conference that covers latest advances in intelligent social networks and collaborative systems, intelligent networking systems, mobile collaborative systems, secure intelligent cloud systems, etc. Additionally, security, authentication, privacy, data trust and user trustworthiness behaviour have become crosscutting features of intelligent collaborative systems. With the fast development of the Internet, we are experiencing a shift from the traditional sharing of information and applications as the main purpose of the Web to an emergent paradigm, which locates people at the very centre of networks and exploits the value of people's connections, relations and collaboration. Social networks are also playing a major role in the dynamics and structure of intelligent Web-based networking and collaborative systems.

Virtual campuses, virtual communities and organizations strongly leverage intelligent networking and collaborative systems by a great variety of formal and informal electronic relations, such as business-to-business, peer-to-peer and many types of online collaborative learning interactions, including the emerging e-learning systems. This has resulted in entangled systems that need to be managed efficiently and in an autonomous way. In addition, latest and powerful technologies based on grid and wireless infrastructure as well as cloud computing are currently enhancing collaborative and networking applications a great deal but also facing new issues and challenges. Well-known social networks lack of knowledge management and adaptive solutions, and the information shared among peers is rather static. INCoS-2017 conference paid a special attention to security, privacy and trust in social networking, cloud computing services and storage, organization, management and autonomic computing, hidden complex networks resource management and optimization, etc.

The principal aim of this conference is to stimulate research that will lead to the creation of responsive environments for networking and, at longer term, the development of adaptive, secure, mobile and intuitive intelligent systems for collaborative work and learning. INCoS-2017 addressed a large number of themes and focuses on the following research tracks:

- Data Mining, Machine Learning and Collective Intelligence
- Fuzzy Systems and Knowledge Management
- Grid And P2P Distributed Infrastructure for Intelligent Networking and Collaborative Systems
- Nature's Inspired Parallel Collaborative Systems
- Security, Organization, Management and Autonomic Computing for Intelligent Networking and Collaborative Systems
- Wireless and Sensor Systems for Intelligent Networking and Collaborative Systems
- Service-based Systems for Enterprise Activities Planning and Management
- Big Data Analytics for Learning, Networking and Collaborative Systems
- Cloud Computing: Services, Storage, Security and Privacy
- Intelligent Collaborative Systems for Work and Learning, Virtual Organization and Campuses
- Social Networking and Collaborative Systems
- Intelligent Sensor Networks for Environmental Monitoring

As in all previous editions, INCoS-2017 counted on with the support and collaboration of a large and internationally recognized TPC covering all main themes of the conference. After a careful peer review process, in which each paper was reviewed by at least two referees (in average three reviews per submission), there were accepted 35 regular papers (30% acceptance rate). Additionally, four workshops were organized in conjunction with the conference, in which 24 workshop papers were accepted.

The successful organization of the conference is achieved thanks to the great collaboration and hard work of many people and conference supporters. First and foremost, we would like to thank all the authors for their continued support to the conference by submitting their research work to the conference, for their presentations and discussions during the conference days. We would like to thank TPC members and external reviewers for their work by carefully evaluating the submissions and providing constructive feedback to authors. We would like to thank the track chairs for their work on setting up the tracks and the respective TPCs and also for actively promoting the conference and their tracks. We would like to appreciate the work of PC co-chairs and workshops co-chairs for the successful organization of workshops in conjunction with main conference.

We would like to acknowledge the excellent work and support by the International Advisory Committee. Our gratitude and acknowledgement for the conference keynotes, Dr. Ann Cavoukian, Ryerson University, Canada, and Prof. Shahrokh Valaee, University of Toronto, Canada, for their interesting and inspiring keynote speeches.

We greatly appreciate the support by Web Administrators Shinji Sakamoto, Yi Liu, Donald Elmazi and Miralda Cuka, Fukuoka Institute of Technology (FIT), Japan, and Finance Chair Makoto Ikeda, Fukuoka Institute of Technology (FIT), Japan.

We would like to give special thanks to the members of the local organizing committee from Ryerson University, Toronto, Canada, for excellent arrangements for the conference.

We are very grateful to Springer as well as several academic institutions for their technical sponsorship, endorsement and assistance.

Finally, we hope that you will find these proceedings to be a valuable resource in your professional, research and educational activities!

We wish you will enjoy your stay in Toronto, Canada, and look forward to seeing you again in forthcoming edition of INCoS conference!

General Co-chairs

Leonard Barolli	Fukuoka Institute of Technology, Japan
Isaac Woungang	Ryerson University, Canada

Message from the INCoS-2017 Workshops Chairs

Welcome to the Workshops of the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

In this edition of the conference, there are held four workshops, which complemented the INCoS main themes with specific themes and research issues and challenges, as follows:

1. The 9th International Workshop on Information Network Design (WIND 2017)
2. The 5th International Workshop on Frontiers in Intelligent Networking and Collaborative Systems (FINCoS-2017)
3. The 3rd International Workshop for Collaborative e-Business Systems (IWCBS-2017)
4. The 3rd International Workshop on Theory, Algorithms and Applications of Big Data Science (BDS-2017)

We would like to thank the workshop organizers for their great efforts and hard work in proposing the workshop, selecting the papers, the interesting programmes and for the arrangements of the workshop during the conference days. We are grateful to the INCoS-2017 Conference Chairs for inviting us to be the workshops co-chairs of the conference.

We hope you will enjoy the workshops programmes and discussions during the conference days and enjoy your stay in Toronto, Canada!

INCoS-2017 Workshops Co-chairs

Xu An Wang	Engineering University of CAPF, P.R. China
Kin Fun Li	University of Victoria, Canada

INCoS-2017 Organizing Committee

Honorary Chairs

Eric Harley	Ryerson University, Canada
Makoto Takizawa	Hosei University, Japan

General Co-chairs

Leonard Barolli	Fukuoka Institute of Technology, Japan
Isaac Woungang	Ryerson University, Canada

Program Co-chairs

Hiroyoshi Miwa	Kwansei Gakuin University, Japan
Omar Hussain	UNSW Canberra, Australia
Kin Fun Li	University of Victoria, Canada

International Advisory Committee

Santi Caballé	Open University of Catalonia, Spain
Ana Azevedo	Polytechnic Institute of Porto/ISCAP, Portugal
Anabela Mesquita	Polytechnic Institute of Porto/ISCAP, Portugal
Vincenzo Loia	University of Salerno, Italy
Masato Tsuru	Kyushu Institute of Technology, Japan
Albert Zomaya	University of Sydney, Australia
Christine Strauss	University of Vienna, Austria

Workshops Co-chairs

Xu An Wang	Engineering University of CAPF, P.R. China
Kin Fun Li	University of Victoria, Canada

Special Issues Co-chairs

Flora Amato	University of Naples, Italy
Farookh Hussain	Univ. of Technology Sydney, Australia

International Liaison Co-chairs

Mehrdad Torandazian	Ryerson University, Canada
Natalia Kryvinska	University of Vienna, Austria
Pavel Kromer	Technical Univ. of Ostrava, Czech Republic

Award Co-chairs

Glaucio Carvalho	Ryerson University, Canada
Marek Ogiela	AGH Univ. of Science and Technology, Poland
Fang-Yie Leu	Tunghai University, Taiwan

Web Administrators

Shinji Sakamoto	Fukuoka Institute of Technology, Japan
Donald Elmazi	Fukuoka Institute of Technology, Japan
Yi Liu	Fukuoka Institute of Technology, Japan
Miralda Cuka	Fukuoka Institute of Technology, Japan

Local Arrangement Co-chairs

Mehrdad Tirandazian	Ryerson University, Canada
Glaucio Carvalho	Ryerson University, Canada

Finance Chair

Makoto Ikeda	Fukuoka Institute of Technology, Japan

Steering Committee

Leonard Barolli	Fukuoka Institute of Technology, Japan

Track 1: Data Mining, Machine Learning and Collective Intelligence

Track Co-chairs

Richi Nayak	Queensland University of Technology, Australia
Carson K Leung	University of Manitoba, Canada

TPC Members

Carson Leung	University of Manitoba, Canada
Alfredo Cuzzocrea	University of Trieste, Italy
Fan Jiang	University of Manitoba, Canada
Wookey Lee	Inha University, Korea
Adam Pazdor	University of Manitoba, Canada
Syed Tanbeer	University of Manitoba, Canada
Yvegen Biletskiy	University of New Brunswick, Canada

Track 2: Fuzzy Systems and Knowledge Management

Track Co-chairs

Marek Ogiela	AGH University of Science and Technology, Poland
Morteza Saberi	UNSW Canberra, Australia
Chang Choi	Chosun University, South Korea

TPC Members

Farookh Khadeer Hussain	University of Technology Sydney, Australia
Hae-Duck Joshua Jeong	Korean Bible University, Korea
Hoon Ko	Sungkyunkwan University, Korea
Natalia Krzyworzeka	AGH University of Science and Technology, Poland
Libor Mesicek	J.E. Purkinje University, Czech Republic
Lidia Ogiela	AGH University of Science and Technology, Poland
Su Xin	Hohai University, China
Ali Azadeh	University of Tehran, Iran
Jin Hee Yoon	Sejong University, South Korea
Hamed Shakouri	University of Tehran, Iran
Jee-Hyong Lee	Sungkyunkwan University, South Korea
Jung Sik Jeong	Mokpo National Maritime University, South Korea
Junho Choi	Chosun University, South Korea

Track 3: Grid and P2P Distributed Infrastructure for Intelligent Networking and Collaborative Systems

Track Co-chairs

Florin Pop	University Politehnica of Bucharest, Romania
Ciprian Dobre	University Politehnica of Bucharest, Romania

TPC Members

Jordi Mongay Batalla	National Institute of Telecommunications, Poland
Nik Bessis	Edge Hill University, UK
Aniello Castiglione	Dipartimento di Informatica, University of Salerno, Italy
Naveen Chilamkurti	La Trobe University, Australia
Radu-Ioan Ciobanu	University Politehnica of Bucharest, Romania
Alexandru Costan	IRISA/INSA Rennes, France
Vladimir-Ioan Ciobanuretu	University Politehnica of Timisoara, Romania
Valentin Cristea	University Politehnica of Bucharest, Romania
Marc Frincu	West University of Timisoara, Romania
Rossitza Ivanova Goleva	Technical University of Sofia, Bulgaria
Dorian Gorgan	Technical University of Cluj-Napoca, Romania
Mauro Iacono	Seconda Universita' degli Studi di Napoli, Italy
George Mastorakis	Technological Educational Institute of Crete, Greece
Constandinos X. Mavromoustakis	University of Nicosia, Cyprus
Gabriel Neagu	National Institute for Research and Development in Informatics, Romania
Rodica Potolea	Technical University of Cluj-Napoca, Romania
Radu Prodan	University of Innsbruck, Austria
Ioan Salomie	Technical University of Cluj-Napoca, Romania
George Suciu	BEIA International, Romania
Nicolae Tapus	University Politehnica of Bucharest, Romania
Sergio L. Toral Marín	University of Seville, Spain
Radu Tudoran	European Research Center, Huawei Technologies Duesseldorf GmbH, Germany
Lucian Vintan	Lucian Blaga University, Romania
Mohammad Younas	Oxford Brookes University, UK

Track 4: Nature's Inspired Parallel Collaborative Systems

Track Co-chairs

Enrique Alba	University of Malaga, Spain
Gabriel Luque	University of Malaga, Spain

TPC Members

Francisco Luna	University of Malaga, Spain
Sergio Nesmachnow	University La Republica, Uruguay
Nouredine Melab	University of Lille 1, France
Julio Ortega	University of Granada, Spain
Domingo Giménez	University of Murcia, Spain
Gregoire Danoy	University of Luxembourg, Luxembourg
Carolina Salto	University of La Pampa, Argentina
Stefka Fidanova	IICT-BAS, Bulgaria
Michael Affenzeller	Upper Austria University, Austria
Hernan Aguirre	Shinshu University, Japan
Francisco Chicano	University of Malaga, Spain
Javid Tahery	Karlstad University, Switzerland
Enrique Domínguez	University of Malaga, Spain
Guillermo Leguizamón	Universidad Nacional de San Luis, Argentina
Konstantinos Parsopoulos	University of Ioannina, Germany
Carlos Segura	CIMAT, Mexico
Eduardo Segredo	Edinburgh Napier University, UK
Javier Arellano	University of Malaga, Spain

Track 5: Security, Organization, Management and Autonomic Computing for Intelligent Networking and Collaborative Systems

Track Co-chairs

Mingwu Zhang	Hubei University of Technology, China
Sudhakar Ganti	University of Victoria, Canada

TPC Members

MarekOgiela	AGH University of Science and Technology, Poland
Chang Choi	Chosun University, Korea

Hsing-Chung (Jack) Chen	Asia University, Taiwan
Been-Chian Chien	National University of Tainan, Taiwan
Junho Choi	Chosun University, Korea
Hae-Duck Joshua Jeongeong	Korean Bible University, Korea

Track 6: Software Engineering, Semantics and Ontologies for Intelligent Networking and Collaborative Systems

Track Co-chairs

Enrahim Bagheri	Ryerson University, Canada
Asil Gill	University of Technology, Sydney, Australia

TPC Members

Tsutomu Kinoshita	Fukui University of Technology, Japan
Kouji Kozaki	Osaka University, Japan
Hiroyoshi Miwa	Kwansei Gakuin University, Japan
Burin Rujjanapan	Nation University, Taiwan
Hiroshi Kanasugi	Tokyo University, Japan
Takayuki Shimotomai	Advanced Simulation Technology of Mechanics R&D, Japan
Jinattaporn Khumsri	Fukui University of Technology, Japan
Rene Witte	Concordia University, Canada
Amal Zouaq	University of Ottawa, Canada
Jelena Jovanovic	University of Belgrade, Serbia
Zeinab Noorian	Ryerson University, Czech Republic
Faezeh Ensan	Ferdowsi University of Mashhad, Iran
Alireza Vazifedoost	Sun Life Financial, Canada
Morteza Mashayekhi	Royal Bank of Canada, Canada
Atif Qureshi	University of Technology Sydney, Australia
Abdul Babar	University of Technology Sydney, Australia
Georges Bou	University of Technology Sydney, Australia
Ghassan Beydoun	University of Technology Sydney, Australia
Mahmood Niazi	KFUMP, Saudi Arabia
Yehia Alzoubi	University of Technology Sydney, Australia
Yvegen Biletskiy	University of New Brunswick, Canada
Akihiro Fujihara	Fukui University of Technology, Japan

Track 7: Wireless and Sensor Systems for Intelligent Networking and Collaborative Systems

Track Co-chairs

Noriki Uchida	Fukuoka Institute of Technology, Japan
Dhananjay Singh	Hankuk (Korea) University of Foreign Studies, Korea

TPC Members

Dhananjay Singh	HUFS, Korea
Shirshu Varma	IIIT-Allahabad, India
B. Balaji Naik	NIT-Sikkim, India
Sayed Chhattan Shah	HUFS, Korea
Madhusudan Singh	Yonsei University, Korea
Irish Singh	Ajou University, Korea
Gaurav Tripathi	Bharat Electronics Limited, India
Hiroyoshi Miwa	Kwansei Gakuin University, Japan

Track 8: Service-Based Systems for Enterprise Activities Planning and Management

Track Co-chairs

Kryvinska Natalia	University of Vienna, Austria
Michal Gregus	Comenius University in Bratislava, Slovakia

TPC Members

Christine Bauer	University of Vienna, Austria
Maria Bohdalova	Comenius University in Bratislava, Slovakia
Alfredo Cuzzocrea	University of Calabria, Italy
Ivan Demydov	Lviv Polytechnic National University, Ukraine
Corinna Engelhardt-Nowitzki	University of Applied Sciences, Austria
Dan Johansson	Umea University, Sweden
Jozef Juhar	Technical University of Košice, Slovakia
Nikolay Kazantsev	National Research University Higher School of Economics, Russia
Wim Laurier	Facultés Universitaires Saint-Louis, Brussels, Belgium
Chris Ling	Monash University, Australia
Viliam Malcher	Comenius University in Bratislava, Slovakia

Michele Melchiori	Universita' degli Studi di Brescia, Italy
Nor Shahniza Kamal Bashah	Universiti Teknologi MARA, Malaysia
Eric Pardede	La Trobe University, Australia
Aneta Poniszewska-Maranda	Lodz University of Technology, Poland
Sushil	Indian Institute of Technology, India
Do van Thanh	Telenor and Norwegian University of Science and Technology, Norway
Sergey Zykov	National Research University Higher School of Economics, Russia

Track 9: Next Generation Secure Network Protocols and Components

Track Co-chairs

Xu An Wang	Engineering University of CAPF, China
Ben Soh	La Trobe University, Australia

TPC Members

Fushan Wei	The PLA Information Engineering University, China
He-Xu Nangjing	University of Posts and Telecommunications, China
Yining Liu	Guilin University of Electronic Technology, China
Yuechuan Wei	Engineering Unviersity of CAPF, China
Weiwei Kong	Xi'an University of Posts and Telecommunications, China
Dianhua Tang	CETC 30, China
Hui Tian	Huaqiao University, China

Track 10: Big Data Analytics for Learning, Networking and Collaborative Systems

Track Co-chairs

Santi Caballe	Universitat Oberta de Catalunya, Spain
Jordi Conesa	Universitat Oberta de Catalunya, Spain

TPC Members

Soumya Barnejee	Institut National des Sciences Appliquées, France
David Bañeres	Open University of Catalonia, Spain
Nicola Capuano	University of Salerno, Italy
Nestor Mora	Open University of Catalonia, Spain
Jorge Moneo	University of San Jorge, Spain
David Gañán	Open University of Catalonia, Spain
Isabel Guitart	Open University of Catalonia, Spain
Elis Kulla	Okayama University of Science, Japan
Evjola Spaho	Polytechnic University of Tirana, Albania
Florin Pop	University Politehnica of Bucharest, Romania
Kin Fun Li	University of Victoria, Canada
Miguel Bote	University of Valladolid, Spain
Pedro Muñoz	University of Carlos III, Spain

Track 11: Cloud Computing: Services, Storage, Security and Privacy

Track Co-chairs

Farookh Hussain	University of Technology Sydney, Australia
Samee U Khan	North Dakota State University, USA

TPC Members

Ejaz Ahmed	National Institute of Standards and Technology, USA
Asad Malik	National University of Science and Technology, Pakistan
Usman Shahid	Comsats Institute of Information Technology, Pakistan
Assad Abbas	North Dakota State University, UK
Nikolaos Tziritas	Chinese Academy of Sciences, China
Osman Khalid	Comsats Institute of Information Technology, Pakistan
Kashif Bilal	Qatar University, Qatar
Javid Taheryaheri	Karlstad University, Switzerland
Sai Rehman	Comsats Institute of Information Technology, Pakistan
Inayat Babar	University of Engineering and Technology, Pakistan
Thanasis Loukopoulos	Technological Educational Institute of Athens, Greece
Mazhar Ali	Comsats Institute of Information Technology, Pakistan
Tariq Umer	Comsats Institute of Information Technology, Pakistan

Track 12: Intelligent Collaborative Systems for Work and Learning, Virtual Organization and Campuses

Track Co-chairs

Hitoshi Sasaki	Takushoku University, Japan
Atanasi Daradoumis Haralabus	Open University of Catalonia, Spain

TPC Members

Luis Alberto Casillas	University of Guadalajara, Mexico
Nestor Mora	University of Cadiz, Spain
Michalis Feidakis	University of Aegean, Greece
Sandra Isabel Enciso	"Fundación Universitaria Juan N. Corpas", Colombia
Nicola Capuano	University of Salerno, Italy
Rafael Del Hoyo	Technological Center of Aragon, Spain
George Caridakis	University of Aegean, Greece
Kazunori Mizuno	Takushoku University, Japan
Satoshi Ono	Kagoshima University, Japan
Yoshiro Imai	Kagawa University, Japan
Takashi Mitsuishi	Tohoku University, Japan
Hiroyuki Mitsuhara	Tokushima University, Japan

Track 13: Social Networking and Collaborative Systems

Track Co-chairs

Nicola Capuano	University of Salerno, Italy
Cherie Ding	Ryerson University, Canada

TPC Members

Santi Caballé	Open University of Catalonia, Spain
Thanasis Daradoumis	University of the Aegean, Greece
Angelo Gaeta	University of Salerno, Italy
Christian Guetl	Graz University of Technology, Austria
Miltiadis Lytras	American College of Greece
Agathe Merceron	Beuth University of Applied Sciences Berlin, Germany
Francis Palma	Screaming Power, Canada
Krassen Stefanov	Sofia University "St. Kliment Ohridski", Bulgaria
Daniele Toti	Roma Tre University, Italy

Jian Wang	Wuhan University, China
Jing Xiao	South China Normal University, China
Jian Yu	Auckland University of Technology, Australia

Track 14: Intelligent Sensor Networks for Environmental Monitoring

Track Co-chairs

Petr Musilek	University of Alberta, Canada
Arturo Sanchez-Azofeifa	University of Alberta, Canada

TPC Members

Tim Brown	Australian National University, Australia
Mario Marcos do Espirito Santo	Universidad Stadual de Montes Claros, Brazil
Jana Heckenbergerova	University Pardubice, Czech Republic
Zdenek Matej	Masaryk University, Czech Republic
Michal Musilek	University Hradec Kralove, Czech Republic
Michal Prauzek	VSB-TU Ostrava, Czech Republic
Vaclav Prenosil	Masaryk University, Czech Republic
Alvin C. Valera	Singapore Management University, Singapore
Nasem Badreldin	University of Manitoba, Canada
Emil Pelikan	Academy of Sciences, China
Joanne Nightingale	National Physical Laboratory, UK
Tomas Barton	University of Alberta, Canada

INCoS-2017 Reviewers

Aleksy Markus
Amato Flora
Barolli Admir
Barolli Leonard
Caballé Santi
Chang Choi
Capuano Nicola
Chen Xiaofeng
Chien Been-Chian
Cui Baojiang
Thanasis Daradoumis
Doug Young Suh
Enokido Tomoya
Fenza Giuseppe
Ficco Massimo
Fiore Ugo
Fujihara Akihiro
Fun Li Kin
Gotoh Yusuke
He Debiao

Hsing-Chung Chen
Hussain Farookh
Hussain Omar
Ikeda Makoto
Joshua Hae-Duck
Kikuchi Hiroaki
Kolici Vladi
Koyama Akio
Kromer Pavel
Kryvinska Natalia
Kulla Elis
Loia Vincenzo
Ma Kun
Matsuo Keita
Messina Fabrizio
Miwa Hiroyoshi
Kryvinska Natalia
Leu Fang-Yie
Juggapong Natwichai
Nadeem Javaid
Nishino Hiroaki
Ogiela Lidia
Ogiela Marek
Palmieri Francesco
Pardede Eric
Rahayu Wenny
Rawat Danda
Rein Wen
Saberi Morteza
Shibata Yoshitaka
Snasel Vaclav
Spaho Evjola
Suganuma Takuo
Sugita Kaoru
Takizawa Makoto
Terzo Olivier
Toral Sergio
Tsuru Masato
Uchida Noriki
Venticinque Salvatore
Wang Xu An
Younas Mohammad
Zhang Mingwu
Zomaya Albert

Welcome Message from WIND-2017 Workshop Organizers

Welcome to the 9th International Workshop on Information Network Design (WIND-2017), which is held in conjunction with the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

Nowadays, the Internet is playing a role of social and economical infrastructure and is expected to support not only comfortable communication and information dissemination but also any kind of intelligent and collaborative activities in a dependable manner. However, the explosive growth of its usage with diversifying the communication technologies and the service applications makes it difficult to manage efficient sharing of the Internet. In addition, an inconsistency between Internet technologies and the human society forces a complex and unpredictable tension among end-users, applications and ISPs (Internet Service Providers).

It is thought, therefore, that the Internet is approaching a turning point, and there might be the need for rethinking and redesigning the entire system composed of the human society, nature and the Internet. To solve the problems across multiple layers on a large-scale and complex system and to design the entire system of systems towards future information networks for human/social orchestration, a new tide of multi-perspective and multidisciplinary research is essential. It will involve not only the network engineering (network routing, mobile and wireless networks, network measurement and management, high-speed networks, etc.) and the networked applications (robotics, distributed computing, human computer interactions, Kansei information processing, etc.), but also the network science (providing new tools to understand and control the huge-scale complex systems based on theories, e.g. graph theory, game theory, information theory, learning theory, statistical physics) and the social science (enabling safe, secure, and human-centric application principles and business models).

The Information Network Design Workshop aims at exploring ongoing efforts in the theory and application on a wide variety of research fields related to the design of information networks and resource sharing in the networks. The workshop provides an opportunity for academic/industry researchers and professionals to share, exchange and review recent advances on information network design research. The workshop seeks original contribution describing recent modelling, analysis, and experiment on network design research related with:

- Large-scale and/or complex networks
- Cross-layered networks
- Overlay and/or P2P networks
- Sensor and/or mobile ad hoc networks
- Delay/disruption tolerant networks
- Social networks
- Applications on networks
- Fundamental theories for network design

Many people contributed to the success of WIND-2017. We would like to thank the organizing committee of INCoS-2017 International Conference for giving us the opportunity to organize the workshop. We also like to thank our program committee members and referees and of course all authors of the workshop for submitting their research works and for their participation.

We wish all participants and contributors to spend an event with high research impact, interesting discussions, exchange of research ideas, to pave future research cooperations.

WIND-2017 Workshop Co-chairs

Masaki Aida	Tokyo Metropolitan University, Japan
Mario Koeppen	Kyushu Institute of Technology, Japan
Hiroyoshi Miwa	Kwansei Gakuin University, Japan
Masato Tsuru	Kyushu Institute of Technology, Japan
Masato Uchida	Waseda University, Japan

WIND-2017 Organizing Committee

WIND-2017 Workshop Co-chairs

Masaki Aida	Tokyo Metropolitan University, Japan
Mario Koeppen	Kyushu Institute of Technology, Japan
Hiroyoshi Miwa	Kwansei Gakuin University, Japan
Masato Tsuru	Kyushu Institute of Technology, Japan
Masato Uchida	Waseda University, Japan

Program Committee

Masaki Aida	Tokyo Metropolitan University, Japan
Yoshiaki Hori	Saga University, Japan
Hideaki Iiduka	Meiji University, Japan
Mario Koeppen	Kyushu Institute of Technology, Japan
Kenichi Kourai	Kyushu Institute of Technology, Japan
Hiroyoshi Miwa	Kwansei Gakuin University, Japan
Kei Ohnishi	Kyushu Institute of Technology, Japan
Masahiro Sasabe	Nara Institute of Science and Technology, Japan
Kazuya Tsukamoto	Kyushu Institute of Technology, Japan
Masato Tsuru	Kyushu Institute of Technology, Japan
Masato Uchida	Chiba Institute of Technology, Japan
Neng-Fa Zhou	The City University of New York, USA

Welcome Message from FINCoS-2017 Workshop Organizers

Welcome to the 5th International Workshop on Frontiers in Intelligent Networking and Collaborative Systems (FINCoS-2017), which is held in conjunction with the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

The FINCoS-2017 covers the latest advances in the interdisciplinary fields of intelligent networking, social networking, collaborative systems, cloud-based systems, business intelligence, etc., which lead to gain competitive advantages in business and academia scenarios. The ultimate aim is to stimulate research that will lead to the creation of responsive environments for networking and, at longer term, the development of adaptive, secure, mobile and intuitive intelligent systems for collaborative work and learning.

Industry and academic researchers, professionals and practitioners are invited to exchange their experiences and present their ideas in this field. Specifically, the scope of FINCoS-2017 comprises research work and findings on intelligent networking, distributed infrastructures, security and privacy and collaborative mobile networks & applications. We would like to thank all authors of the workshop for submitting their research works and their participation. We would like to express our appreciation to the reviewers for their timely review and constructive feedback to authors. Finally, we would like to thank the Local Arrangement Chairs for facilitating the workshop organization.

We are looking forward to meet you again in the forthcoming editions of the workshop.

FINCoS-2017 Workshop Organizer

Leonard Barolli	Fukuoka Institute of Technology, Japan

FINCoS-2017 Organizing Committee

Workshop Organizer

Leonard Barolli	Fukuoka Institute of Technology, Japan

Program Committee

Santi Caballé	Open University of Catalonia
Xiaofeng Chen	Xidian University, China
Makoto Ikeda	Fukuoka Institute of Technology, Japan
Kin Fun Li	University of Victoria, Canada
Shengli Liu	Shanghai Jiaotong University, China
Janusz Kacpryzk	Polish Academy of Science, Poland
Hiroaki Nishino	University of Oita, Japan
Makoto Takizawa	Hosei University, Japan
David Taniar	Monash University, Australia
Xu An Wang	CAPF Engineering University, P.R. China

Welcome Message from IWCBS-2017 International Workshop Organizers

Welcome to the 3rd International Workshop for Collaborative e-Business Systems (IWCBS-2017), which is held in conjunction with the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

The rapid expansion of business relationships and processes involved led to the emerging standards and infrastructure for business collaborations. Business large or small can no longer survive alone. The efficient and effective links with the business partners and consumers become critical. Overall, the collaborations occur between the communities of buyers, i.e. service consumers and sellers, i.e. service providers.

As much of the competition occurs between services providers and services consumers along the e-business value chains, the main theme of IWCBS is on collaborative e-business systems through aspects of business-IT alignment, business process integration, mobility, technology and tools, platforms and architectures, and applications.

The workshop aims to address the resources planning, modelling, coordination and integration in order to develop long-term sustainable and beneficial business relationships among all the partners and consumers along the value chains. Development of well-cooperated and coordinated e-business environment is crucial. Information technology has significant roles in supporting more competitive collaborative and integrated e-business systems. For business stakeholders, the long-term sustainability and efficiency are to be increasingly important. Indeed, to appropriately address the balance between the community of buyers and sellers through collaboration and support is becoming urgent. The technology trend of supply chain management and logistics is heading towards all aspects of the integration, coordination and intelligent use of the network-based resources. In practical deployment of the solutions, mobility and hand-held devices are to be involved.

We are looking forward to meet you again in the forthcoming editions of the workshop.

IWCBS-2017 Workshop Organizer

Leonard Barolli	Fukuoka Institute of Technology, Japan

IWCBS-2017 Workshop PC Chair

Natalia Kryvinksa	University of Vienna, Austria

IWCBS-2017 Organizing Committee

Workshop Organizer

Leonard Barolli	Fukuoka Institute of Technology, Japan

Workshop PC Chair

Natalia Kryvinksa	University of Vienna, Austria

Workshop PC Members

Wei Dai	University of Victoria, Australia
Serguei Dobrinevski	Hypersoft, Germany
Ciprian Dobre	University Politehnica of Bucharest, Romania
Antonio Gentile	University of Palermo, Italy
Makoto Ikeda	Fukuoka Institute of Technology, Japan
Kin Fun Li	University of Victoria, Canada
Hiroaki Nishino	University of Oita, Japan
Tomoya Enokido	Rissho University, Japan
Makoto Takizawa	Hosei University, Japan

Welcome Message from BDS-2017 Workshop Organizer

Welcome to the 3rd International Workshop on Theory, Algorithms and Applications of Big Data Science, which is held in conjunction with the 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017), which will be held from 24 to 26 August 2017, at Ryerson University, Toronto, Canada.

Diverse multidisciplinary approaches are being continuously developed and advanced to address the challenges that Big Data research raises. In particular, the current academic and professional environments are working to produce algorithms, theoretical advance in Big Data science, to enable the full utilization of its potential and better applications.

The proposed workshop focuses on the dissemination of original contributions to discuss and explore theoretical concepts, principles, tools, techniques and deployment models in the context of Big Data. Via the contribution of both academics and industry practitioners, the current approaches for the acquisition, interpretation and assessment of relevant information will be addressed to advance the state-of-the-art Big Data technology.

The workshop covers the following topics:

- Contributions should focus on (but not limited to) the following topics:
- Statistical and dynamical properties of Big Data;
- Applications of machine learning for information extraction;
- Hadoop and Big Data;
- Data and text mining techniques for Big Data;
- Novel algorithms in classification, regression, clustering and analysis;
- Distributed systems and cloud computing for Big Data;
- Big Data applications;
- Theory, applications and mining of networks associated with Big Data;
- Large-scale network data analysis;
- Data reduction, feature selection and transformation algorithms;
- Data visualization;
- Distributed data analysis platforms;

- Scalable solutions for pattern recognition;
- Stream and real-time processing of Big Data;
- Information quality within Big Data;
- Threat detection in Big Data.

We would like to thank the organizing committee of INCoS-2017 International Conference for giving us the opportunity to organize the workshop and the Local Arrangement Chairs for facilitating the workshop organization.

We are looking forward to meet you again in the forthcoming editions of the workshop.

Workshop Organizer

Marcello Trovati	Edge Hill University, UK

BDS-2017 Organizing Committee

Workshop Organizer

Marcello Trovati	Edge Hill University, UK

Workshop PC Members

Georgios Kontonatsios	Edge Hill University, UK
Richard Conniss	University of Derby, UK
Ovidiu Bagdasar	University of Derby, UK
Peter Larcombe	University of Derby, UK
Stelios Sotiriadis	University of Toronto, Canada
Jer Hayes	IBM Research, Dublin Lab, Ireland
Xiaolong Xu	Nanjing University of Post and Telecommunications, China
Nan Hu	Nanjing University of Post and Telecommunications, China
Tao Lin	Nanjing University of Post and Telecommunications, China

INCoS-2017 Keynote Talks

Network and Information Systems Need to Embed Privacy and Security, by Design

Ann Cavoukian

Ryerson University, Toronto, Canada

Abstract. Privacy is under siege. With the growth of ubiquitous computing, online connectivity, social media, and wireless and wearable devices, people are being led to believe they have no choice but to give up on privacy. Not so! In this Keynote Talk, I will outline a privacy framework called Privacy by Design that will enable our privacy and our freedom, to live now and well into the future. While some believe that fundamental privacy protections will be challenged by the operation of Big Data and the Internet of Things, I dispel the notion that privacy acts as a barrier to data analytics and the innovations they can spark. I argue that the limiting paradigm of "zero-sum"—that you can have either privacy or innovation, but not both—is an outdated, win/lose model of approaching the question of privacy in the age of Big Data, IoT, network and information systems. Instead, a "positive-sum" solution is needed in which the interests of both sides may be met, in a doubly enabling, "win-win" manner through Privacy by Design (PbD). PbD is predicated on the rejection of zero-sum propositions by proactively identifying the risks and embedding the necessary protective measures into the IT and data architecture involved. I will demonstrate how you can embed privacy into virtually any system or operation to achieve positive-sum outcomes, enabling both privacy and data utility—not one at the expense of the other.

Cooperative Self-driving Vehicles

Shahrokh Valaee

University of Toronto, Toronto, Canada

Abstract. Can autonomous driving fully replace the driver with a smart control system? We are witnessing the emergence of self-driving vehicles, which intend to be an assistant to, or completely replace the driver. Unfortunately, we also hear about accidents that such autonomous vehicles are involved in. In this talk, we will show that autonomous driving will indeed be the start of a new chapter for automobiles that will pave the path for the more advanced connected car technology. Despite much progress in manufacturing sophisticated vehicles, the communication methods among drivers on roads are still primitive, through visual contact. Autonomous driving addresses some of these issues by using advance sensing to enhance safety. However, sensing quickly loses its effectiveness in high speeds, severe weather conditions and non-line of sight. This talk will explore the need for communication in smart vehicles and discusses the research challenges. We will investigate both the WiFi and 5G cellular technologies and study their role in cooperative self-driving vehicles.

Contents

The 5th International Workshop on Frontiers in Intelligent Networking and Collaborative Systems (FINCoS-2017)

The 3rd International Workshop for Collaborative e-Business Systems (IWCBS-2017)

The 3rd International Workshop on Theory, Algorithms and Applications of Big Data Science (BDS-2017)

The 9th International Conference on International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017)

Fine Tuning a Bayesian Network and Fairly Allocating Resources to Improve Procurement Performance

Mohammad Hassan Abolbashari(✉), Omar Khadeer Hussain, Morteza Saberi, and Elizabeth Chang

School of Business, University of New South Wales, Canberra, BC 2610, Australia
m.abolbashari@student.adfa.edu.au

Abstract. Procurement is one of the most important activities in any organization. Hence it is vital for an organization to track procurement practices. Through performance measurement, the organization will have a clear understanding on how it's performing as well as the effect of any action that it makes towards improvement. In our previous work, we proposed a Bayesian Network (BN) model to measure the level of procurement performance in an organization. This paper extends that model in two ways. First it uses the Best-Worst Method (BWM) to adjust the impact of each KPI on the procurement performance according to its importance to the overall business strategy. Second is by using the relative importance of the KPIs, it demonstrates how procurement can be improved by re-allocating the available resources among the KPIs in a fair way.

Keywords: Procurement performance management · Fair resource allocation · Bayesian network

1 Introduction

Procurement is the act of buying goods or services from an external supplier. In the recent past, it has evolved from a single task of purchasing to a multi-step time consuming process. (Aissaoui et al. 2007) have pointed out the significance of procurement by stating that since there has been a huge increase in outsourcing the acquirement of products and services, procurement has become an inevitable part of any business. This key strategic activity (Abdollahi et al. 2015) is not just a backup function for normal daily operations, but vital for securing a competitive edge in the market. Given the importance of procurement, research also shows that it consumes a significant share of any organization's costs (Nair et al. 2015; Quinn 2005). This means that a slight reduction in procurement costs through improving the procurement process could lead to a significant amount of monetary savings. However for that we need a framework that is capable of determining the procurement performance of an organization and then determine ways to improve it in order to secure a competitive edge in the market. This paper proposes a solution to this problem by extending our previous work (Abolbashari et al. 2017) in which we developed a Bayesian Network model to

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_1

measure the current procurement performance of an organization. The structure of this paper is as follows. In Sect. 2 we will discuss the previous work related to this study followed by the problem statement. In Sect. 3 we discuss our solution to improve the procurement performance in organizations. Section 4 includes a case study to demonstrate the functionality of our proposed solution. In Sect. 5, we conclude the paper with suggestions for future work.

2 Related Work

The majority of the research in the literature which aim to investigate and improve the procurement process only focus on individual activities of this integrated process (Moschuris 2015; Trapp and Sarkis 2016; Zimmer et al. 2015; Malik et al. 2016; Venkatesh and Luthra 2016; Awasthi and Kannan 2016) and less on the procurement process as a whole. Although the former approaches are useful and beneficial in improving each individual activity, but increasing procurement performance is a combined effort; requiring all participating activities to be considered simultaneously, in a single model. Resources within the procurement office are limited to a fixed amount. In other words, focusing on a single task and trying to improve it in silo without considering other tasks will end up in consuming all the available resources (which are finite) for that particular task only, leaving no more resources for the improvement of other tasks. This strategy results in suboptimal performance, meaning that a specific part of the system is functioning well as a significant amount of resources have been allocated to it. However, other parts are left with no attention and might not be functioning as they should be. This scenario can result into an overall poor performance of the system which in this case is procurement performance. To tackle this issue, in our previous work a BN modelling approach was proposed that considers all different processes of the procurement process simultaneously and uses them to measure the procurement performance. Figure 1 shows a snapshot of our developed BN model which is the starting point for this paper. As shown in the figure, once the BN is formed and values are assigned to the root nodes, the current level of performance is measured. The value of each root node is the result of a certain amount of resources assigned to it. The root nodes (KPIs) then accumulate through the Bayesian Network to ascertain an organization's performance in procurement. In Fig. 1, the following KPIs are used to measure the procurement performance: ***Procurement Cycle time:*** The time it takes to complete procurement expressed in two levels (Short, Long). ***E-procurement:*** The use of e-procurement platforms in the process expressed as (Yes, No). ***Supplier performance:*** The supplier's performance level expressed in two levels as (High, Low). ***Quality:*** The level in which the procured service/product meets the end user's expectations expressed as (High, Low). ***Effectiveness:*** The amount the right tasks re being done expressed in two levels as (High, Low). ***Efficiency:*** The amount of doing the tasks in a right way expressed in two levels as (High, Low). ***Buyer's smartness level:*** The procurement performance level expressed in three levels as (Smart, Normal, Poor).

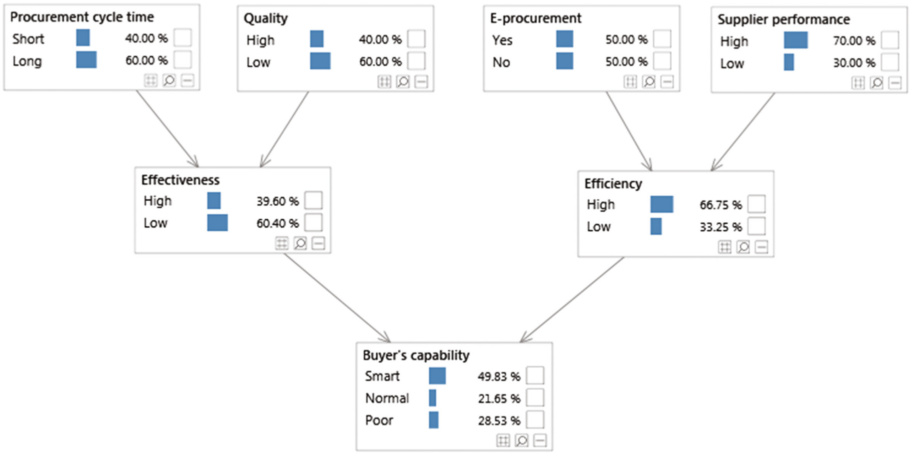


Fig. 1. The BN model for quantifying procurement performance

In this study, we capture and apply the insight of optimization from Operations Research (OR) for procurement performance improvement through fair resource allocation (Ye et al. 2017).

The objective of this paper is twofold. First is to rank the KPIs based on their importance in the criteria which the organization thinks as important. The ranking results will be used to adjust the impact that they will have on the organization's procurement process to determine the accurate overall level of performance. Second, the ranking results will be used to conduct a fair allocation of resources among the activity associated with each KPI. We will then show how this merit-based allocation of the resources will result into a higher level of procurement performance from what was being determined earlier.

The research question in this paper is as follows: Given a certain number of KPIs in the context of procurement performance measurement, (a) what impact should each of them have on determining the accurate level of performance measurement and (b) what portion of the total amount of available resources should be allocated to each KPI so that a certain (not necessarily the maximum) level of performance is achieved in each KPI which leads to achieving the maximum feasible level of procurement performance? This objective of allocation is not necessarily a cost driven optimization problem since cost is not always the only concern.

3 Solution Proposal

The steps to answer the research question are as follows:

1. **Prioritization:** In this step, we have to determine the importance of each KPI used to measure the procurement performance with respect to the organization's business and strategic priorities. Each KPI measures how well the organization satisfies a criterion. However the evolution of time brings changes to the number of criteria an

organization considers as important as well as the importance of each KPI. Hence it is vital to update the procurement performance measurement system. This will help in two ways. First, is to update and determine the new level of procurement performance whereas the second is to allocate the resources to each KPI according to its level of importance. As mentioned before, this prioritization is not just cost oriented but can involve other criteria which the organization considers as important. Determining the relevant priority of each criterion is a MCDM ranking problem which the output is a numerical weight for each KPI that demonstrates not only its importance but also the level of its superiority among other KPIs. The importance of each KPI is the amount of contribution it has (or should have) in the overall performance output.

2. **Model adjustment:** Once the relative weightings of the KPIs with regard to the criteria are identified, the impact which these KPIs have on the overall procurement performance now is adjusted to correctly reflect the final output. The adjusted final output value will show the updated level of performance which could either be better or worse than the initial result. This is because the KPIs which are more important to the organization and should be mapping the overall performance score now have a higher impact on the output. If the value of these KPIs are either low or high, then the negative or positive impact of them will be intensified and reflected on the output.
3. **Resource re-allocation:** In this step, the amount of resources allocated to each KPI is adjusted to be proportional to the level of significance of that KPI. This allocation is fair and has some characteristics such as being pareto efficient which will be explained further in the following sections. In this step, the overall performance level is expected to improve. This is because from step 2 the KPIs which are more important to the organization and now have a higher impact on determining the performance level are now receiving more resources in this step which will increase their value and the overall score respectively.

The solution proposal is demonstrated in Fig. 2. By following the three steps mentioned in this section, the aim of this paper is to improve the current performance of an organization without calling for additional resources to be spent. In the next section, the feasibility of our proposed approach to achieve the objective is demonstrated through a case study. The three steps mentioned earlier will be implemented in the model and results will be explored.

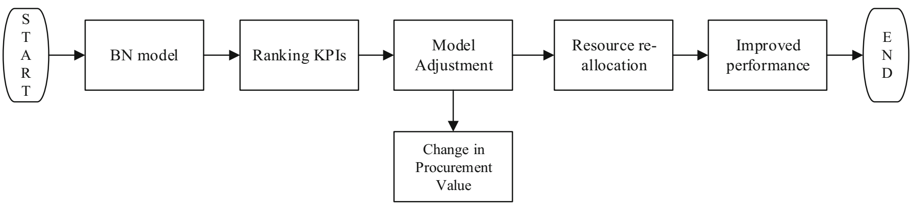


Fig. 2. Flowchart of the solution proposal

4 Case Study

As shown in Fig. 1, the considered KPIs in this study are Procurement Cycle time, E-procurement, Supplier performance, Quality, Effectiveness, Efficiency and the overall performance level which in this study we name it as Buyer's smartness. In a real world procurement scenario, there are more KPIs involved rather than the seven mentioned above. However since the aim of this paper is to demonstrate the improvement process through BN modelling rather than the context which it is applied in, only a few number of KPIs have been selected. The nodes (KPIs) are linked according to conditional probability distributions between them and the current level of performance is determined. We consider that the procurement office has distributed its available budget among the four KPIs (Procurement cycle time, Quality, E-procurement and Supplier performance) so that a certain amount of performance is acquired for each one of them as shown in Fig. 1.

We also assume that the current distribution of the budget among the four KPIs is Pareto-efficient; which means the total amount of budget has been completely allocated among the KPIs and adding to the expenditure of a KPI requires the same amount to be cut off from the allocated resource of another KPI (Ghodsi et al. 2011). At this stage, we use the Best-Worst method (BWM) (Rezaei 2015) to identify the relative importance of each KPI to the organization according to different criteria so that a more robust budget allocation could be carried out. Although other ranking methods such as AHP, TOPSIS, etc. could also be used here, BWM is one of the most recent MCDM approaches and provides more consistent results when compared to other classical models (Rezaei 2015). The five steps associated with the BWM which need to be carried out are as follows:

1. Determining the decision criteria: $C_1, C_2, \ldots, C_n$
2. Specifying the best (most desirable/significant) and worst (least desirable/significant) criteria
3. Comparing the best criterion against all the other criteria: $a_{B1}, a_{B2}, \ldots, a_{Bn}$ where a_{Bi} is the preference of the best criterion against criterion i on a scale of 1 to 10 (any other scale could be assumed) and $a_{BB} = 1$
4. Comparing all the criteria against the worst criteria: $a_{1W}, a_{2W}, \ldots, a_{nW}$ where a_{iW} is the preference of the each criterion against the worst criterion on a scale of 1 to 10 $a_{WW} = 1$
5. Determining the weights of the criteria: $W_1, W_2, \ldots, W_n$ and $\sum_{i=1}^{n} W = 1$

The aim here is to rank the four root node KPIs by allocating a weight to each of them. Prior to doing so, we first need to determine the criteria against which these four KPIs will be ranked. Moreover, the criteria will not be at the same importance and the five steps of the BWM mentioned earlier need to be used in order to identify the weight of each criterion too.

We consider that a CEO of the company whose procurement performance is being assessed has suggested the following four criteria as the most important ones to the organization namely Reputation[1] (C_1), Environment[2] (C_2), Customer satisfaction[3] (C_3) and undependability[4] (C_4) (Step 1). These criteria are not fixed and may differ from case to case. The expert was then asked to assign the most and least important criteria to his organization that led to the ranking of as reputation the most important and environment as the least important (Step 2). Following the BWM, the best criterion was compared against all the other criteria on a scale of 1 to 10 where 10 is the maximum value and 1 is the least. Results were as in Table 1.

Table 1. Comparing the best criterion with other criteria on a scale of 1 to 10

	C_1	C_2	C_3	C_4
C_B: Reputation	1	10	3	8

In the next step, all of the criteria were compared against the worst criteria regarding their favorability and results are as in Table 2.

Table 2. Comparing the criteria with the worst criterion

	C_W: Environment
C_1	10
C_2	1
C_3	10/3 = 3.33
C_4	10/8 = 1.25

In this paper, since the scale of the example is not big, the scorings are accurately performed to accomplish full consistency (refer to the scores given for C_3 and C_4 in Table 2). However, if the consistency is not met at this stage (specially for larger scale problems) then to determine the optimum weights with an acceptable level of consistency, an additional step as mentioned by Rezaei (2015) needs to be carried out which includes finding the optimum weights through solving a mathematical problem that secures an acceptable level of consistency among the results. The final relative weight of each criterion is calculated as from $W_i^* = \frac{W_i}{\sum W_i}$ as follows:

$$W_1^* = \frac{10}{22} = 0.45 \quad W_2^* = \frac{1}{22} = 0.04 \quad W_3^* = \frac{3}{22} = 0.13 \quad W_4^* = \frac{8}{22} = 0.36$$

where W_i^* is the final weight of criteria i. At this stage, the four procurement KPIs are compared against each criterion individually.

[1] The organization's reputation in the community against competitors.

[2] Conserving the environment and aligning business strategies with environmental regulatory schemes.

[3] The satisfactory level of end users.

[4] The level of the organization's undependability to suppliers and business partners.

BWM comparison of the KPIs against C_1: Reputation
In the first step, the best and worst KPIs are determined with regard to the first criterion, reputation. The best (most important) KPI in this regard was chosen by the expert as Quality and the worst (least important) was chosen as E-procurement. Next, the best KPI is compared against the other KPIs as shown in Table 3.

Table 3. Comparing the best KPI with other KPIs with regard to reputation

	KPI_1: procurement cycle time	KPI_2: quality	KPI_3: E-procurement	KPI_4: supplier performance
KPI_B: Quality	7	1	10	3

The final weightings for the KPIs with regard to reputation are as $W_{11}^* = 0.33$, $W_{21}^* = 0.04$, $W_{31}^* = 0.47$ and $W_{41}^* = 0.14$ where W_{ij} is the weight of KPI i regarding criteria j.

BWM comparison of the KPIs against C_2: Environment
Considering the second criteria environment, E-procurement was selected as the best KPI and Procurement cycle time as the worst KPI as shown in Table 4. The remaining steps are similar to what has been undertaken for the previous criteria and the weights of the KPIs are:

Table 4. Comparing the best KPI with other KPIs with regard to environment

	KPI_1	KPI_2	KPI_3	KPI_4
KPI_B: E-procurement	10	9	1	7

$$W_{12}^* = 0.37, W_{22}^* = 0.33, W_{32}^* = 0.03 \; and \; W_{42}^* = 0.25$$

BWM comparison of the KPIs against C_3: Customer satisfaction
With respect to customer satisfaction, Quality was selected as the best KPI and supplier performance as the worst KPI as shown in Table 5 and the weights of the KPIs are:

Table 5. Comparing the best KPI with other KPIs with regard to customer satisfaction

	KPI_1	KPI_2	KPI_3	KPI_4
KPI_B: Quality	7	1	2	10

$$W_{13}^* = 0.35, W_{23}^* = 0.05, W_{33}^* = 0.1 \; and \; W_{43}^* = 0.5$$

BWM comparison of the KPIs against C_4: Undependability

With respect to undependability, supplier performance was selected as the best KPI and Procurement cycle time as the worst KPI as shown in Table 6.

Table 6. Comparing the best KPI with other KPIs with regard to undependability

	KPI_1	KPI_2	KPI_3	KPI_4
KPI_B: Supplier performance	10	2	4	1

$$W_{14}^* = 0.58, W_{24}^* = 0.11, W_{34}^* = 0.23 \; and \; W_{44}^* = 0.05$$

At this stage, the final weight for each KPI could be calculated from $W_i^* = \sum W_j W_{ij}$ where W_i^* is the final weight for KPI i, W_j is the weight of criteria j and W_{ij} is the weight of KPI i regarding criteria j. Results are as:

$$W_1^* = 0.16, \; W_2^* = 0.31, \; W_3^* = 0.08 \; and \; W_4^* = 0.42$$

Hence, Supplier performance is the most important KPI, followed by Quality, Procurement cycle time and E-procurement. According to these weights, the Conditional Probability Distributions (CPDs) in the BN are adjusted so that they capture the actual level of importance of the KPIs. For instance, from Fig. 1, Supplier performance should now have a significant impact on determining the efficiency level. To explain how the new CPD values are generated, we will take the example of efficiency here. According to Fig. 1, the probability distribution of efficiency is conditioned to e-procurement and supplier performance as in Table 7.

Table 7. CPD of *efficiency*

E-procurement (EP)	Supplier performance (SP)	Efficiency	
		High	Low
Yes	High	0.95	0.05
No	High	0.7	0.3
Yes	Low	0.4	0.6
No	Low	0.2	0.8

To adjust the CPD values based on the relative importance of the root nodes, we need to configure a relationship between the importance (weights) of the root nodes and the probability distribution assigned to them. The values in Table 7 have been primarily elicited and suggested by the domain experts. In the second row of this table, although E-procurement (EP) is “No” and Supplier Performance (SP) is “High” and vice versa in the table’s third row, the probability distribution has not been suggested as

0.5 and 0.5 respectively, which reveals that the expert, at the time of constructing the BN, had given SP a higher priority than EP, since there is more alignment between the state of SP and efficiency in either cases (i.e. when SP is "High" then efficiency is expected to be "High" as well). To reveal what weight did the expert have in mind for SP, we conduct a simple reasoning for both cases. If SP and EP have been assumed to have the same importance among the four root nodes ($W_3 = W_4 = 0.25$), then the probability distribution would have been suggested as "0.5, 0.5"; but now that the probability distribution is suggested as "0.7, 0.3" the presumed weight for SP could be calculated from $\frac{0.25}{x=?} = \frac{0.5}{0.7}$ and $\frac{0.25}{x=?} = \frac{0.5}{0.6}$ which will result in $x = 0.35$ and 0.3 respectively ($x \cong 0.325$). Given the fact that if $W_3 = 0$ and $W_4 = 0.5$, the probability distribution would have been considered as the maximum threshold of (0, 1), we now have enough points $(x,\ y) = (0.325, 0.7), (0.5, 1)$ to perform a simple linear regression ($y = ax + b$) to map the relationship between the root nodes' weights and the probability values as $Y = 1.7x + 0.14$. Hence to derive the probability value for the case which $W_4 = 0.42$, we let $x = 0.42$ which will give $y = 0.85$.

After the new probability distributions are obtained and the model is adjusted, the updated BN will appear as in Fig. 3. As it can be seen from this figure, only by model adjustment the level of smartness has increased from 49.83% to 51.78%. That is because the effect of Supplier performance which itself has a good score is empowered and now has a higher effect in determining the overall output, due to it achieving the highest rank among the KPIs ($W_4^* = 0.42$). As mentioned before, this could have had a reverse effect too. If for instance Procurement cycle time and Quality which have the lowest values among the KPIs were assigned a very high rank in the prioritization process, their negative effect on the overall score would have been intensified, causing

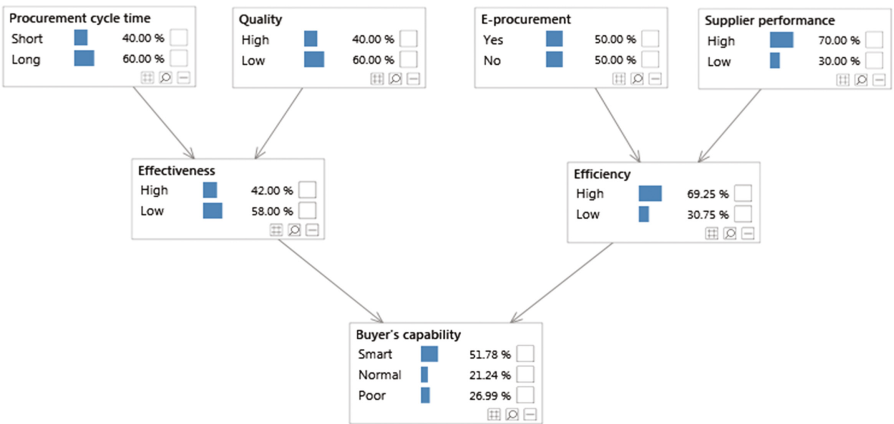


Fig. 3. The adjusted BN model based on the importance of each KPI

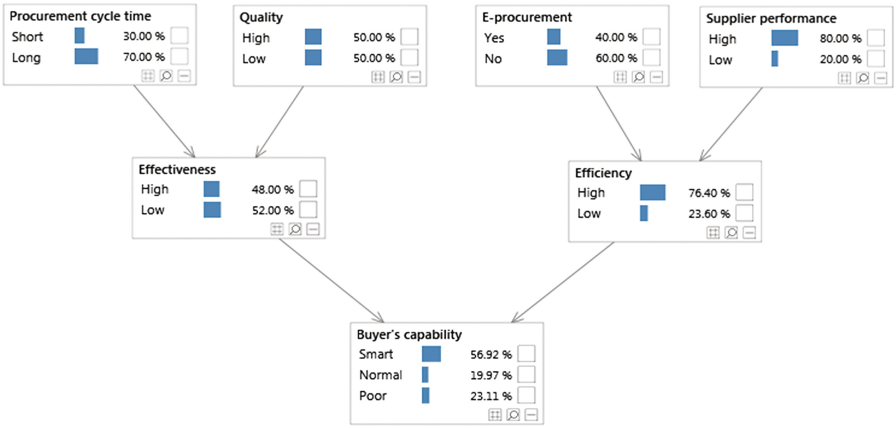


Fig. 4. The final BN model demonstrating improvement in the Buyer's capability

the buyer's smartness level to drop. Respectively, KPIs with higher level of importance now can be allocated more resources while the same amount is cut off from the less important KPIs. This action will result into improving the performance level on the KPIs which are more important and vice versa. Results are shown in the updated BN as depicted in Fig. 4.

By comparing the results from Figs. 1 and 4, the smartness level has improved from 49.83% to 56.92%. To validate the results, two other scenarios are also presented. In each of these scenarios, different weightings for both the criteria and the KPIs are considered. In scenario 2, Environment was chosen as the most important criterion by the organization while in scenario 3, Undependability was considered as the best criterion. The calculations and results are shown in Tables 8 and 9.

Table 8. Comparing the best criteria with other criteria and the final weighting of the criteria in scenarios 2 & 3 (C_1: Reputation, C_2: Environment, C_3: Customer satisfaction, C_4: Undependability)

		C_1	C_2	C_3	C_4
Scenario2	C_B: Environment	10	1	5	3
	W_j^*	0.52	0.05	0.26	0.15
Scenario 3	C_B: Undependability	7	8	10	1
	W_j^*	0.26	0.30	0.38	0.03

Table 9. The final ranking of the KPIs in scenario 2 and 3 (KPI_1: Procurement cycle time, KPI_2: Quality, KPI_3: E-procurement, KPI_4: Supplier performance)

	Criteria	KPI_B	KPI_1	KPI_2	KPI_3	KPI_4
Scenario 2	C_1: Reputation	E-procurement	3	6	1	10
		W^*_{i1}	0.15	0.3	0.05	0.5
	C_2: Environment	Procurement cycle time	1	10	2	1
		W^*_{i2}	0.0714	0.7142	0.1428	0.0714
	C_3: Customer satisfaction	Supplier performance	5	5	10	1
		W^*_{i3}	0.2380	0.2380	0.4761	0.0476
	C_4: Undependability	Quality	10	1	8	6
		W^*_{i4}	0.4	0.04		04
	W^*_i		0.2034	0.2596	0.2049	0.3119
Scenario 3	C_1: Reputation	Supplier performance	8	9	10	1
		W^*_{i1}	0.285714	0.321429	0.357143	0.035714
	C_2: Environment	Quality	1	1	10	4
		W^*_{i2}.	0.0625	0.0625	0.625	0.25
	C_3.: Customer satisfaction	E-procurement	7	10	1	3
		W^*_{i3}.	0.3333	0.4761	0.0476	0.1428
	C_4: Undependability	Quality	3	1	2	10
		W^*_{i4}	0.1875	0.0625	0.125	0.625
	W^*_i		0.2253	0.28514	0.3022	0.1573

Once the initial BN model in Fig. 1 is modified based on the results of scenarios 2 and 3, the performance level in each scenario is acquired at two stages namely model adjustment and resource re-allocation. Results are shown in Table 10 and depicted in Fig. 5.

Table 10. Comparison of the performance level in all scenarios at three stages: (1) Initial (2) After model adjustment (3) After resource re-allocation

	Procurement performance level		
	Initial	After model adjustment	After resource re-allocation
Scenario 1	49.83	51.78	56.92
Scenario 2	49.83	50.72	54.35
Scenario 3	49.83	46.07	51.25

According to Table 10, by model adjustment there is the possibility for the performance level to whether increase or decrease, however, after re-allocating the resources based on the retrieved importance of the KPIs (a merit-based allocation), the results in all scenarios show improvement in the level of performance.

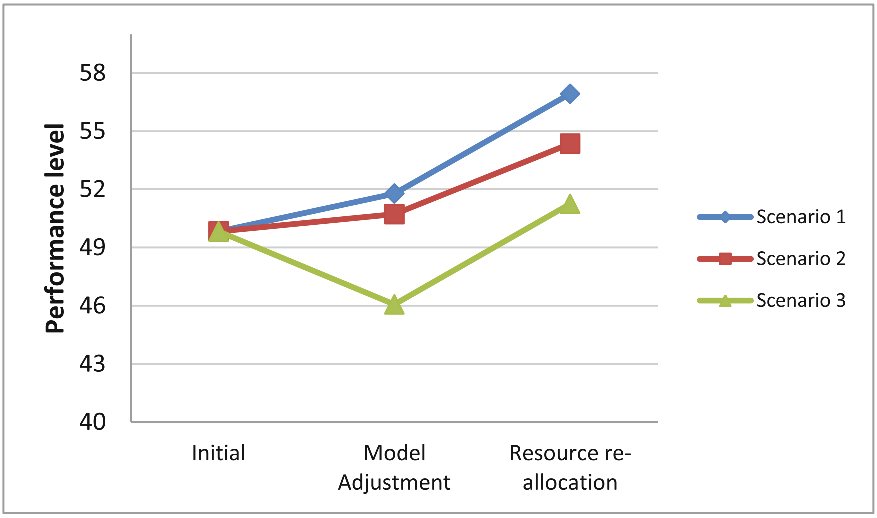


Fig. 5. The change in performance level in scenario 1, 2 & 3.

5 Conclusion and Future Work

In this paper, by revisiting the norm of resource allocation in the organization and conducting a merit-based fair resource allocation among the KPIs rather than a casual one, a higher level of performance is achieved only by using the available resources and not by calling for more. In our future work, we will extend this work by conducting a fair resource allocation to the different KPIs using mathematical techniques. This will enable the demonstration of a more accurate trade-off between the values of the KPIs with regard to the level of resources allocated to them. Moreover, to increase the accuracy of the performance level, more KPIs at different tiers are encouraged to be considered in the BN model. Although the novel methodology proposed in this paper has been performed in the context of procurement performance development, however it could be applied for improving performance in any other context as well.

References

Abdollahi, M., Arvan, M., Razmi, J.: An integrated approach for supplier portfolio selection: lean or agile? Exp. Syst. Appl. **42**(1), 679–690 (2015)

Abolbashari, M.H., et al.: A Bayesian network system for measuring and improving procurement performance in organizations. Knowl. Based Syst. (2017, submitted)

Aissaoui, N., Haouari, M., Hassini, E.: Supplier selection and order lot sizing modeling: a review. Comput. Oper. Res. **34**(12), 3516–3540 (2007)

Awasthi, A., Kannan, G.: Green supplier development program selection using NGT and VIKOR under fuzzy environment. Comput. Ind. Eng. **91**, 100–108 (2016)

Ghodsi, A., et al.: Dominant resource fairness: fair allocation of multiple resource typestle. In: NSDI, p. 24 (2011)

Malik, M.M., Abdallah, S., Hussain, M.: Assessing supplier environmental performance: applying analytical hierarchical process in the United Arab Emirates healthcare chain. Renew. Sustain. Energy Rev. **55**, 1313–1321 (2016)
Moschuris, S.J.: Decision-making criteria in tactical make-or-buy issues: an empirical analysis. EuroMed J. Bus. (2015)
Nair, A., Jayaram, J., Das, A.: Strategic purchasing participation, supplier selection, supplier evaluation and purchasing performance. Int. J. Prod. Res. **53**(20), 6263–6278 (2015)
Quinn, F.J.: The power of procurement. Supply Chain Manage. Rev. **9**(9), 6–8 (2005)
Rezaei, J.: Best-worst multi-criteria decision-making method. Omega **53**, 49–57 (2015)
Trapp, A.C., Sarkis, J.: Identifying robust portfolios of suppliers: a sustainability selection and development perspective. J. Clean. Prod. **112**, 2088–2100 (2016)
Venkatesh, V.G., Luthra, S.: Role of sustainable procurement in sustainable manufacturing operations: an indian insight. In: Dubey, R., Gunasekaran, A. (eds.) Strategic Management of Sustainable Manufacturing Operations, pp. 132–148. IGI Global (2016)
Ye, Q.C., Zhang, Y., Dekker, R.: Fair task allocation in transportation. Omega **68**, 1–16 (2017)
Zimmer, K., Fröhling, M., Schultmann, F.: Sustainable supplier management – a review of models supporting sustainable supplier selection, monitoring and development. Int. J. Prod. Res. **54**(5), 1412–1442 (2015)

A Fuzzy-Based System for Selection of IoT Devices in Opportunistic Networks Considering IoT Device Speed, Storage and Remaining Energy Parameters

Miralda Cuka[1(✉)], Donald Elmazi[1], Tetsuya Oda[2], Elis Kulla[2], Makoto Ikeda[3], and Leonard Barolli[3]

[1] Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-ku, Fukuoka 811-0295, Japan
mcuka91@gmail.com, donald.elmazi@gmail.com

[2] Department of Information and Computer Engineering, Okayama University of Science, 1-1 Ridai-cho, Kita-Ku, Okayama 700-0005, Japan
oda.tetsuya.fit@gmail.com, kulla@ice.ous.ac.jp

[3] Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-ku, Fukuoka 811-0295, Japan
makoto.ikd@acm.org, barolli@fit.ac.jp

Abstract. The opportunistic networks are the variants of Delay Tolerant Networks (DTNs). These networks can be useful for routing in places where there are few base stations and connected routes for long distances. In an opportunistic network, when nodes move away or turn off their power to conserve energy, links may be disrupted or shut down periodically. These events result in intermittent connectivity. When there is no path existing between the source and the destination, the network partition occurs. Therefore, nodes need to communicate with each other via opportunistic contacts through store-carry-forward operation. In this work, we consider the IoT device selection problem in opportunistic networks. We propose a fuzzy-based system consisting of three input parameters: IoT Device Speed (IDS), IoT Device Storage (IDST) and IoT Device Remaining Energy (IDRE). The output parameter is IoT Device Selection Decision (IDSD). We evaluate the performance of the proposed system by simulations. The simulation results show that the proposed system makes a proper selection decision of IoT-devices in opportunistic networks.

1 Introduction

The Internet is dramatically evolving and creating various connectivity methodologies. The Internet of Things (IoT) is one of those methodologies which transforms current Internet communication to Machine-to-Machine (M2M) basis.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_2

Hence, IoT can seamlessly connect the real world and cyberspace via physical objects that embed with various types of intelligent sensors. A large number of Internet-connected machines will generate and exchange an enormous amount of data that make daily life more convenient, help to make a tough decision and provide beneficial services. The IoT probably becomes one of the most popular networking concepts that has the potential to bring out many benefits [1].

Opportunistic Networks are the variants of Delay Tolerant Networks (DTNs). It is a class of networks that has emerged as an active research subject in the recent times. Owing to the transient and un-connected nature of the nodes, routing becomes a challenging task in these networks. Sparse connectivity, no infrastructure and limited resources further complicate the situation. Hence, the challenges for routing in opportunistic networks are very different from the traditional wireless networks. However, their utility and potential for scalability makes them a huge success. These networks can be useful for routing in places where one is not likely to find base stations and connected routes for long distances [2].

The Fuzzy Logic (FL) is unique approach that is able to simultaneously handle numerical data and lingustic knowledge. It is a nonlinear mapping of an input data (feature) vector into a scalar output. Fuzzy set theory and FL establish the specifics of the nonlinear mapping.

In this paper, we propose and implement a simulation system for selection of IoT devices in opportunistic networks. The system is based on fuzzy logic and considers three parameters for IoT device selection. We show the simulation results for different values of parameters.

The remainder of the paper is organized as follows. In the Sect. 2, we present a brief introduction of IoT. In Sect. 3, we describe the basics of opportunistic networks including research challenges and architecture. In Sect. 4, we introduce the proposed system model and its implementation. Simulation results are shown in Sect. 5. Finally, conclusions and future work are given in Sect. 6.

2 IoT

2.1 IoT Architecture

The typical IoT architecture can be divided into five layers as shown in Fig. 1. Each layer is briefly described below.

Perception Layer: The perception layer is similar to physical layer in OSI model which consists of the different types of sensor devices and environmental elements. This layer generally deals with identification and collection of specific information by each type of sensor devices. The gathered information can be location, wind speed, vibration, pH level, humidity, amount of dust in the air and so on. The gathered information is transmitted through Network layer toward central information processing system.

Network Layer: The Network layer plays an important role in securely transferring and keeping the sensitive information confidential from sensor devices to the

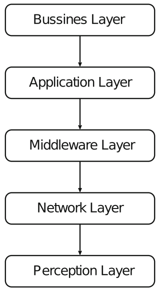


Fig. 1. IoT Architecture Layers

central information processing system through 3G, 4G, UMTS, WiFi, WiMAX, RFID, Infrared and Satellite dependent on the type of sensors devices. Thus, this layer is mainly responsible for transfering the information from Perception layer to upper layer.

Middleware Layer: The devices in the IoT system may generate various type of services when they are connected and communicate with others. Middleware layer has two essential functions, including service management and store the lower layer information into the database. Moreover, this layer has capability to retrieve, process, compute information, and then automatically decide based on the computational results.

Application Layer: Application layer is responsible for inclusive applications management based on the processed information in the Middleware layer. The IoT applications can be smart postal, smart health, smart car, smart glasses, smart home, smart independent living, smart transportation, etc.

Business Layer: This layer functions cover the whole IoT applications and services management. It can create practically graphs, business models, flow chart and executive report based on the amount of accurate data received from lower layer and effective data analysis process. Based on the good analysis results, it will help the functional managers or executives to make more accurate decisions about the business strategies and roadmaps.

2.2 IoT Protocols

In following we will briefly describe about the most frequently used protocols for Machine-to-Machine (M2M) communication.

The Message Queue Telemetry Transport (MQTT) is a Client Server publishes or subscribes messaging transport protocol. It is light weight, open, simple and designed so as to be easy to implement. The protocol runs over TCP/IP or over other network protocols that provide ordered, lossless, bi-directional connections. The MQTT features include the usage of the publish/subscribe message pattern which provides one-to-many message distribution, a messaging transport that is agnostic to the content of the payload. Furthermore, the MQTT protocol has not only minimized transport overhead and protocol exchange to reduce network traffic but also has an extraordinary mechanism to notify interested parties when an abnormal disconnection occurs as well.

The Constraint Application Protocol (CoAP) is a specialized web transfer protocol for use with constrained nodes and constrained networks. The nodes often have 8-bit microcontroller with small amounts of ROM and RAM, while constrained network often have high packet error rate and typical throughput is 10 kbps. This protocol designed for M2M application such as smart city and building automation. The CoAP provides a request and response interaction model between application end points, support build-in discovery services and resources, and includes key concepts of the Web such as URIs and Internet media types. CoAP is designed to friendly interface with HTTP for integration with the Web while meeting specialized requirements such as multicast support, very low overhead and simplicity for constrained environments.

3 Opportunistic Networks

3.1 Opportunistic Networks Challenges

In an opportunistic network, when nodes move away or turn off their power to conserve energy, links may be disrupted or shut down periodically. These events result in intermittent connectivity. When there is no path existing between the source and the destination, the network partition occurs. Therefore, nodes need to communicate with each other via opportunistic contacts through store-carry-forward operation. In this section, we consider two specific challenges in an opportunistic network: the contact opportunity and the node storage.

- *Contact Opportunity:* Due to the node mobility or the dynamics of wireless channel, a node can make contact with other nodes at an unpredicted time. Since contacts between nodes are hardly predictable, they must be exploited opportunistically for exchanging messages between some nodes that can move between remote fragments of the network. The routing methods for opportunistic networks can be classified based on characteristics of participants' movement patterns. The patterns are classified according to two independent properties: their inherent structure and their adaptiveness to the demand in the network. Other approaches proposed message ferries to provide communication service for nodes in the deployment areas. In addition, the contact capacity needs to be considered [3,4].
- *Node Storage:* As described above, to avoid dropping packets, the intermediate nodes are required to have enough storage to store all messages for an unpredictable period of time until next contact occurs. In other words, the required storage space increases as a function of the number of messages in the network. Therefore, the routing and replication strategies must take the storage constraint into consideration [5].

3.2 Opportunistic Networks Architectures

In an opportunistic network, a network is typically separated into several network partitions called regions. Traditional applications are not suitable for this kind of environment because they normally assume that the end-to-end connection must exist from the source to the destination.

The opportunistic network enables the devices in different regions to interconnect by operating message in a store-carry-forward fashion. The intermediate nodes implement the store-carry-forward message switching mechanism by overlaying a new protocol layer, called the bundle layer, on top of heterogeneous region-specific lower layers.

In an opportunistic network, each node is an entity with a bundle layer which can act as a host, a router or a gateway. When the node acts as a router, the bundle layer can store, carry and forward the entire bundles (or bundle fragments) between the nodes in the same region. On the other hand, the bundle layer of gateway is used to transfer messages across different regions. A gateway can forward bundles between two or more regions and may optionally be a host, so it must have persistent storage and support custody transfers.

4 Proposed System

4.1 System Parameters

Based on Opportunistic Networks characteristics and challenges, we consider the following parameters for implementation of our proposed system.

IoT Device Speed (IDS): There are different types of IoT devices in opportunistic networks scenarios such as: mobile phone terminals, computers, cars, trains, plains, robots and so on. Considering that high speed IoT devices can transfer the information faster, they will be selected with high probability.

IoT Device Storage (IDST): In delay tolerant networks data is carried by the IoT device until a communication opportunity is available. Considering different IoT devices have different storage capabilities, the selection decision is made based on the storage capacity.

IoT Device Remaining Energy (IDRE): The IoT devices in opportunistic networks are active and can perform tasks and exchange data in different ways from each other. Consequently, some IoT devices may have a lot of remaining power and other may have very little, when an event occurs.

IoT Device Selection Decision (IDSD): The proposed system considers the following levels for IoT device selection:

- Very Low Selection Possibility (VLSP) - The IoT device will have very low probability to be selected.
- Low Selection Possibility (LSP) - There might be other IoT devices which can do the job better.

- Middle Selection Possibility (MSP) - The IoT device is ready to be assigned a task, but is not the "chosen" one.
- High Selection Possibility (HSP) - The IoT device takes responsibility of completing the task.
- Very High Selection Possibility (VHSP) - The IoT device has almost all the required information and potential to be selected and then allocated in an appropriate position to carry out a job.

4.2 System Implementation

Fuzzy sets and fuzzy logic have been developed to manage vagueness and uncertainty in a reasoning process of an intelligent system such as a knowledge based system, an expert system or a logic control system [6–19]. In this work, we use fuzzy logic to implement the proposed system.

The structure of the proposed system is shown in Fig. 2. It consists of one Fuzzy Logic Controller (FLC), which is the main part of our system and its basic elements are shown in Fig. 3. They are the fuzzifier, inference engine, Fuzzy Rule Base (FRB) and defuzzifier.

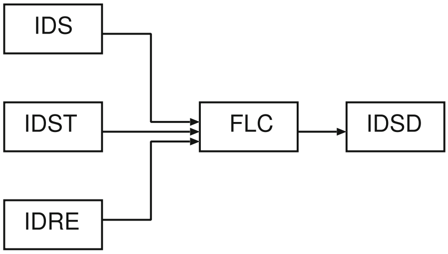


Fig. 2. Proposed system model.

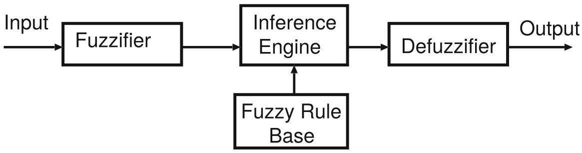


Fig. 3. FLC structure.

As shown in Fig. 4, we use triangular and trapezoidal membership functions for FLC, because they are suitable for real-time operation [20]. The x_0 in $f(x)$ is the center of triangular function, $x_0(x_1)$ in $g(x)$ is the left (right) edge of trapezoidal function, and $a_0(a_1)$ is the left (right) width of the triangular or trapezoidal function. We explain in details the design of FLC in following.

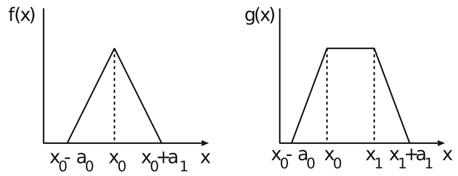


Fig. 4. Triangular and trapezoidal membership functions.

We use three input parameters for FLC:

- IoT Device Speed (IDS);
- IoT Device Storage (IDST);
- IoT Device Remaining Energy (IDRE).

The term sets for each input linguistic parameter are defined respectively as shown in Table 1.

$$
\begin{aligned}
T(IDS) &= \{Slow(Sl), Medium(Md), Fast(Fa)\} \\
T(IDST) &= \{Low(Lo), Medium(Me), High(Hi)\} \\
T(IDRE) &= \{Low(Lw), Medium(Mdm), High(Hg)\}
\end{aligned}
$$

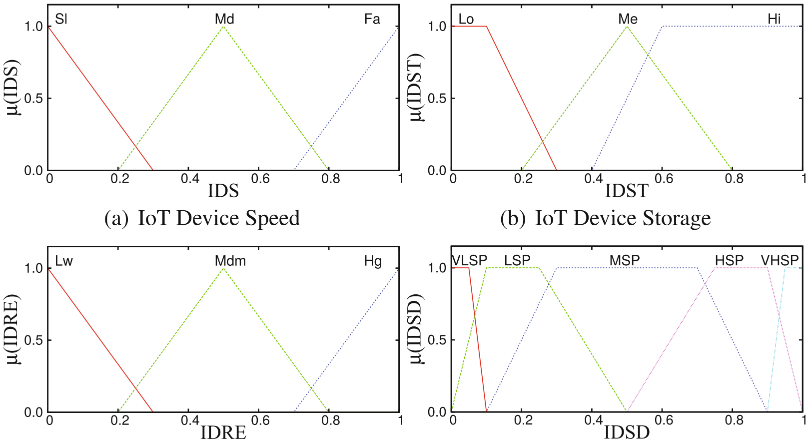
The membership functions for input parameters of FLC are defined as:

$$
\begin{aligned}
\mu_{Sl}(IDS) &= f(IDS; Sl_0, Sl_{w0}, Sl_{w1}) \\
\mu_{Md}(IDS) &= f(IDS; Md_0, Md_{w0}, Md_{w1}) \\
\mu_{Fa}(IDS) &= f(IDS; Fa_0, Fa_{w0}, Fa_{w1}) \\
\mu_{Lo}(IDST) &= g(IDST; Lo_0, Lo_1, Lo_{w0}, Lo_{w1}) \\
\mu_{Me}(IDST) &= f(IDST; Me_0, Me_{w0}, Me_{w1}) \\
\mu_{Hi}(IDST) &= g(IDST; Hi_0, Hi_1, Hi_{w0}, Hi_{w1}) \\
\mu_{Lw}(IDRE) &= f(IDRE; Lw_0, Lw_{w0}, Lw_{w1}) \\
\mu_{Md}(IDRE) &= f(IDRE; Mdm_0, Mdm_{w0}, Mdm_{w1}) \\
\mu_{Hg}(IDRE) &= f(IDRE; Hg_0, Hg_{w0}, Hg_{w1})
\end{aligned}
$$

The small letters *w0* and *w1* mean left width and right width, respectively.

Table 1. Parameters and their term sets for FLC.

Parameters	Term sets
IoT Device Speed (IDS)	Slow (Sl), Medium (Md), Fast (Fa)
IoT Device Storage (IDST)	Low (Lo), Medium (Me), High (Hi)
IoT Device Remaining Energy (IDRE)	Low (Lw), Medium (Mdm), High (Hg)
IoT Device Selection Decision (IDSD)	VLSP, LSP, MSP, HSP, VHSP



(a) IoT Device Speed (b) IoT Device Storage

(c) IoT Device Remaining Energy (d) IoT Device Selection Decision

Fig. 5. Fuzzy membership functions.

The output linguistic parameter is the Actor Node Selection Decision (IDSD). We define the term set of IDSD as:

$$\{Very\ Low\ Selection\ Possibility\ (VLSP),\\ Low\ Selection\ Possibility\ (LSP),\\ Middle\ Selection\ Possibility\ (MSP),\\ High\ Selection\ Possibility\ (HSP),\\ Very\ High\ Selection\ Possibility\ (VHSP)\}.$$

The membership functions for the output parameter *IDSD* are defined as:

$$\begin{aligned}
\mu_{VLSP}(IDSD) &= g(IDSD; VLSP_0, VLSP_1, VLSP_{w0}, VLSP_{w1})\\
\mu_{LSP}(IDSD) &= g(IDSD; LSP_0, LSP_1, LSP_{w0}, LSP_{w1})\\
\mu_{MSP}(IDSD) &= g(IDSD; MSP_0, MSP_1, MSP_{w0}, MSP_{w1})\\
\mu_{HSP}(IDSD) &= g(IDSD; HSP_0, HSP_1, HSP_{w0}, HSP_{w1})\\
\mu_{VHSP}(IDSD) &= g(IDSD; VHSP_0, VHSP_1, VHSP_{w0}, VHSP_{w1}).
\end{aligned}$$

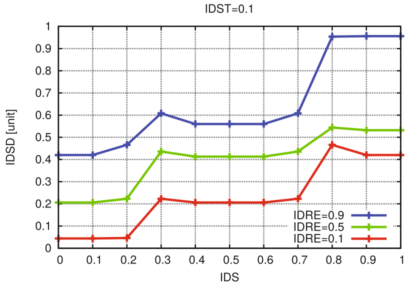
The membership functions are shown in Fig. 5 and the Fuzzy Rule Base (FRB) is shown in Table 2. The FRB forms a fuzzy set of dimensions $|T(IDS)| \times |T(IDST)| \times |T(IDRE)|$, where $|T(x)|$ is the number of terms on $T(x)$. The FRB has 27 rules. The control rules have the form: IF "conditions" THEN "control action".

Table 2. FRB of proposed fuzzy-based system.

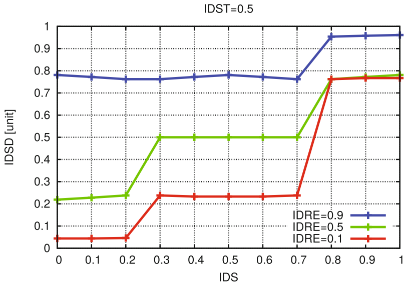
No.	IDS	IDST	IDRE	IDSD
1	Sl	Lo	Lw	VLSP
2	Sl	Lo	Mdm	VLSP
3	Sl	Lo	Hg	LSP
4	Sl	Me	Lw	VLSP
5	Sl	Me	Mdm	LSP
6	Sl	Me	Hg	MSP
7	Sl	Hi	Lw	LSP
8	Sl	Hi	Mdm	HSP
9	Sl	Hi	Hg	HSP
10	Md	Lo	Lw	VLSP
11	Md	Lo	Mdm	LSP
12	Md	Lo	Hg	MSP
13	Md	Me	Lw	VLSP
14	Md	Me	Mdm	MSP
15	Md	Me	Hg	HSP
16	Md	Hi	Lw	VLSP
17	Md	Hi	Mdm	HSP
18	Md	Hi	Hg	VHSP
19	Fa	Lo	Lw	LSP
20	Fa	Lo	Mdm	HSP
21	Fa	Lo	Hg	VHSP
22	Fa	Me	Lw	MSP
23	Fa	Me	Mdm	HSP
24	Fa	Me	Hg	VHSP
25	Fa	Hi	Lw	VHSP
26	Fa	Hi	Mdm	VHSP
27	Fa	Hi	Hg	VHSP

5 Simulation Results

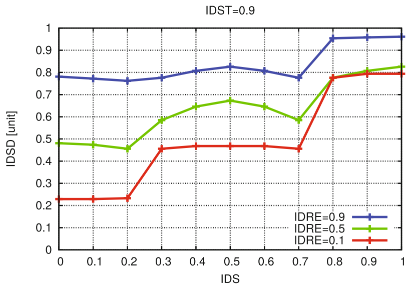
We present the simulation results in Fig. 6. In Fig. 6(a) is shown the relation between IDSD and IDS for different IDRE values. The IDST is considered 0.1. We see that when the speed is increased, the possibility of the present IoT device to be selected for carrying out a job is increased. By increasing the IDRE value, the IDSD is also increased. This means that the IoT device with higher remaining energy will be selected. The value of IDSD is increased faster when the IDS is from 0.2 to 0.8.



(a) IDST=0.1



(b) IDST=0.5



(c) IDST=0.9

Fig. 6. Results for different values of IDWT.

In Fig. 6(b) and (c), we increase the IDST value to 0.5 and 0.9, respectively. We see that with the increase of the IDST parameter, the possibility of an IoT device to be selected is increased.

6 Conclusions and Future Work

In this paper, we proposed and implemented a fuzzy-based simulation system for selection of IoT devices in opportunistic networks. We considered three parameters IDS, IDST and IDRE to select an IoT device to carry out a required task.

We evaluated the proposed system by some simulation results. The simulation results show that the highest the speed, the greater is the possibility of IoT device to be selected for carrying out a job. We can see that by increasing IDRE and IDST, the IDSD is also increased. When the IDST parameter is increased, then IDSD also is increased.

In the future work, we will consider also other parameters for IoT device selection in opportunistic networks and make extensive simulations to evaluate the proposed system.

References

1. Kraijak, S., Tuwanut, P.: A survey on internet of things architecture, protocols, possible applications, security, privacy, real-world implementation and future trends. In: IEEE 16th International Conference on Communication Technology (ICCT), pp. 26–31. IEEE (2015)
2. Dhurandher, S.K., Sharma, D.K., Woungang, I., Bhati, S.: HBPR: history based prediction for routing in infrastructure-less opportunistic networks. In: IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp. 931–936. IEEE (2013)
3. Akbas, M., Turgut, D.: APAWSAN: actor positioning for aerial wireless sensor and actor networks. In: IEEE 36th Conference on Local Computer Networks, LCN-2011, pp. 563–570, October 2011
4. Akbas, M., Brust, M., Turgut, D.: Local positioning for environmental monitoring in wireless sensor and actor networks. In: IEEE 35th Conference on Local Computer Networks, LCN-2010, pp. 806–813, October 2010
5. Melodia, T., Pompili, D., Gungor, V., Akyildiz, I.: Communication and coordination in wireless sensor and actor networks. IEEE Trans. Mob. Comput. **6**(10), 1126–1129 (2007)
6. Inaba, T., Sakamoto, S., Kolici, V., Mino, G., Barolli, L.: A CAC scheme based on fuzzy logic for cellular networks considering security and priority parameters. In: The 9th International Conference on Broadband and Wireless Computing, Communication and Applications, BWCCA-2014, pp. 340–346 (2014)
7. Spaho, E., Sakamoto, S., Barolli, L., Xhafa, F., Barolli, V., Iwashige, J.: A fuzzy-based system for peer reliability in JXTA-overlay P2P considering number of interactions. In: The 16th International Conference on Network-Based Information Systems, NBiS-2013, pp. 156–161 (2013)
8. Matsuo, K., Elmazi, D., Liu, Y., Sakamoto, S., Mino, G., Barolli, L.: FACS-MP: a fuzzy admission control system with many priorities for wireless cellular networks and its performance evaluation. J. High Speed Netw. **21**(1), 1–14 (2015)
9. Grabisch, M.: The application of fuzzy integrals in multicriteria decision making. Eur. J. Oper. Res. **89**(3), 445–456 (1996)
10. Inaba, T., Elmazi, D., Liu, Y., Sakamoto, S., Barolli, L., Uchida, K.: Integrating wireless cellular and Ad-Hoc networks using fuzzy logic considering node mobility and security. In: The 29th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA-2015, pp. 54–60 (2015)
11. Kulla, E., Mino, G., Sakamoto, S., Ikeda, M., Caballé, S., Barolli, L.: FBMIS: A fuzzy-based multi-interface system for cellular and Ad Hoc networks. In: International Conference on Advanced Information Networking and Applications, AINA-2014, pp. 180–185 (2014)
12. Elmazi, D., Kulla, E., Oda, T., Spaho, E., Sakamoto, S., Barolli, L.: A comparison study of two fuzzy-based systems for selection of actor node in wireless sensor actor networks. J. Ambient Intell. Humaniz. Comput. **6**(5), 635–645 (2015)
13. Zadeh, L.: Fuzzy logic, neural networks, and soft computing. ACM Commun. **37**(3), 77–84 (1994)
14. Spaho, E., Sakamoto, S., Barolli, L., Xhafa, F., Ikeda, M.: Trustworthiness in P2P: performance behaviour of two fuzzy-based systems for JXTA-overlay platform. Soft. Comput. **18**(9), 1783–1793 (2014)

15. Inaba, T., Sakamoto, S., Kulla, E., Caballe, S., Ikeda, M., Barolli, L.: An integrated system for wireless cellular and Ad-Hoc networks using fuzzy logic. In: International Conference on Intelligent Networking and Collaborative Systems, INCoS-2014, pp. 157–162 (2014)
16. Matsuo, K., Elmazi, D., Liu, Y., Sakamoto, S., Barolli, L.: A multi-modal simulation system for wireless sensor networks: a comparison study considering stationary and mobile sink and event. J. Ambient Intell. Humaniz. Comput. **6**(4), 519–529 (2015)
17. Kolici, V., Inaba, T., Lala, A., Mino, G., Sakamoto, S., Barolli, L.: A fuzzy-based CAC scheme for cellular networks considering security. In: International Conference on Network-Based Information Systems, NBiS-2014, pp. 368–373 (2014)
18. Liu, Y., Sakamoto, S., Matsuo, K., Ikeda, M., Barolli, L., Xhafa, F.: A comparison study for two fuzzy-based systems: improving reliability and security of JXTA-overlay P2P platform. Soft. Comput. **20**(7), 2677–2687 (2015)
19. Matsuo, K., Elmazi, D., Liu, Y., Sakamoto, S., Mino, G., Barolli, L.: FACS-MP: a fuzzy admission control system with many priorities for wireless cellular networks and its perforemance evaluation. J. High Speed Netw. **21**(1), 1–14 (2015)
20. Mendel, J.M.: Fuzzy logic systems for engineering: a tutorial. Proc. IEEE **83**(3), 345–377 (1995)

Performance Analysis of WMNs by WMN-GA Simulation System for Different WMN Architectures and TCP Congestion-Avoidance Algorithms Considering Normal and Uniform Distributions

Keita Matsuo[1], Shinji Sakamoto[2], Tetsuya Oda[3(✉)], Admir Barolli[4], Makoto Ikeda[1], and Leonard Barolli[1]

[1] Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
{kt-matsuo,barolli}@fit.ac.jp, makoto.ikd@acm.org

[2] Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
shinji.t.sakamoto@gmail.com

[3] Department of Information and Computer Engineering, Okayama University of Science (OUS), 1-1 Ridaicho, Kita-ku, Okayama 700–0005, Japan
oda.tetsuya.fit@gmail.com

[4] Department of Information Technology, Aleksander Moisiu University of Durres, L.1, Rruga e Currilave, Durres, Albania
admir.barolli@gmail.com

Abstract. In this paper, we evaluate the performance of two Wireless Mesh Networks (WMNs) architectures considering throughput, delay, jitter and fairness index metrics. For simulations, we used ns-3, Distributed Coordination Function (DCF) and Optimized Link State Routing (OLSR). We compare the performance of Transmission Control Protocol (TCP) Tahoe, Reno and NewReno for normal and uniform distributions of mesh clients by sending multiple Constant Bit Rate (CBR) flows in the network. The simulation results show that the PDR for both distributions and architectures is almost the same, but the PDR of I/B WMN for uniform distribution is a little bit higher than normal distribution. For both WMN architectures, the throughput of normal distribution is better than uniform distribution. The delay of Tahoe is a little bit lower compared with other algorithms for normal distribution and Hybrid WMN, but in case of uniform distribution, the NewReno performs better than other algorithms for both architectures. The fairness index of normal distribution is higher than uniform distribution.

1 Introduction

Wireless Mesh Networks (WMNs) [1] are important networking infrastructures. These networks are made up of wireless nodes, organized in a mesh topology,

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_3

where mesh routers are interconnected by wireless links and provide Internet connectivity to mesh clients.

WMNs distinguish for their low cost nature that makes them attractive for providing wireless Internet connectivity. Moreover, such infrastructure can be used to deploy community networks, metropolitan area networks, municipal and, corporative networks, and to support applications for urban areas, medical, transport and surveillance systems.

The main issue of WMNs is to achieve network connectivity and stability as well as QoS in terms of user coverage. This problem is very closely related to the family of node placement problems in WMNs [2–5], among them, the mesh router mesh nodes placement. We consider the version of the mesh router nodes placement problem in which we are given a grid area where to deploy a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. The objective is to find a location assignment for the mesh routers to the cells of the grid area that maximizes the network connectivity and client coverage.

As node placement problems are known to be computationally hard to solve for most of the formulations [6,7], Genetic Algorithms (GAs) has been recently investigated as an effective resolution method.

In our previous work [8–10], we used mesh router nodes placement system that is based on Genetic Algorithms (GAs) to find an optimal location assignment for mesh routers in the grid area in order to maximize the network connectivity and client coverage.

In this work, we use the topology generated by WMN-GA system and evaluate by simulations the performance of normal and uniform distributions of mesh clients considering TCP congestion-avoidance algorithms, two architectures and Distributed Coordination Function (DCF) protocol by sending multiple Constant Bit Rate (CBR) flows in the network. For simulations, we use ns-3 and Optimized Link State Routing (OLSR). As evaluation metrics we considered packet delivery ratio (PDR), throughput, delay and fairness.

The rest of the paper is organized as follows. Architectures of WMNs are presented in Sect. 2. In Sect. 3, we show the description and design of the simulation system. In Sect. 4, we discuss the simulation results. Finally, conclusions and future work are given in Sect. 5.

2 Architectures of WMNs

In this section, we describe the architectures of WMN. The architecture of the nodes in WMNs [11–14] can be classified according to the functionalities they offer as follows:

Infrastructure/Backbone WMNs: This type of architecture (also known as infrastructure meshing) is the most used and consists of a grid of mesh routers which are connected to different clients. Moreover, routers have gateway functionality thus allowing Internet access for clients. This architecture enables integration with other existing wireless networks and is widely used in neighboring communities.

Client WMNs: Client meshing architecture provides a communications network based on peer-to-peer over client devices (there is no the role of mesh router). In this case we have a network of mesh nodes which provide routing functionality and configuration as well as end-user applications, so that when a packet is sent from one node to another, the packet will jump from node to node in the mesh of nodes to reach the destination.

Hybrid WMNs: This architecture combines the two previous ones, so that mesh clients are able to access the network through mesh routers as well as through direct connection with other mesh clients. Benefiting from the advantages of the two architectures, Hybrid WMNs can connect to other networks (Internet, Wi-Fi, and sensor networks) and enhance the connectivity and coverage due to the fact that mesh clients can act as mesh routers.

3 Simulation System Description and Design

3.1 GUI of WMN-GA System

The WMN-GA system can generate instances of the problem using different distributions of client and mesh routers.

The GUI interface of WMN-GA is shown in Fig. 1. The left site of the interface shows the GA parameters configuration and on the right side are shown the network configuration parameters.

For the network configuration, we use: distribution, number of clients, number of mesh routers, grid size, radius of transmission distance and the size of subgrid.

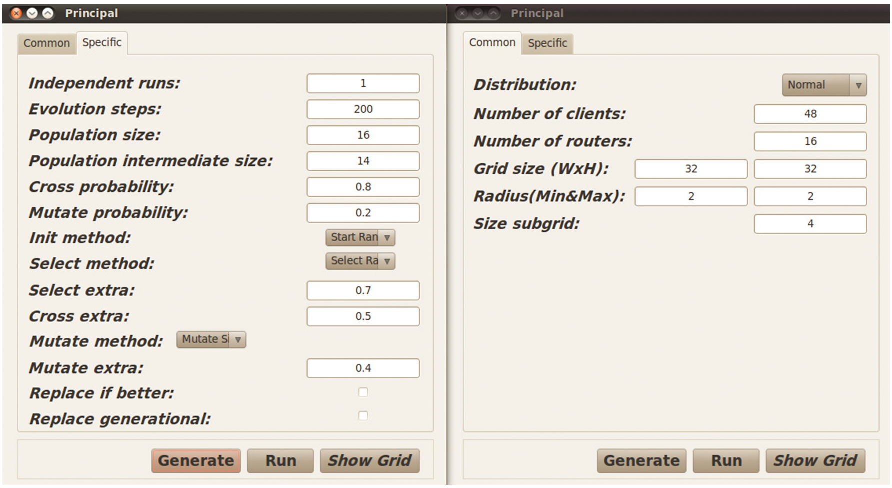


Fig. 1. GUI tool for WMN-GA system.

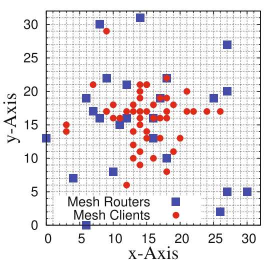
For the GA parameter configuration, we use: number of independent runs, GA evolution steps, population size, population intermediate size, crossover probability, mutation probability, initial methods, select method.

3.2 Positioning of Mesh Routers by WMN-GA System

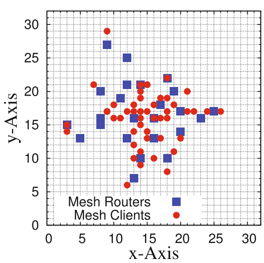
We use WMN-GA system for node placement problem in WMNs. A bi-objective optimization is used to solve this problem by first maximizing the number of connected routers in the network and then the client coverage. The input parameters of WMN-GA system are shown in Table 1. In Figs. 2 and 3, we show the location of mesh routers and clients for first generations and the optimized topologies generated by WMN-GA system for normal and exponentiall distributions, respectively.

Table 1. Input parameters of WMN-GA system.

Parameters	Values
Number of clients	48
Number of routers	16, 20, 24, 28, 32
Grid width	32 [*units*]
Grid height	32 [*units*]
Independent runs	10
Number of generations	200
Population size	64
Selection method	Linear ranking
Crossover rate	80 [%]
Mutate method	Single
Mutate rate	20 [%]
Distribution of clients	Normal, Uniform

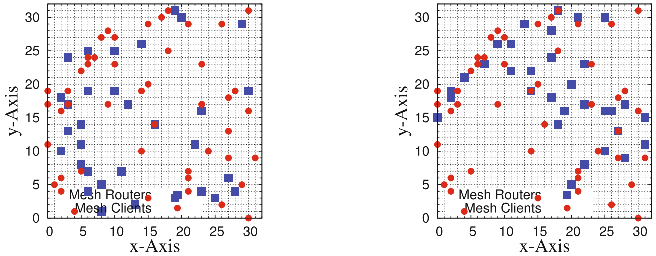


(a) Number of generations: 1 (8, 12)



(b) Number of generations: 200 (32, 35)

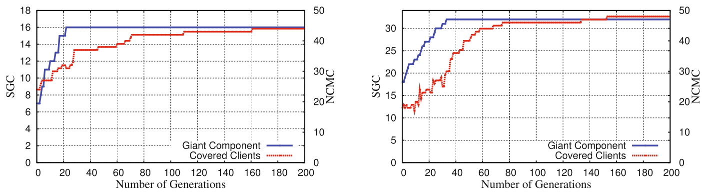
Fig. 2. Location of mesh routers by WMN-GA system for normal distribution; (m, n): m is number of connected mesh routers, n is number of covered mesh clients.



(a) Number of generations: 1 (8, 12) (b) Number of generations: 200 (32, 35)

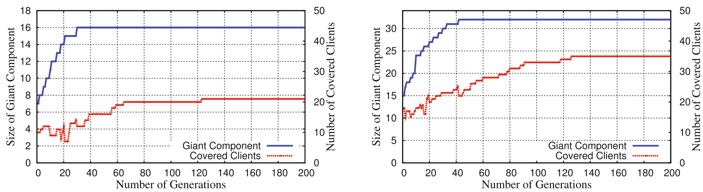
Fig. 3. Location of mesh routers by WMN-GA system for uniform distribution; (m, n): m is number of connected mesh routers, n is number of covered mesh clients.

In Figs. 4 and 5 are shown the simulation results of Size of Giant Component (SGC) and Number of Covered Mesh Clients (NCMC) vs. number of generations. After few generations, all routers are connected with each other.



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 4. SGC and NCMC vs. number of generations for normal distribution.



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 5. SGC and NCMC vs. number of generations for uniform distribution.

Then, we optimize the position of routers in order to cover as many mesh clients as possible. The simulation results of SGC and NCMC are shown in Table 2.

Table 2. Evaluation of WMN-GA system.

Number of mesh routers	Normal distribution		Uniform distribution	
	SGC	NCMC	SGC	NCMC
16	16	44	16	21
20	20	46	20	22
24	24	47	24	27
28	28	48	28	33
32	32	48	32	35

3.3 Simulation Description

We conduct simulations using ns-3 simulator. The simulations in ns-3 are done for number of generations 1 and 200. The area size is considered 640 $[m] \times 640$ $[m]$ (or 32 $[units] \times 32$ $[units]$) and the number of mesh routers is from 16 to 32. We used DCF, OLSR protocols and sent multiple CBR flows over different TCP congestion-avoidance algorithms. The pairs source-destination are the same for all simulation scenarios. Log-distance path loss model and constant speed delay model are used for the simulation and other parameters are shown in Table 3.

Table 3. Simulation parameters for ns-3.

Parameters	Values
Area Size	640 [m] × 640 [m]
Distributions of mesh clients	Normal, Uniform
Number of mesh routers	20
Number of mesh clients	48
PHY protocol	IEEE 802.11a
Propagation loss model	Log-distance path loss model
Propagation delay model	Constant speed model
MAC protocols	DCF
Maximum queue size	400
Routing protocol	OLSR
Transport protocol	TCP
TCP version	Tahoe, Reno, NewReno
Application type	CBR
Packet size	1024 [Bytes]
Number of source nodes	10
Number of destination node	1
Transmission current	17.4 [mA]
Receiving current	19.7 [mA]
Simulation time	600 [s]

3.4 NS-3

The ns-3 simulator [15] is developed and distributed completely in the C++ programming language, because it better facilitated the inclusion of C-based implementation code. The ns-3 architecture is similar to Linux computers, with internal interface and application interfaces such as network interfaces, device drivers and sockets. The goals of ns-3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users of ns-3 are free to write their simulation scripts as either *C++ main()* programs or *Python* programs. The ns-3's low-level API is oriented towards the power-user but more accessible "helper" APIs are overlaid on top of the low-level API.

In order to achieve scalability of a very large number of simulated network elements, the ns-3 simulation tools also support distributed simulation. The ns-3 support standardized output formats for trace data, such as the pcap format used by network packet analyzing tools such as tcpdump, and a standardized input format such as importing mobility trace files from ns-2 [16].

The ns-3 simulator is equipped with *Pyviz* visualizer, which has been integrated into mainline ns-3, starting with version 3.10. It can be most useful for debugging purposes, i.e. to figure out if mobility models are what you expect, where packets are being dropped. It is mostly written in Python and it works both with Python and pure C++ simulations. The function of ns-3 visualizer is more powerful than network animator (*nam*) of ns-2 simulator.

The ns-3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card or a more complex wireless IEEE 802.11 device.

The ns-3 is intended as an eventual replacement for popular ns-2 simulator. The ns-3's wifi models a wireless network interface controller based on the IEEE 802.11 standard [17]. The ns-3 provides models for these aspects of 802.11:

1. Basic 802.11 DCF with infrastructure and ad hoc modes.
2. 802.11a, 802.11b, 802.11g and 802.11s physical layers.
3. QoS-based EDCA and queueing extensions of 802.11e.
4. Various propagation loss models including Nakagami, Rayleigh, Friis, LogDistance, FixedRss, and so on.
5. Two propagation delay models, a distance-based and random model.
6. Various rate control algorithms including Aarf, Arf, Cara, Onoe, Rraa, ConstantRate, and Minstrel.

3.5 Overview of DCF Protocol

DCF is a random access scheme based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) scheme. A legacy DCF station with a packet to send will first sense the medium for activity. If the channel is idle for a Distributed Inter-Frame Space (DIFS), the station will attempt to transmit after a random back-off period. This period is referred as the Contention Window (CW). The value for the CW is chosen randomly from a range $[0, 2^n - 1]$, i.e.

$$CW_{min} \leq CW \leq CW_{max} \tag{1}$$

where n is PHY dependent. Initially, CW is set to the minimum number of slot times CW_{min}, which is defined per PHY in microseconds [18]. The randomly chosen CW value, referred as the back-off counter, is decreased each slot time if the medium remains idle. If during any period the medium becomes busy, the back-off counter is paused and resumed only when the medium becomes idle. On reaching zero, the station transmits the packet in the physical channel and awaits an acknowledgment (ACK). The transmitting station then performs a post back-off, where the back-off procedure is repeated once more. This is to allow other stations to gain access to the medium during heavy contention.

If the ACK is not received within a Short Inter-Frame Space (SIFS), it assumes that the frame was lost due to collision or being damaged. The CW value is then increased exponentially and the back-off begins once again for retransmission. This is referred as the Automatic Repeat Request (ARQ) process. If the following retransmission attempt fails, the CW is again increased exponentially, up until the limit CW_{max}. The retransmission process will repeat for up to 4 or 7 times, depending on whether the short retry limit or long retry limit is used. Upon reaching the retry limit the packet is considered lost and discarded. The retry limit is manufacturer dependent and can vary considerably.

3.6 Overview of OLSR Routing Protocol

The OLSR protocol [19] is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information, which is exchanged by means of Topology Control (TC) packets.

OLSR makes use of HELLO messages to find its one hop neighbours and its two hop neighbours through their responses. The sender can then select its Multi Point Relays (MPR) based on the one hop node which offer the best routes to the two hop nodes. By this way, the amount of control traffic can be reduced. Each node has also an MPR selector set which enumerates nodes that have selected it as an MPR node. OLSR uses TC messages along with MPR forwarding to disseminate neighbour information throughout the network. Host Network Address (HNA) messages are used by OLSR to disseminate network route advertisements in the same way TC messages advertise host routes.

3.7 Overview of TCP Congestion-Avoidance Algorithms

TCP is transport layer is the reliable connection orientated protocol that provides reliable transfer of data between the nodes [20]. It ensures that the data is reached the destination correctly without any loss or damage. The data is transmitted in the form of continuous stream of octets. The reliable transfer of octets is achieved through the use of a sequence number to each octet. Another aspect of TCP is the tree way handshakes mechanism to establish a connection between the nodes [21]. Furthermore, TCP uses the port assignment as an addressing mechanism to differentiate each connection for the cases of more TCP connection between nodes are required. After the introduction of first version of TCP several different TCP variants exist. The most famous implementation of TCP are Tahoe, Reno and NewReno.

Modern TCP implementations contain a number of algorithms aimed at controlling network congestion while maintaining good user throughput. Early TCP implementations followed a go-back model using cumulative positive acknowledgment and requiring a retransmission timer expiration to re-send data lost during transport. These TCPs did little to minimize network congestion. In our study we concentrate on three TCP congestion-avoidance algorithms [21]: TCP Tahoe [22], TCP Reno [23] and TCP NewReno [24].

3.7.1 TCP Tahoe

The Tahoe TCP implementation added a number of new algorithms and refinements to earlier implementations. The new algorithms include Slow-Start, Congestion Avoidance, and Fast Retransmit [22]. The refinements include a modification to the round-trip time estimator used to set retransmission timeout values. All modifications have been described elsewhere [22,25]. The Fast Retransmit algorithm is modified in subsequent versions of TCP. With Fast Retransmit, after receiving a small number of duplicate acknowledgments for the same TCP segment (dup ACKs), the data sender infers that a packet has been lost and retransmits the packet without waiting for a retransmission timer to expire, leading to higher channel utilization and connection throughput.

3.7.2 TCP Reno

TCP Reno retains the basic principle of Tahoe, such as slow starts and the coarse grain retransmit timer. However it adds some intelligence over it so that lost packets are detected earlier and the pipeline is not emptied every time a packet is lost [26]. Reno requires immediate acknowledgement whenever a segment is received. If we receive a number of duplicate acknowledgements then that means that sufficient time have passed and even if the segment had taken a longer path, it arrives to the receiver [27]. There is a very high probability that it was lost. So Reno suggests an algorithm called “Fast Retransmit” [28].

3.7.3 TCP NewReno

NewReno is a slight modification over TCP Reno. It is able to detect multiple packet losses and thus is much more efficient that Reno in the event of multiple packet losses. Like Reno, NewReno also enters into fast-retransmit when it receives multiple duplicate packets, however it differs from Reno in that it does not exit fast-recovery until all the data which was out standing at the time it entered fast recovery is acknowledged. Thus it overcomes the problem faced by

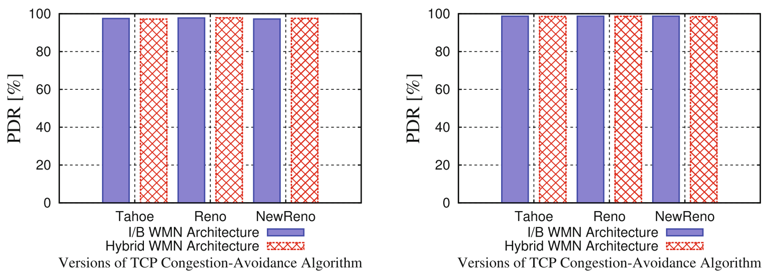
Reno of reducing the cwnd multiples times. The fast-transmit phase is the same as in Reno. The difference in the fast recovery phase which allows for multiple re-transmissions in NewReno. Whenever NewReno enters fast recovery it notes the maximums segment which is outstanding. It exits fast recovery when all the data in the window are acknowledged [24].

4 Simulation Results

We used the PDR, throughput, delay and fairness metrics to evaluate the performance of WMNs for two architectures considering TCP congestion-avoidance algorithms and normal and uniform distributions of mesh clients.

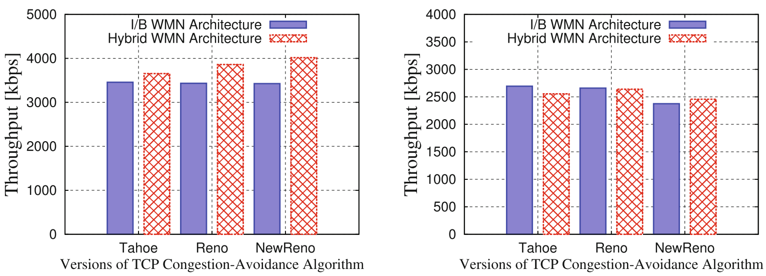
In Fig. 6, we show the simulation results of PDR. The simulation results show that the PDR for both distributions and architectures is almost the same, but the PDR of I/B WMN for uniform distribution is a little bit higher than normal distribution.

In Fig. 7, we show the simulation results of throughput. For both WMN architectures, the throughput of normal distribution is better than uniform distribution.



(a) Normal distribution (b) Uniform distribution

Fig. 6. Results of average PDR.

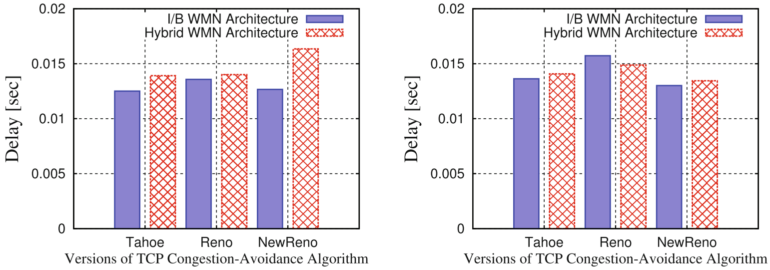


(a) Normal distribution (b) Uniform distribution

Fig. 7. Results of average throughput.

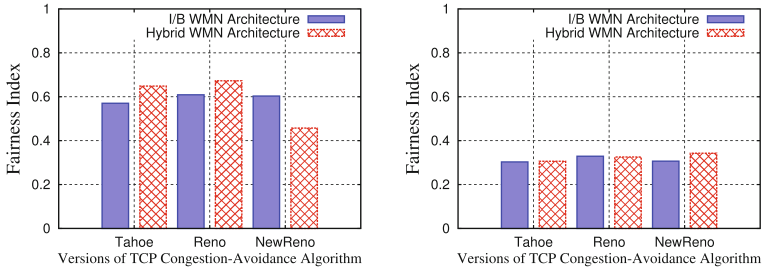
In Fig. 8, the delay of Tahoe is a little bit lower compared with other algorithms for normal distribution and Hybrid WMN, but in case of uniform distribution, the NewReno performs better than other algorithms for both architectures.

In Fig. 9, we show the fairness index. The fairness index of normal distribution is higher than uniform distribution.



(a) Normal distribution (b) Uniform distribution

Fig. 8. Results of average delay.



(a) Normal distribution (b) Uniform distribution

Fig. 9. Results of average fairness index.

5 Conclusions

In this work, we presented WMN-GA system and applied it for node placement problem in WMNs. We evaluated the performance of WMNs by WMN-GA simulation system for normal and uniform distributions of mesh clients considering DCF, OLSR and different TCP congestion-avoidance algorithms.

From the simulations we found that:

- The simulation results show that the PDR for both distributions and architectures is almost the same, but the PDR of I/B WMN for uniform distribution is a little bit higher than normal distribution.

- For both WMN architectures, the throughput of normal distribution is better than uniform distribution.
- The delay of Tahoe is a little bit lower compared with other algorithms for normal distribution and Hybrid WMN, but in case of uniform distribution, the NewReno performs better than other algorithms for both architectures.
- The fairness index of normal distribution is higher than uniform distribution.

In the future work, we would like to implement other systems and compare the performance with the proposed system.

References

1. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: a survey. Comput. Netw. **47**(4), 445–487 (2005)
2. Franklin, A., Murthy, C.: Node placement algorithm for deployment of two-tier wireless mesh networks. In: Proceedings of IEEE GLOBECOM-2007, pp. 4823–4827 (2007)
3. Muthaiah. S.N., Rosenberg, C.: Single gateway placement in wireless mesh networks. In: Proceedings of 8th International IEEE Symposium on Computer Networks, Turkey, pp. 4754–4759 (2008)
4. Tang, M.: Gateways placement in backbone wireless mesh networks. Int. J. Commun. Netw. Syst. Sci. **2**(1), 45–50 (2009)
5. Vanhatupa, T., Hännikäinen, M., Hämäläinen, T.D.: Genetic algorithm to optimize node placement and configuration for WLAN planning. In: Proceedings of 4th International Symposium on Wireless Communication Systems, pp. 612–616 (2007)
6. Lim, A., Rodrigues, B., Wang, F., Xua, Z.: k-center problems with minimum coverage. Theoret. Comput. Sci. **332**(1–3), 1–17 (2005)
7. Wang, J., Xie, B., Cai, K., Agrawal, D.P.: Efficient Mesh Router Placement in Wireless Mesh Networks. MASS, Pisa (2007)
8. Oda, T., Barolli, A., Xhafa, F., Barolli, L., Ikeda, M., Takizawa, M.: WMN-GA: a simulation system for WMNs and its evaluation considering selection operators. J. Ambient Intell. Humanized Comput. (JAIHC) **4**(3), 323–330 (2013). Springer
9. Ikeda, M., Oda, T., Kulla, E., Hiyama, M., Barolli, L., Younas, M.: Performance evaluation of WMN considering number of connections using NS-3 simulator. In: The 3-rd International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2012), pp. 498–502 (2012)
10. Oda, T., Elmazi, D., Barolli, A., Sakamoto, S., Barolli, L., Xhafa, F.: A genetic algorithm based system for wireless mesh networks: analysis of system data considering different routing protocols and architectures. J. Soft Comput. (SOCO) **20**(7), 2627–2640 (2016). Springer
11. Xhafa, F., Sanchez, C., Barolli, L.: Locals search algorithms for efficient router nodes placement in wireless mesh networks. In: International Conference on Network-Based Information Systems (NBiS-2009), pp. 572–579 (2009)
12. Oda, T., Barolli, A., Spaho, E., Barolli, L., Xhafa, F.: Analysis of mesh router placement in wireless mesh networks using Friedman test. In: The 28-th IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA-2014), pp. 289–296 (2014)

13. Oda, T., Sakamoto, S., Barolli, A., Ikeda, M., Barolli, L., Xhafa, F.: A GA-based simulation system for WMNs: performance analysis for different WMN architectures considering TCP. In: The 8-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2014), pp. 120–126 (2014)
14. Oda, T., Barolli, A., Spaho, E., Xhafa, F., Barolli, L., Takizawa, M.: Evaluation of WMN-GA for different mutation operators'. Int. J. Space-Based Situated Comput. (IJSSC) **2**(3), 149–157 (2012). Inderscience
15. ns-3. https://www.nsnam.org/
16. The Network Simulator-ns-2. http://www.isi.edu/nsnam/ns/
17. IEEE 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Computer Society Std. (2007). http://standards.ieee.org/getieee802/download/802.11-2007.pdf
18. IEEE-SA: IEEE 802.11 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (1999)
19. Clausen, T., Jacquet, P.: Optimized Link State Routing Protocol (OLSR), RFC 3626 (Experimental) (2003)
20. Kaur, R., Josan, G.S.: Performance evaluation of congestion control TCP variants in vanet using Omnet++. Int. J. Eng. Res. Appl. (IJERA) **2**(5), 1682–1688 (2012)
21. Fall, K., Floyd, S.: Simulation-based comparisons of Tahoe, Reno and SACK TCP. ACM SIGCOMM Comput. Commun. Rev. **26**(3), 5–21 (1996)
22. Jacobson, V.: Congestion avoidance and control. In: SIGCOMM 1988, pp. 314–329 (1988)
23. Allman, M., Paxson, V., Stevens, W.: TCP Congestion Control, RFC 2581 (1999)
24. Henderson, T., Floyd, S., Gurtov, A., Nishida, Y.: The newreno modification to TCP's fast recovery algorithm, RFC 6582 (2012)
25. Stevens, W.R.: TCP/IP Illustrated, Volume 1: The Protocols. Addison Wesley, Reading (1994)
26. Subedi, L., Najiminaini, M., Trajkovi, L.: Performance Evaluation of TCP Tahoe, Reno, Reno with SACK, and New Reno Using OPNET Modeler. Simon Fraser University Vancouver, Vancouver (2004)
27. Anjum, F., Tassiulas, L.: Comparative study of various TCP versions over a wireless link with correlated losses. IEEE/ACM Trans. Netw. **11**(3), 370–383 (2003)
28. Jacobson, V.: Modified TCP congestion avoidance algorithm (1990). ftp://ftp.ee.lbl.gov/email/vanj.90apr30.txt

Performance Evaluation of a Deep Q-Network Based Simulation System for Actor Node Mobility Control in Wireless Sensor and Actor Networks Considering Three-Dimensional Environment

Tetsuya Oda[1(✉)], Donald Elmazi[2], Miralda Cuka[2], Elis Kulla[1], Makoto Ikeda[3], and Leonard Barolli[3]

[1] Department of Information and Computer Engineering, Okayama University of Science (OUS), 1-1 Ridaicho, Kita-ku, Okayama 700–0005, Japan
oda.tetsuya.fit@gmail.com, eliskulla@gmail.com

[2] Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
donald.elmazi@gmail.com, mcuka91@gmail.com

[3] Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
makoto.ikd@acm.org, barolli@fit.ac.jp

Abstract. A Wireless Sensor and Actor Network (WSAN) is a group of wireless devices with the ability to sense physical events (sensors) or/and to perform relatively complicated actions (actors), based on the sensed data shared by sensors. This paper presents design and implementation of a simulation system based on Deep Q-Network (DQN) for actor node mobility control in WSANs. DQN is a deep neural network structure used for estimation of Q-value of the Q-learning method. We implemented the proposed simulating system by Rust programming language. We evaluated the performance of proposed system for normal distribution of events considering three-dimensional environment. For this scenario, the simulation results show that for normal distribution of events and the best episode all actor nodes are connected but one event is not covered.

1 Introduction

Wireless Sensor and Actor Networks (WSANs) have emerged as the-state-of-the-art technology in data gathering from remote locations by interacting with physical phenomena and relying on collaborative efforts by a few resource rich devices and a large number of low cost devices [1]. In WSANs sensors are low cost resource constrained devices that have limited energy, communication and computation capabilities. Once deployed, sensor nodes collect the information

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_4

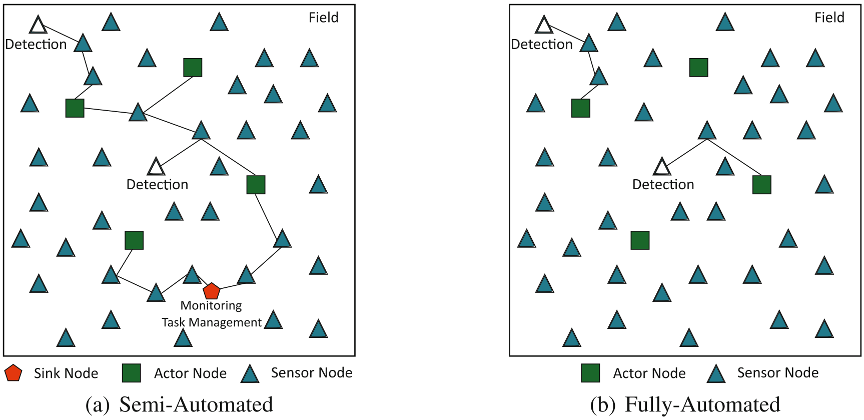


Fig. 1. WSAN architectures.

of interest from their on board sensors, perform local processing of these data including quantization and compression. The actor nodes, on the other hand are resource rich nodes that are equipped with better processing power, higher transmission power, more energy resources and may contain additional capabilities such as mobility.

The concept of a mobile actor node is relatively new in the area of WSAN [2]. The main advantage of having a mobile actor node is that more effective action can be taken, since the actor can get as close as possible to the event location. Additionally, in a densely connected network the presence of a mobile actor eases the problem of overburdened forwarding nodes by balancing the load among all the nodes in the network. In a sparsely connected network, a mobile actor node can bridge the connectivity between groups of isolated nodes. But, there is no guarantee that the existing mobility models will work efficiently for a mobile actor node in a sparsely connected network. An example application is coastal monitoring where sensor nodes may be deployed to monitor water temperature, wave characteristics, water level or meteorological conditions. When an event occurs in this environment, a possible action taken by the actor node could be of collecting a water sample for further analysis.

One of the main advantages of WSANs is the ability to exploit node mobility for various purposes [3]. Several performance metrics including dependability, connectivity, deployment convenience, coverage, energy, and accuracy can be improved by moving and relocating various nodes in these networks [4,5]. Our focus in this paper is connectivity which is crucial in WSANs. In WSANs, the connectivity of actor node is needed to organize collaborative tasks. Mobility helps improving dependability of these networks in several ways. For instance, failed nodes can be replaced with the other spare nodes by moving the spare nodes to their locations. Similarly, if the network is partitioned, mobility can be exploited to restore connectivity by moving one or multiple nodes to the selected locations. Recently, the aforementioned dependability issues have been studied extensively in the context of WSANs [6,7]. The main focus of these works was to deal with individual failures and restore the connectivity of the actor nodes or the connectivity of the WSN with the sink node.

There are many critical issues arising in WSANs. Based on the specific application, different objectives can be taken into account such as energy consumption, throughput, delay, coverage, etc. [8,9]. Also many schemes have been proposed in order to optimize a specific Quality of Service (QoS) parameter. We propose a simulation system based on Deep Q-Network (DQN) as a controller for actor node mobility and implement by Rust programming language [10]. We evaluate the performance of proposed system for normal distribution of events considering three-dimensional environment.

The rest of the paper is organized as follows. In Sect. 2, we describe the basics of WSANs including architectures, research challenges and mobile control for actor node. In Sect. 3, we present the overview of DQN. In Sect. 4, we show the description and design of the simulation system. Simulation results are shown in Sect. 5. Finally, conclusions and future work are given in Sect. 6.

2 WSAN

2.1 WSAN Architecture

The development of smart sensors and the spread of Micro Electro Mechanical Systems (MEMS) have contributed in a major way to the recent expansion of WSANs [11]. WSANs are normally composed of a large number of nodes, which are low power devices generally equipped with one or more sensing units (composed of sensors and, eventually, analog to digital converters), an actuator, a processing unit that includes a limited memory, a radio transceiver, and the power supply.

The main functionality of WSANs is to make actors perform appropriate actions in the environment, based on the data sensed from sensors and actors [12]. When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actors' tasks from distance or transmit their data to actors, which can perform actions independently from the sink node. Here, the former scheme is called Semi-Automated Architecture and the latter one Fully-Automated Architecture, as seen in Fig. 1(b) and (a), respectively. Obviously, both architectures can be used in different applications. Fully-Automated Architecture, which is considered in this paper, emerges the need to develop new sophisticated algorithms, in order to provide appropriate coordination between nodes of WSAN. On the other hand, it has advantages, such as *low latency*, *low energy consumption*, *long network lifetime*, *higher local position accuracy*, *higher reliability* and so on.

2.2 WSAN Challenge

Some of the key challenges in WSAN are related to the presence of actors and their functionalities [12].

Deployment and Positioning: WSAN are heterogeneous networks [13], where actors and sensors have different processing powers, mobility abilities and functionalities. Thus, at the moment of node deployment, algorithms must consider to optimize the number of sensors and actors and their initial positions based on application [14,15].

Architecture: The main functionality of WSANs is to make actors perform appropriate actions in the environment, based on the data sensed from sensors and actors [16]. When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actors' tasks from distance or transmit their data to actors, which can perform actions independently from the sink node.

Real-Time: The purpose of using WSANs in most of the applications is mainly related to their ability to react independently to situations where human intervention is physically difficult or time-restricted [17]. In other words, there are a lot of applications that have strict real-time requirements. In order to fulfill them, real-time limitations must be clearly defined for each application and system.

Coordination: Unlike WSN, where sensors coordinate with each-other to send data to the sink, in WSAN, sensor-actor coordination occurs as well, because all sensed data controls actor's behavior. Also, actor-actor coordination is important in cases when actors collaborate on performing tasks together. In order to provide effective sensing and acting, a distributed local coordination mechanism is necessary among sensors and actors [18].

Power Management: Similar to energy-constrained WSNs, in WSANs sensors have limited power supplies, which limits the network lifetime. Actors have more powerful power supplies but their functionalities are more sophisticated, so they spend more energy when completing complicated tasks. Thus, WSAN protocols should be designed with minimized energy consumption for both sensors and actors [19].

Mobility: In WSANs, nodes, especially actors can be mobile. For example, robots used in industrial monitoring sites or flying drones aver a disaster recovery area [20]. Therefore, protocols developed for WSANs should support the mobility of nodes [21], where dynamic topology changes, unstable routes and network isolations are present.

Self Healing: One of the main problems in mobile Self-Organizing Networks (SON) is the high probability of node isolations during network runtime. An actor failure may lead to partitioning the network and thus hinder the fulfillment

Algorithm 1. Deep Q-learning with Experience Replay

Initialize replay memory D to capacity N
Initialize action-value function Q with random weights
for *episode* $= 1, M$ **do**
  Initialise sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$
  for $t = 1, T$ **do**
    With probability ϵ select a random action a_t
    otherwise select at $= \max_a Q^* (\phi(s_t), a; \theta)$
    Execute action at in emulator and observe reward r_t and image x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
    Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D
    Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D
    Set$y_j = \begin{cases} r_j & \text{for terminal}\phi_{j+1} \\ r_j + \max_a(\phi_{j+1}, a_j; \theta) & \text{for non-terminal}\phi_{j+1} \end{cases}$
    Perform a gradient descent step on $(y_j - Q(\phi_j, a'_j; \theta))^2$ according to Eq. 3
  end for
end for

of the application requirements. Many works have been done on connectivity restoration, by using actors ability to move without using much energy [6]. Actors may also be specialized to carry extra energy supplies, in order to charge sensors or other actors in the network.

Scalability: Smart Cities are emerging fast and WSAN, with its practical functions of simultaneous sensing and acting, are a key technology. The heterogeneity is not limited and most of the systems will continue to grow together with cities. In order to keep the functionality of WSAN applicable, scalability should be considered when designing WSAN protocols and algorithms. Data replication, clustering and so on, can be used in order to support growing networks [15].

3 DQN

The algorithm for training Deep Q-learning is presented in Algorithm 1. The DQN defined the tasks between the agents and the environments [22,23] in Atari 2600 games emulator [24]. Because Q maps history-action pairs to scalar estimates of their Q-value, the history and the action have been used as inputs to the neural network by some previous approaches [25,26]. The environment was set as ξ. At each step, the agent selected an action a_t from the action sets of the game and observed a displayed image x_t from the current screen [27]. The change of the game score r_t was regarded as the reward for the action. For a standard reinforcement learning method, we can complete all of these game sequences s_t as Markov decision process directly, where sequences of actions and observations $s_t = x_1, a_1, x_2, \ldots, a_{t-1}, x_t$. Also, it uses a technique known

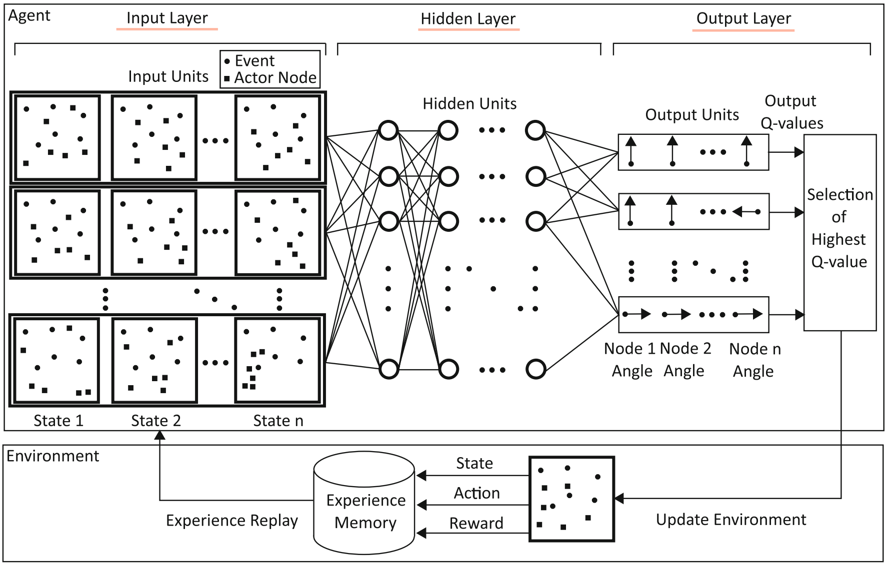


Fig. 2. The structure of DQN based mobile actor node control simulation system.

as experience replay [28] in which it store the agent's experiences at each time-step, $e_t = (s_t, a_t, r_t, s_{t+1})$ in a data-set $D = e_1, \ldots, e_N$, pooled over many episodes into a replay memory. Defining the discounted reward for the future by a factor γ, the sum of the future reward until the end would be $R_t = \sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}$. T means the termination time-step of the game. After performing experience replay, the agent selects and executes an action according to an ϵ-greedy policy. Since using histories of arbitrary length as inputs to a neural network can be difficult, Q-function of DQN instead works on fixed length representation of histories produced by a function ϕ. The target was to maximize the action-value function $Q^*(s, a) = max_\pi E[R_t | s_t = s, a_t = a, \pi]$, where π is the strategy for choosing of best action. From the Bellman equation, it is equal to maximize the expected value of $r + \gamma Q^*(s', a')$, if the optimal value $Q^*(s', a')$ of the sequence at the next time step is known.

$$Q^*(s', a') = E_{s' \sim \xi}[r + \gamma \underset{a'}{max} Q^*(s', a') | s, a] \tag{1}$$

Not using iterative updating method to optimal the equation, it is common to estimate the equation by using a function approximator. Q-network in DQN was such a neural network function approximator with weights θ and $Q(s, a; \theta) \approx Q^*(s, a)$. The loss function to train the Q-network is:

$$L_i(\theta_i) = E_{s,a \sim \rho(.)}[(y_i - Q(s, a; \theta_i))^2]. \tag{2}$$

The y_i is the target, which is calculated by the previous iteration result θ_{i-1}. $\rho(s,a)$ is the probability distribution of sequences s and a. The gradient of the loss function is shown in Eq. (3):

$$\nabla_{\theta_i} L_i(\theta_i) = E_{s,a\sim\rho(.);s'\sim\xi}[(y_i - Q(s,a;\theta_i))\nabla_{\theta_i} Q(s,a;\theta_i)]. \tag{3}$$

4 Design and Implementation of Proposed Simulation System

In this section, we present design and implementation of proposed simulation system based on DQN for mobile actor node control in WSANs. The simulation system structure is shown in Fig. 2. The proposed simulating system is implemented by Rust programming language [29,30]. Rust is a system programming language focused on three goals: safety, speed, and concurrency [31]. Rust supports a mixture of programming styles: imperative procedural, concurrent actor, object-oriented and functional.

We consider tasks in which an agent interacts with an environment. In this case, the mobile actor node moves step by step in a sequence of actions, observations and rewards. We took in consideration the mobility and connectivity of actor nodes.

For a mobile actor node are considered 7 patterns (forward, back, left, right, up, down, stop). The actor nodes have mobility, networking, sensing and actuation mechanisms. In order to decide the reward function, we considered the Number of Connected Actor Nodes (NCAN) and Number of Sensing Events (NSE) parameters. The reward function r is defined as follows:

$$r = \begin{cases} 5 \times (NCAN + NSE) & (if\ NCAN \geq 1) \\ -5 \times (NCAN + NSE) & (if\ NCAN = 0). \end{cases} \tag{4}$$

The initial weights values are considered as Normal Initialization [32]. The input layer is using the position of events and actor nodes, mobile actor node patterns and total reward values in Experience Memory. The hidden layer is connected with 256 rectifier units in Rectified Linear Units (ReLU) [33]. The output Q-values are mobile actor node patterns.

5 Simulation Results

The simulation parameters are shown in Table 1. The simulations are done for 4 actor nodes and 24 events for normal distribution. One episode has 200 iterations.

In Table 2, we show the simulation results of total reward of episodes. There are best, median and worst episodes. In Fig. 3, we show the simulation results

Table 1. Simulation parameters.

Parameters	Values
Field size	32 × 32
Types of actor node mobile	Forward, back, left, right, up, down, stop
Number of episode	30000
Number of iteration	200
Number of hidden layers	3
Number of hidden units	15
Initial weight value	Normal Initialization
Activation function	ReLU
Action selection probability (ϵ)	0.999 - (t/Number of episode) (t = 0, 1, 2, ..., Number of episode)
Learning rate (α)	0.04
Discount rate (γ)	0.9
Experience memory size	300 × 100
Batch size	32
Positions of events	Normal distribution
Initial positions of actor nodes	Center
Number of events	24
Number of actor nodes	4

Table 2. Results of total reward.

Episode	Total reward
Best episode	10600
Median episode	4890
Worst episode	−2440

of reward vs. number of iteration. For normal dirstribution of events, the higher reward value means that the actor node can move and keep connection with other actor nodes. In Fig. 4 is shown visualization interface of implemented simulation system for normal distribution considering three-dimensional environment. For this scenario, the simulation results show that for normal distribution of events and the best episode all actor nodes are connected but one event is not covered.

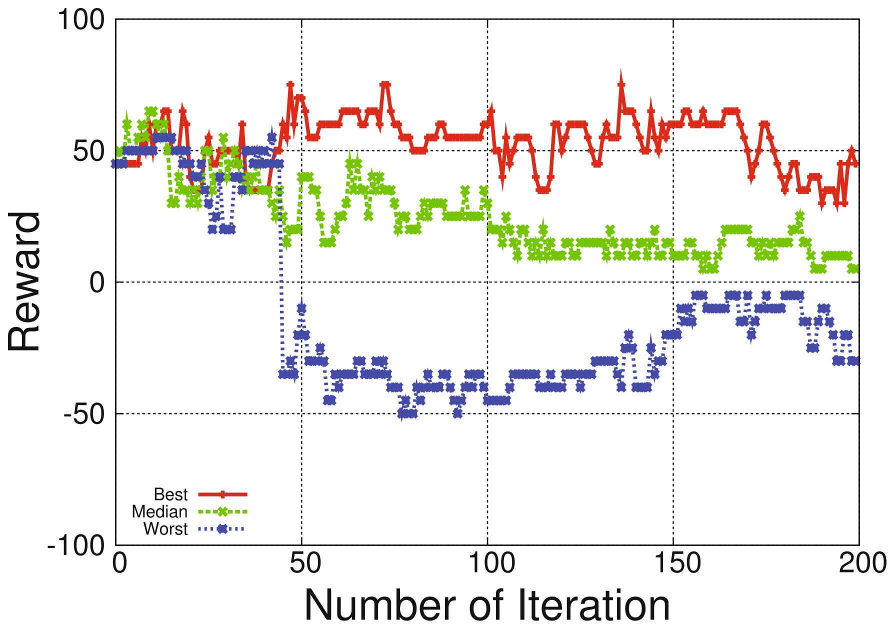


Fig. 3. Results of reward.

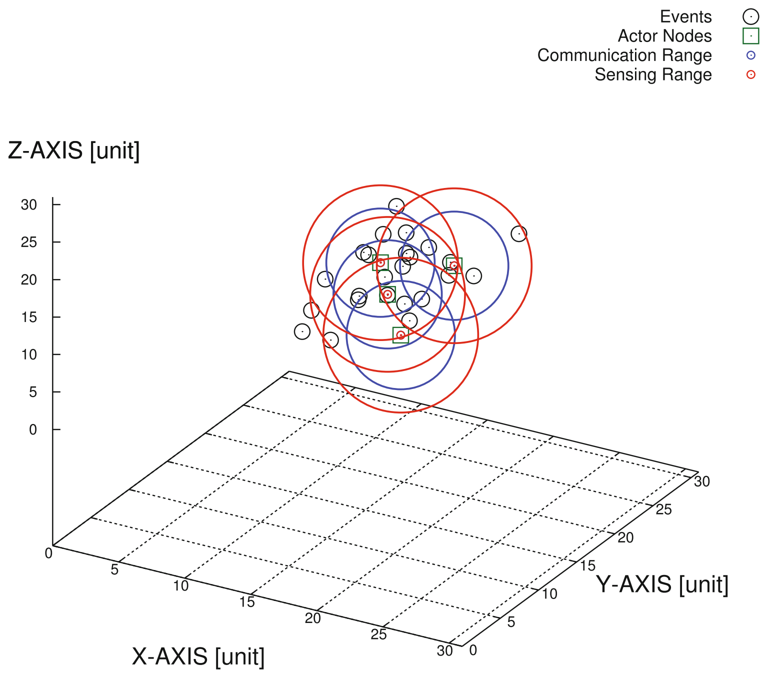


Fig. 4. Visualization interface.

6 Conclusions

In this work, we implemented a simulation system based on DQN for mobile actor node control in WSANs. We presented the implementation of the proposed simulation system and have shown also the interface in a simulation scenario for normal distribution considering three-dimensional environment. For this scenario, the simulation results show that for normal distribution of events and the best episode all actor nodes are connected but one event is not covered.

In the future, we would like to make extensive simulations for different simulation scenarios and improve the simulation system.

References

1. Akyildiz, I.F., Kasimoglu, I.H.: Wireless sensor and actor networks: research challenges. Ad Hoc Netw. **2**(4), 351–367 (2004)
2. Krishnakumar, S.S., Abler, R.T.: Intelligent actor mobility in wireless sensor and actor networks. In: IFIP WG 6.8 1-st International Conference on Wireless Sensor and Actor Networks (WSAN-2007), pp. 13–22 (2007)
3. Sir, M.Y., Senturk, I.F., Sisikoglu, E., Akkaya, K.: An optimization-based approach for connecting partitioned mobile sensor/actuator networks. In: IEEE Conference on Computer Communications Workshops, pp. 525–530 (2011)
4. Younis, M., Akkaya, K.: Strategies and techniques for node placement in wireless sensor networks: a survey. Ad-Hoc Netw. **6**(4), 621–655 (2008)
5. Liu, H., Chu, X., Leung, Y.-W., Du, R.: Simple movement control algorithm for bi-connectivity in robotic sensor networks. IEEE J. Sel. Areas Commun. **28**(7), 994–1005 (2010)
6. Abbasi, A., Younis, M., Akkaya, K.: Movement-assisted connectivity restoration in wireless sensor and actor networks. IEEE Trans. Parallel Distrib. Syst. **20**(9), 1366–1379 (2009)
7. Akkaya, K., Senel, F., Thimmapuram, A., Uludag, S.: Distributed recovery from network partitioning in movable sensor/actor networks via controlled mobility. IEEE Trans. Comput. **59**(2), 258–271 (2010)
8. Costanzo, C., Loscri, V., Natalizio, E., Razafindralambo, T.: Nodes self-deployment for coverage maximization in mobile robot networks using an evolving neural network. Comput. Commun. **35**(9), 1047–1055 (2012)
9. Li, Y., Li, H., Wang, Y.: Neural-based control of a mobile robot: a test model for merging biological intelligence into mechanical system. In: IEEE 7-th Joint International Information Technology and Artificial Intelligence Conference (ITAIC-2014), pp. 186–190 (2014)
10. Oda, T., Obukata, R., Ikeda, M., Barolli, L., Takizawa, M.: Design and implementation of a simulation system based on deep Q-network for mobile actor node control in wireless sensor and actor networks. In: The 31-st IEEE International Conference on Advanced Information Networking and Applications Workshops (IEEE WAINA-2017) (2017)
11. Llaria, A., Terrasson, G., Curea, O., Jiménez, J.: Application of wireless sensor and actuator networks to achieve intelligent microgrids: a promising approach towards a global smart grid deployment. Appl. Sci. **6**(3), 61–71 (2016)

12. Kulla, E., Oda, T., Ikeda, M., Barolli, L.: SAMI: a sensor actor network matlab implementation. In: The 18-th International Conference on Network-Based Information Systems (NBiS-2015), pp. 554–560 (2015)
13. Kruger, J., Polajnar, D., Polajnar, J.: An open simulator architecture for heterogeneous self-organizing networks. In: Canadian Conference on Electrical and Computer Engineering (CCECE-2006), pp. 754–757 (2006)
14. Akbas, M., Turgut, D.: APAWSAN: actor positioning for aerial wireless sensor and actor networks. In: IEEE 36-th Conference on Local Computer Networks (LCN-2011), pp. 563–570 (2011)
15. Akbas, M., Brust, M., Turgut, D.: Local positioning for environmental monitoring in wireless sensor and actor networks. In: IEEE 35-th Conference on Local Computer Networks (LCN-2010), pp. 806–813 (2010)
16. Melodia, T., Pompili, D., Gungor, V., Akyildiz, I.: Communication and coordination in wireless sensor and actor networks. IEEE Trans. Mob. Comput. **6**(10), 1126–1129 (2007)
17. Gungor, V., Akan, O., Akyildiz, I.: A real-time and reliable transport (RT2) protocol for wireless sensor and actor networks. IEEE/ACM Trans. Netw. **16**(2), 359–370 (2008)
18. Mo, L., Xu, B.: Node coordination mechanism based on distributed estimation and control in wireless sensor and actuator networks. J. Control Theor. Appl. **11**(4), 570–578 (2013)
19. Selvaradjou, K., Handigol, N., Franklin, A., Murthy, C.: Energy-efficient directional routing between partitioned actors in wireless sensor and actor networks. IET Commun. **4**(1), 102–115 (2010)
20. Kantaros, Y., Zavlanos, M.M.: Communication-aware coverage control for robotic sensor networks. In: IEEE Conference on Decision and Control, pp. 6863–6868 (2014)
21. Melodia, T., Pompili, D., Akyldiz, I.: Handling mobility in wireless sensor and actor networks. IEEE Trans. Mob. Comput. **9**(2), 160–173 (2010)
22. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518**, 529–533 (2015)
23. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing Atari with Deep Reinforcement Learning, arXiv:1312.5602v1, pp. 1–9 (2013)
24. Lei, T., Ming, L.: A robot exploration strategy based on Q-learning network. In: IEEE International Conference on Real-time Computing and Robotics (RCAR-2016), pp. 57–62 (2016)
25. Riedmiller, M.: Neural fitted Q iteration – first experiences with a data efficient neural reinforcement learning method. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) ECML 2005. LNCS, vol. 3720, pp. 317–328. Springer, Heidelberg (2005). doi:10.1007/11564096_32
26. Lange, S., Riedmiller, M.: Deep auto-encoder neural networks in reinforcement learning. In: The 2010 International Joint Conference on Neural Networks (IJCNN-2010), pp. 1–8 (2010)
27. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artif. Intell. **101**(1–2), 99–134 (1998)
28. Lin, L.J.: Reinforcement Learning for Robots Using Neural Networks. Technical report, DTIC Document (1993)

29. The Rust Programming Language (2007). https://www.rust-lang.org/
30. GitHub - rust-lang/rust: a safe, concurrent, practical language (2007). https://github.com/rust-lang/
31. rust tag wiki - Stack Overflow (2007). http://stackoverflow.com/tags/rust/info/
32. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: The 13-th International Conference on Artificial Intelligence and Statistics (AISTATS-2010), pp. 249–256 (2010)
33. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: The 14-th International Conference on Artificial Intelligence and Statistics (AISTATS-2011), pp. 315–323 (2011)

Space-Filling Curves based on Residue Number System

Jan Platoš(✉), Jana Nowaková, Pavel Krömer, and Václav Snášel

Department of Computer Science, FEECS, VŠB-Technical University of Ostrava, 17. listopadu 15/2172, 708 00 Ostrava-Poruba, Czech Republic
{jan.platos,jana.nowakova,pavel.kromer,vaclav.snasel}@vsb.cz

Abstract. Space-filling curves are a useful tool for fast multidimensional space indexing, dimension reduction, and fast optimization of complex problems. Several curves such as Hilbert, Peano, Gray, Morton or Z-order were discovered, and their properties and features were intensely studied. In this paper, a new space-filling curve is described, and its features are analyzed and compared with the other space-filling curves. The novel algorithm for a space-filling curve is based on the Residue Number System that is extensively studied during the last thirty years. The proposed curve has specific behavior and may be controlled by several parameters.

1 Introduction

Space-filling curves (SFC) are interesting and useful tools, which can serve for space indexing, dimension reduction or fast optimization of complex problems. As a typical problem, which can be solved using SFC can be considered a traveling salesman problem to find the shortest path through a series of nodes [15] or neighbor-finding [5].

In 1968, a conjunction of space-filling curves and mathematical programming was described in [3]. There were some limits in comparison with the convex programming techniques such as smaller tolerance of errors.

During the next years, many space-filling curves approaches was designed, such as Hilbert's [4] space-filling curves with quantification of the locality and locality boundary of multidimensional SFC [6] with four alternative patterns [12]. The comparison of SFC used for mapping the multi-dimensional space into the one-dimensional space was presented by Mokbel et al. [14].

It was proposed the usage of SFC in such areas as compressing the bandwidth of waveforms [2]. SFC can also be implemented for a location of basis function centers in radial basis function networks serving for an evolutionary neural network training algorithm [19]. The 6-D SFC is proposed for the transformation of six number parameters on the unit circle [17]. SFC are also suggested for a design of geometric data structures [1], easy representation of two-dimensional space [8], for genetic string generation [10] or for path planning of cooperative

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_5

mobile robots [21] and robotic data collection [20]. Very interesting approach, proposed by Wang et al. [18], is based on an expert system for SFC creating. Hilbert SFC can also be determined for lossless compression of the medical images [11]. Unfortunately, there does not exist face-continuous SFC constructed by reptilings of acute triangles [7]. The idea based on Morton SFC is chosen for crystal structure modelling used in solving of chemical problems [9].

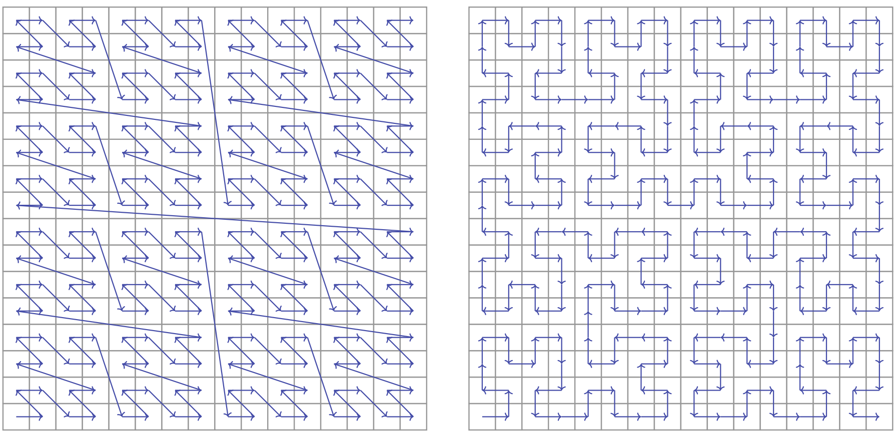
The paper is organized as follows. The following section describes the necessary background theory about Residue Number System and space-filling curves. Moreover, the section contains a definition of the novel SFC. The last section discusses the pros and cons of the novel approach.

2 Space-Filling Curves Based on Residue Number System

The main benefit of this paper is the definition of the new type of SFC that is based on the special number representation called Residue Number System (RNS). Both areas have to be defined before algorithm for the RNS-based curves is going to be suggested.

2.1 Space-Filling Curves

Space-filling curves are special types of curves that are able to cover a constrained space. Some curves are designed for two-dimensional spaces, but others are generalizable to three or even more dimensions. The most popular curves Z-curve and Hilbert curve are depicted in Fig. 1. Both of them can be generalized to higher dimensions. The usage of the SFC was described above.



(a) Z-curve (b) Hilbert curve

Fig. 1. Examples of two space-filling curves.

2.2 Residue Number Systems

Residue number systems [13,16] provide means for an alternative representation of integers that do not rely on the correspondence between sequences of digits and numbers in the usual sense [16]. RNSs is based on the congruence relation and represent numbers by sets of remainders (residues) after integer division by a fixed set of divisors (moduli). Suppose q and r are the quotient and remainder of integer division and $a = q \cdot m + r$. Then r is the residue of a with respect to the modulus m and we write $r = |a|_m$. The RNS is a number representation and arithmetic system defined by a set of two or more moduli [13,16].

Definition 1 (Residue number system). *Consider a set of positive integers (moduli),* $\mathbf{M} = \{m_1, m_2, \ldots, m_N\}$*, for which it holds that* $\forall j, k \in N : j \neq k \Rightarrow m_j$ *and* m_k *have no common divisor greater than 1. Every two moduli in* $\mathbf{M}$ *are relatively prime (co–prime) to each other and any number,* x*, can be represented by a set of residues,* $\mathbf{r} = \{|x|_{m_i} : 1 \leq i \leq N\}$*. Let* $\mathcal{DR} = \prod_{i \in N}(m_i)$ *be the dynamic range of the RNS. Every* $x < \mathcal{DR}$ *has an unique representation in the RNS defined by* $\mathbf{M}$*. The RNS can represent positive integers from the range* $[0, \mathcal{DR} - 1]$*.*

We adopt the notation from [16] and write

$$x \cong \langle x_1, x_2, \ldots, x_N \rangle \tag{1}$$

to indicate that x *is a number uniquely represented by the residue set* $\{|x|_{m_1}, \ldots, |x|_{m_1}\}$ *in the RNS defined by the moduli set* $\{m_1, m_2, \ldots m_N\}$*.*

Basic arithmetic operations are in RNSs defined on residue sets [13,16]. Addition of two numbers, $x \cong \langle x_1, x_2, \ldots, x_N \rangle$ and $y \cong \langle y_1, y_2, \ldots, y_N \rangle$, is in a RNS defined by

$$\begin{aligned} x + y &\cong \langle x_1, x_2, \ldots, x_N \rangle + \langle y_1, y_2, \ldots, y_N \rangle \\ &= \langle |x_1 + y_1|_{m_1}, |x_2 + y_2|_{m_2}, \ldots, |x_N + y_N|_{m_N} \rangle. \end{aligned} \tag{2}$$

Multiplication of x and y is defined in an analogous way

$$\begin{aligned} x \times y &\cong \langle x_1, x_2, \ldots, x_N \rangle \times \langle y_1, y_2, \ldots, y_N \rangle \\ &= \langle |x_1 \times y_1|_{m_1}, |x_2 \times y_2|_{m_2}, \ldots, |x_N \times y_N|_{m_N} \rangle. \end{aligned} \tag{3}$$

In the simplest case, RNSs work with unsigned non–negative integers. However, they can be easily extended on negative integers and subtraction operation can be defined. The realization of other arithmetic operations such as comparison and division is in general problematic in RNSs [16].

Although RNSs are generally more complex than traditional number systems, the definitions of addition and multiplication clearly illustrate why they are attractive for an efficient realization of fast arithmetics. The computations with potentially large numbers are split into N less complex independent tasks in a divide–and–conquer manner [13] and can be executed concurrently. They are performed over individual residues that have significantly smaller range than the RNS as a whole. Because of that and because there is no carry–propagation for addition and multiplication in RNSs, the operations can be realized in hardware fast [13] and at low–power costs [16].

The conversions from traditional to residue number systems and back are associated with a computational overhead. However, they are required only when the operations of an algorithm 'switch' to and from a RNS and efficient conversion methods are known for RNSs with specific moduli sets [16]. Applications that involve a large number of consecutive addition or multiplication operations can benefit from RNSs and achieve a significant speedup by residue arithmetic. RNSs have been used in digital signal processing to implement finite and infinite impulse response filters, table–lookup–based filters, discrete Fourier transform, error–detection [16], digital to analog converters, and e.g. for digital frequency synthesis [13].

2.3 Algorithm for RNS-Based Space-Filling Curves

The RNS system is a solid system for number representation, that may be used for acceleration of the computation using parallelization. Moreover, the RNS has a very specific behavior that may be used in many special cases. In this paper, we developed a new type of space-filling curve that is based on the residue arithmetics. More specific, the algorithm is based on the conversion from the RNS into traditional number system.

The definition of the RNS guarantees that any assignment into x that fulfill the constrains defined above may be converted into traditional system. The nature of the RNS is that the final conversion uses a lot of multiplication the inverse coefficients and the final modulo operation [16]. The consequence of this

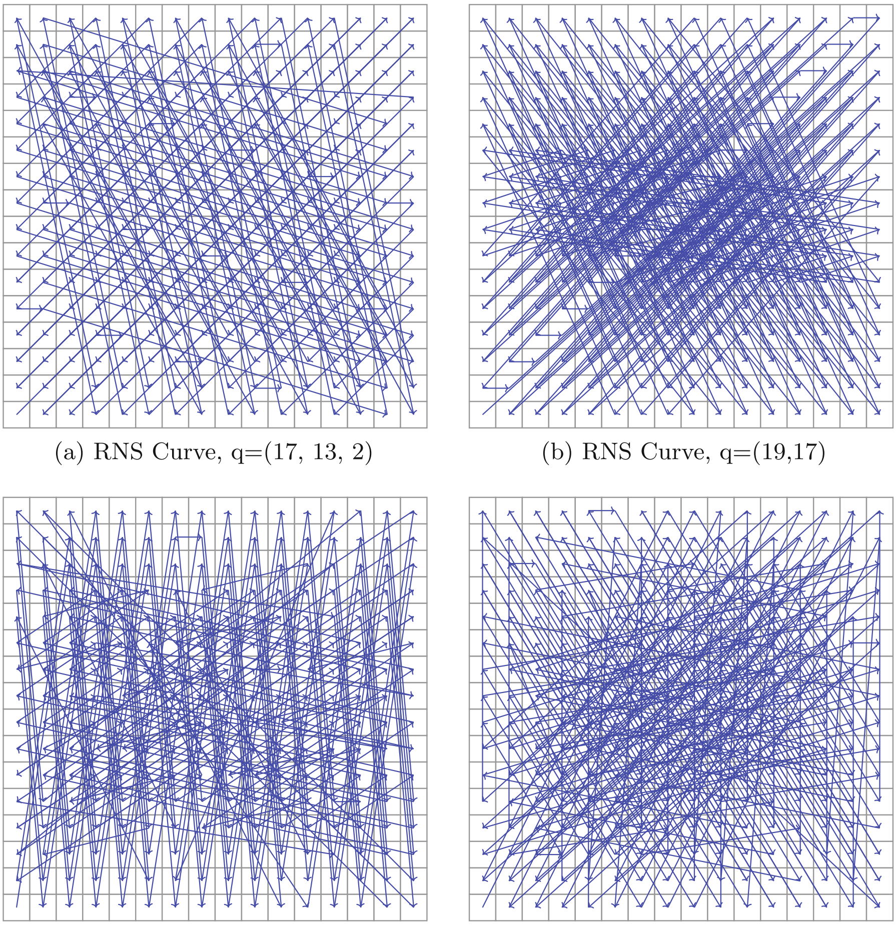
Algorithm 1. RNS-Curve(Size: N).

begin
    Find the moduli set M such that $\sum M_i \geq N^2$, where $|M|$ is the size of set M;
    Define an array R such that $|R| = |M|$ and $\forall i, i < |M| : R_i = 0$;
    do
        Increment the first residue $R[0]$ by one;
        When $R_0 = M_0$ set $R_0 = 0$ and increment R_1;
        Continue with incrementation until a valid residue representation is found;
        Compute $index = Inverse(R)$;
        Compute coordinates $(x, y) = (index \mod N, index/N)$;
        Append point (x, y) to the curve;
    while *all $\mathcal{DR}$ numbers were tested*;
    return *the coverage of the defined area by the curve;*
end

complexity of the reverse conversion is that when x and y are two different RNS numbers that differ in only residue by one than the inverse conversion of x and y differs significantly.

This behavior is the main idea of the space-filling curves generation using the RNS. The algorithm for the creating of the curve that covers a square area of size N is depicted in Algorithm 1.

As it may be seen, the pseudocode in the Algorithm 1 does not take into account rectangular areas, but the RNS curve may also cover non-squared areas. The line no. 9, where the coordinates are computed, may lead into coordinate outside of the computation area, but such point may be discarded. The set of moduli M is not unique, and many different sets may be found and used for the



(a) RNS Curve, q=(17, 13, 2)

(b) RNS Curve, q=(19,17)

(c) RNS Curve, q=(5, 7, 11)

(d) RNS Curve, q=(11, 7, 5)

Fig. 2. Four RNS-based space-filling curves with different quotients.

same size of the area. The example of four different curves that covers an area of 16×16 points is depicted in Fig. 2, where the numbers in the brackets represent the moduli set.

As it can be seen from the Fig. 2, the behavior of the curves or the position of the consecutive points is different than for the standard curves. The cases (a) and (b) are highly regular. The case (d) is almost chaotic. All cases except the example (a) show that the are of the center is denser than the area around borders. This is really unusual behavior that does not follow the usual layout of the space-filling curves. This behavior, where the following points are not necessarily close to the previous may be used for neighborhood search that does not stuck too close to actual one, e.g. in the optimization algorithms.

2.4 Comparison of the Space-Filling Curves Properties

The space-filling curves that are used in real applications satisfy a locality property. It means that the near points in the original space are mapped onto near point in the curve. In the other words, when a small neighborhood is taken, e.g. classical 8-neighborhood of a point, then this neighborhood should be placed in a relatively close area in the curve. The results of such mapping are depicted in Fig. 3 which shows the cumulative frequency of average distance from the source point to the point in the neighborhood.

As it may be seen, the Z-curve and Hilbert curve are able to concentrate all points very close to each other, because about 50% of points is located closer than 10. On the contrary, the RNS-based curves spread the points from the neighborhood much farther. For the first two curves, more than 50% points are in the average on the same distance from the selected point.

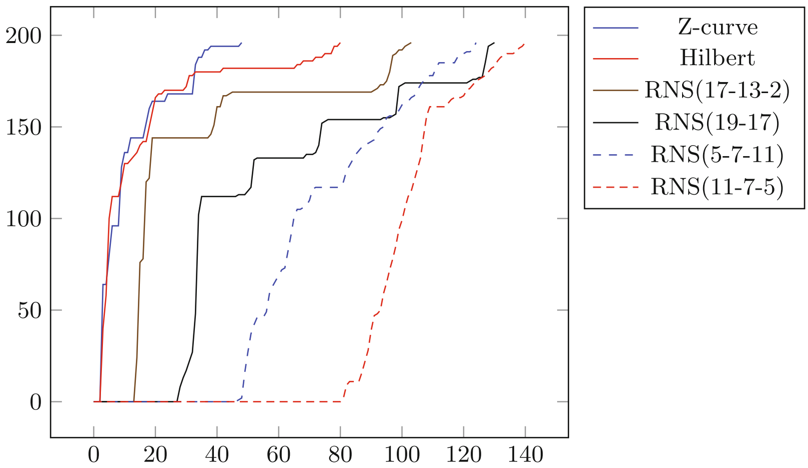


Fig. 3. A cumulative frequency of average distances in the 8-neighborhood for a 16×16 space.

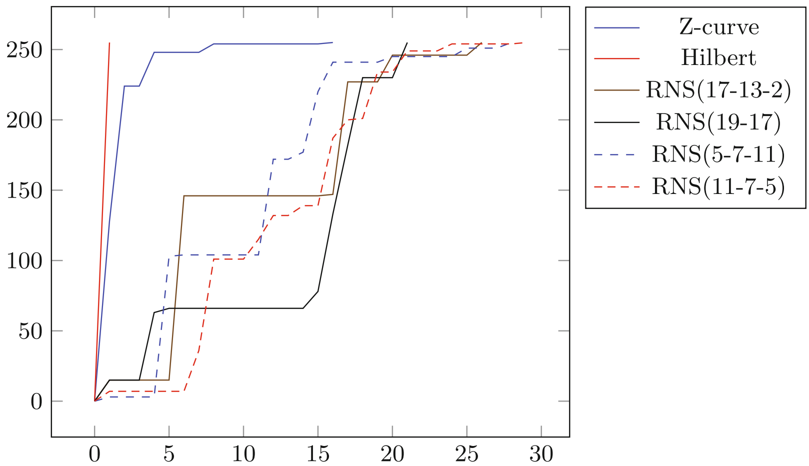


Fig. 4. A cumulative frequency of distances along the curves (mapping from index into 2D space of size 16×16).

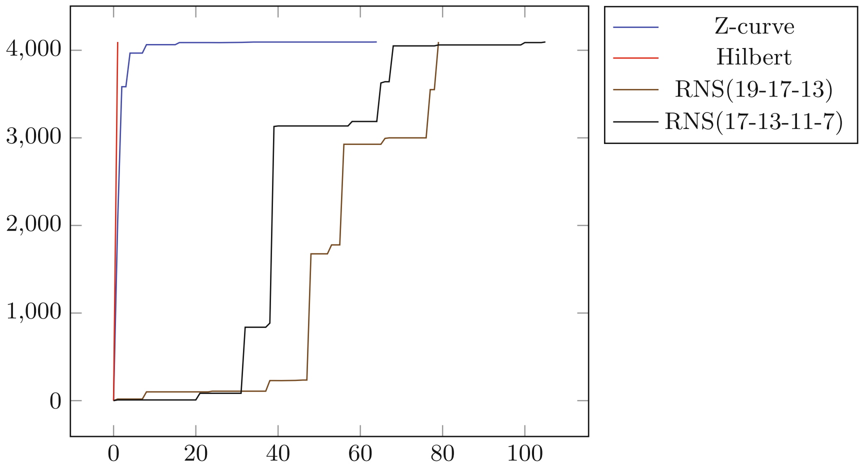


Fig. 5. A cumulative frequency of distances along the curves (mapping from index into 2D space of size 64×64).

The second histogram depicts the inverse behavior because RNS curves show how far the following indices on the curve are placed in the 2D space. The cumulative frequency of the distances between the following indices is depicted in Fig. 4. This distance is measured using well-known Manhattan measure to maintain integer values. The similar plot for larger space is depicted in Fig. 5.

As it can be seen, the distances along the curve are in the case of RNS much bigger, and, moreover, it may increases with the size of the covered space (compare Figs. 4 and 5). This behavior is unusual and it is opposite to the standard behavior of the boths, Z- and Hilbert curves.

3 Conclusion

In this paper, a novel approach for space-filling curves generation based on the Residue Number System is introduced. The presented novel curves have a special behavior which can be customized by the selection of the Moduli set. The coverage of the defined region may look chaotic or systematic according to the used moduli set. The main difference of the proposed space-filling curves and the standard curves such as Hilbert's or Z-curves is that the locality is not satisfied, when the mapping is applied. This property may be used in many areas, such as optimization algorithms. Even when the behavior may look chaotic and or random, it is deterministic and replicable. Everything is controlled by the moduli set that is used for curve generation.

Acknowledgment. This work was supported by the Czech Science Foundation under the grant no. GJ16-25694Y and by the projects SP2017/100 "Parallel processing of Big Data IV"and SP2017/85 "Processing and advanced analysis of bio-medical data II", of the Student Grant System, VŠB-Technical University of Ostrava.

References

1. Asano, T., Ranjan, D., Roos, T., Welzl, E., Widmayer, P.: Space-filling curves and their use in the design of geometric data structures. Theoret. Comput. Sci. **181**(1), 3–15 (1997)
2. Bially, T.: Space-filling curves: their generation and their application to bandwidth reduction. IEEE Trans. Inf. Theor. **15**(6), 658–664 (1969)
3. Butz, A.R.: Space filling curves and mathematical programming. Inf. Control **12**(4), 314–330 (1968)
4. Butz, A.R.: Alternative algorithm for Hilbert's space-filling curve. IEEE Trans. Comput. **100**(4), 424–426 (1971)
5. Chen, H.L., Chang, Y.I.: Neighbor-finding based on space-filling curves. Inf. Syst. **30**(3), 205–226 (2005)
6. Gotsman, C., Lindenbaum, M.: On the metric properties of discrete space-filling curves. IEEE Trans. Image Process. **5**(5), 794–797 (1996)
7. Gottschau, M., Haverkort, H., Matzke, K.: Reptilings and space-filling curves for acute triangles. arXiv preprint arXiv:1603.01382 (2016)
8. Jagadish, H.V.: Analysis of the Hilbert curve for representing two-dimensional space. Inf. Process. Lett. **62**(1), 17–22 (1997)
9. Jasrasaria, D., Pyzer-Knapp, E.O., Rappoport, D., Aspuru-Guzik, A.: Space-filling curves as a novel crystal structure representation for machine learning models. arXiv preprint arXiv:1608.05747 (2016)
10. Kochhar, J.S., Foster, B.T., Heragu, S.S.: Hope: a genetic algorithm for the unequal area facility layout problem. Comput. Oper. Res. **25**(7), 583–594 (1998)
11. Liang, J.Y., Chen, C.S., Huang, C.H., Liu, L.: Lossless compression of medical images using Hilbert space-filling curves. Comput. Med. Imaging Graph. **32**(3), 174–182 (2008)

12. Liu, X.: Four alternative patterns of the Hilbert curve. Appl. Math. Comput. **147**(3), 741–752 (2004)
13. Ananda Mohan, P.V.: Residue Number Systems: Theory and Applications. Springer, Cham (2016)
14. Mokbel, M.F., Aref, W.G., Kamel, I.: Performance of multi-dimensional space-filling curves. In: Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems, pp. 149–154. ACM (2002)
15. Norman, M.G., Moscato, P.: The Euclidean traveling salesman problem and a space-filling curve. Chaos Solitons Fractals **6**, 389–397 (1995)
16. Omondi, A., Premkumar, B.: Residue Number Systems: Theory and Implementation. Advances in Computer Science and Engineering: Texts, Imperial College Press, London (2007)
17. Pan, C.H., Liu, S.Y.: A comparative study of order batching algorithms. Omega **23**(6), 691–700 (1995)
18. Wang, M.J., Hu, M.H., Ku, M.Y.: A solution to the unequal area facilities layout problem by genetic algorithm. Comput. Ind. **56**(2), 207–220 (2005)
19. Whitehead, B.A., Choate, T.D.: Evolving space-filling curves to distribute radial basis functions over an input space. IEEE Trans. Neural Netw. **5**(1), 15–23 (1994)
20. Yan, Y., Mostofi, Y.: Efficient clustering and path planning strategies for robotic data collection using space-filling curves. IEEE Trans. Contr. Netw. Syst. (99) (2016). http://ieeexplore.ieee.org/document/7480435/
21. Yu, Z., Jinhai, L., Guochang, G., Rubo, Z., Haiyan, Y.: An implementation of evolutionary computation for path planning of cooperative mobile robots. In: Proceedings of the 4th World Congress on Intelligent Control and Automation, 2002, vol. 3, pp. 1798–1802. IEEE (2002)

Performance Evaluation of an IoT-Based E-Learning Testbed Using Mean-Shift Clustering Approach Considering Theta Type of Brain Waves

Masafumi Yamada[1(✉)], Miralda Cuka[1], Yi Liu[1], Tetsuya Oda[2], Keita Matsuo[3], and Leonard Barolli[3]

[1] Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
masafumi00835563@gmail.com, mcuka91@gmail.com, ryuui1010@gmail.com

[2] Department of Information and Computer Engineering, Okayama University of Science (OUS), 1-1 Ridaicho, Kita-ku, Okayama 700-0005, Japan
oda.tetsuya.fit@gmail.com

[3] Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
{kt-matsuo,barolli}@fit.ac.jp

Abstract. Due to the opportunities provided by the Internet, people are taking advantage of e-learning courses and enormous research efforts have been dedicated to the development of e-learning systems. So far, many e-learning systems are proposed and used practically. However, in these systems the e-learning completion rate is low. One of the reasons is the low study desire and motivation. In this work, we present an IoT-Based E-Learning testbed using Raspberry Pi mounted on Raspbian. We carried out some experiments with a student of our laboratory for theta type of brain waves. We used MindWave Mobile (MWM) to get the data and considered four situations: sleeping, relaxing, active and moving. Then, we used mean-shift clustering algorithm to cluster the data. The evaluation results show that our tesbed can judge the human situation by using theta waves.

Keywords: Internet of Things · Testbed · Raspberry Pi · Raspbian · MindWave Mobile · Mean-shift · Theta type

1 Introduction

The Internet is growing every day and the performance of computers is significantly increased [1,2]. Also, with appearance of new technologies such as ad-hoc networks, sensor networks, body networks, home networking, new network devices and application are appearing. Therefore, it is very important to monitor and control the network devices via communication channels and exploit their

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_6

capabilities for the everyday real life activities. However, in large scale networks such as Internet, it is very difficult to control the network devices.

So for many e-learning systems are proposed and used practically. In [3], the authors presents a work-in-progress intending to enhance the learning experience for distance university students enrolled at the Open University of Catalonia (UOC). The UOC virtual campus has an integrated e-learning environment that allows students to pursue their studies completely online with the exception of taking final exams. By integrating the technologies of the IoT, they want to expand the learning environment and add a new learning place to the one existing on the computer. The authors hope to combine both the virtual and the physical environments in order to provide a better learning experience to their students. The authors consider two applications types: one related to the learning process and learning materials, the other related to creating a university community as well as fighting dropout and loneliness.

In [4], the authors present the context-aware and culture-oriented aspects of an adaptability approach called Adapt-SUR. Adapt-SUR is an international joint project between Argentina and Brazil. The approach is designed to be integrated into two distinct E-learning environments (ELEs): the AdaptWeb (Adaptive Web based learning Environment) system [5] and the eTeacher+SAVER (Software de Asistencia Virtual para Educacion Remota) environment [6]. This study describes the main features of the context-aware and culture-oriented aspects of a student profile and shows how to organize this contextual information in a multidimensional space where each dimension is represented by a different ontology, which may be handled separately or jointly. Finally the authors use some examples to discuss and illustrate how to use cultural information to provide context-based e-learning personalization.

In this work, we present an IoT-based e-learning testbed using Raspberry Pi. We carried out some experiments with a student of our laboratory for theta type of brain waves. We used MWM to get the data and considered four situations: sleeping, relaxing, active and moving. Then, we used mean-shift clustering algorithm to cluster the data.

The paper is organized as follows. In Sect. 2, we explain the overview of IoT and ULE. In Sect. 3, we present an overview of mean-shift clustering algorithm. In Sect. 4, we present the testbed. In Sect. 5, we discuss the experiments and simulation results. Finally, conclusions and future work are given in Sect. 6.

2 Overview of IoT and ULE

2.1 Internet of Things (IoT)

The Internet of Things (IoT) is a recent communication paradigm that envisions a near future, in which the objects of everyday life will be equipped with microcontrollers, transceivers for digital communication, and suitable protocol stacks that will make them able to communicate with one another and with the users, becoming an integral part of the Internet [7,8]. The IoT concept aims at making the Internet even more immersive and pervasive. Furthermore, by enabling easy access and interaction with a wide variety of devices such as, for instance, home appliances, surveillance cameras, monitoring sensors, actuators, displays, vehicles, and so on, the IoT will foster the development of a number of applications that make use of the potentially enormous amount and variety of data generated by such objects to provide new services to citizens, companies, and public administrations. This paradigm indeed finds application in many different domains, such as home automation, industrial automation, medical aids, mobile healthcare, elderly assistance, intelligent energy management and smart grids, automotive, traffic management, and many others [9].

2.2 Ubiquitous Learning Environment (ULE)

Ubiquitous learning is a seamless learning whenever it is in information space or in physics space, through ubiquitous computing information space and physics space are converged. In ULE Learning, learning demands and learning resources are everywhere; study, life and work are connected each other. When learners meet any practice problem ubiquitous computing help them to resolve it at anytime, anywhere. In the future, school, library, classroom, meeting room, museum, and the circulation fields send their information and knowledge to the learner through all kinds of technology, every learner immerse into information ecology surroundings that the real world and digital world intermingle. The learners can easily perception and obtaining learning objects detailed information and content through situational perception of mobile devices. Using dialogue, living community, cooperation studies, social process of internalization, participate in joint activity to realize social learning. An effective ubiquitous learning depends on founding of learning environment.

2.3 Role of IoT in ULE

According to learning environment classification, ubiquitous learning environment belong to a kind of learning environment that are deeper, and the highest flexibility. While the basic elements of constructing the learning environment mainly include three parts: ubiquitous communication network, learning terminal device, learning resources. The traditional single point centralized resource storage mode is unable to meet with the ubiquitous learning requirements whether the resource storage or the promptness of obtaining resources. IoT make not only real world are connected, but also the real world (physical narrow room) and virtual worlds (digital information space) are all interconnected, and it support effectively M2M interaction. IoT make every things of learning environment digital, intelligence and networking, make learning seamless integration, learner study what they need at any time, at anyplace, and adjust corresponding learning content, and make learning environment intelligence. For example, monitor and control light brightness by sensor; learn outdoor things by RFID, and so on.

3 Mean-Shift Clustering Algorithm

Mean-shift represents a general non-parametric mode finding/clustering procedure [10]. In contrast to the classic k-means clustering approach, there are no embedded assumptions on the shape of the distribution nor the number of modes/clusters. Mean-shift was first proposed by Fukunaga and Hostetler, later adapted by Cheng for the purpose of image analysis and more recently extended by Comaniciu, Meer and Ramesh to low level vision problems, including, segmentation adaptive smoothing and tracking.

The main idea behind mean-shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local optima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

Here is briefly described the variable bandwidth mean-shift procedure [11,12]. Given n data points x_i on a d-dimensional space R^d and the associated bandwidths $h_i = h(x_i)$, $i = 1, \ldots, n$, the sample point density estimator obtained with profile $k(x)$ is given by:

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} k\Big(\parallel \frac{x - x_i}{h_i} \parallel^2\Big). \tag{1}$$

It is utilized multivariate normal profile:

$$k(x) = e^{-\frac{1}{2}x}, x \geq 0. \tag{2}$$

Taking the gradient of Eq. (1), the stationary points of the density function satisfy:

$$\frac{2}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+2}} (x_i - x) g\Big(\parallel \frac{x - x_i}{h_i} \parallel^2\Big) = 0, \tag{3}$$

where $g(x) = -k'(x)$. The solution can be found iteratively via the fixed point algorithm

$$\overline{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{h_i^{d+2}} g\Big(\parallel \frac{x-x_i}{h_i} \parallel^2\Big)}{\sum_{i=1}^{n} \frac{1}{h_i^{d+2}} g\Big(\parallel \frac{x-x_i}{h_i} \parallel^2\Big)}, \tag{4}$$

which is called mean-shift procedure. Comaniciu and Meer [13] show that the convergence to a local mode of the distribution is guaranteed when the mean-shift iterations are started at a data point. The mean-shift procedure for a given point x_i is shown in Algorithm 1.

Algorithm 1. The process of mean-shift algorithm.

1: Compute the mean-shift vector $m(x_i^t)$
2: Translate density estimation window: $x_i^{t+1} = x_i^t + m(x_i^t)$
3: Iterate steps 1. and 2. until convergence, i.e., $\nabla f(x_i) = 0$

4 Testbed Description

In Fig. 1 is shown the structure of IoT-based e-learning testbed. Our testbed is composed of a Raspberry Pi B+ [14,15], MWM and SmartBox. The Raspberry Pi is a credit card-sized single-board computer developed by the Raspberry Pi Foundation. In the implemented system, we use MWM to get human EEG data. These data are sent to Raspberry Pi computer and are processed by mean-shift clustering algorithm. Then, the results are send to the SmartBox, which can change the mood of humans using its functions.

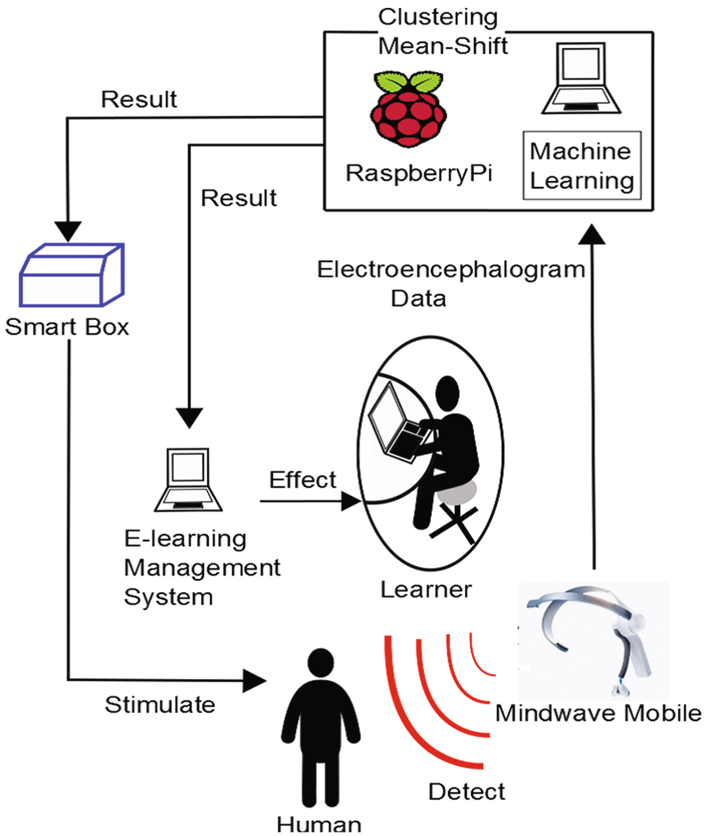


Fig. 1. Structure of IoT-Based E-Learning testbed.

4.1 MWM

A snapshot of the MWM is shown in Fig. 2. MWM is a device capable of acquiring the human EEG data [16]. The device measures the raw signal, power spectrum (Delta, Theta, Alpha, Beta and Gamma waves), Attention level, Mediation level and blink detection. The raw EEG data are received at a rate of 512 Hz. Other measured values are made every second. Therefore, raw EEG data are the main source of information on EEG signals using MWM. By MWM can be determined how effectively the user is engaging Attention (similar to concentration) by decoding the electrical signals and applying algorithms to provide readings on a scale of 0 to 100. These values are described in Table 1.

Fig. 2. A snapshot of MWM.

Table 1. Descriptions of eSense meter values.

Values	Description
0–20	Strongly lowered levels
20–40	Reduced levels
40–60	Neutral/Baseline levels
60–80	Slightly elevated/higher than normal levels
80–100	Elevated/heightened levels

4.2 EEG

4.2.1 Delta Wave

Delta waves in humans are between 0 and 4 Hz. In general, low frequency oscillations dominate the human EEG during early developmental stages [17]. Low frequency oscillations are presumably generated when humans are sleeping.

4.2.2 Theta Wave

Human theta wave lies between 4 and 7.5 Hz. The crucial finding is that with increasing task demands theta wave synchronizes. If EEG power in a resting condition is compared with a test condition, alpha power decreases and desynchronizes, while theta power increases and synchronizes.

4.2.3 Alpha Wave

Alpha wave is the dominant frequency in the human scalp EEG of adults [18]. When a healthy adult relaxes, with eyes closed, rhythmic electric activity of around 10 Hz can be recorded over the posterior scalp regions. The fact that alpha wave clearly is an oscillatory component of the human EEG has led to a recent "renaissance" in the interest of EEG alpha activity. Usually, alpha frequency is defined in terms of peak or gravity frequency within the traditional alpha frequency range of about 7.5–12.5 Hz.

4.2.4 Beta Wave

Human beta wave between 12 and 30 Hz. EEG is sensitive to a continuum of states ranging from stress state, alertness to resting state, hypnosis, and sleep [19]. During normal state of wakefulness with open eyes beta waves are dominant. In relaxation or drowsiness alpha activity rises and if sleep appears power of lower frequency bands increase. Sleep is generally divided into two broad types: Non-Rapid Eye Movement sleep (NREM) and REM sleep. NREM and REM occur in alternating cycles, when slower dominant frequencies responsiveness to stimuli decreases.

4.2.5 Gamma Wave

Human gamma wave lies between 30 and 80 Hz. A dramatic increase of activity in the gamma band in association with meditation-visible in the raw EEG-was reported in a study using trained practitioners of meditation [20].

4.3 SmartBox Description

We implemented a SmartBox device [1,2]. The size of the SmartBox is $35 \times 7 \times 12$ [cm]. The SmartBox is equipped with different sensors (for sensing learner situation) and devices (used for stimulating learner's motivation). The SmartBox has the following sensors and functions.

- Body Sensor: for detecting the learner's body movement.
- Chair Vibrator Control: for vibrating the learner's chair.
- Light Control: for adjusting the room light for study.
- Smell Control: for controlling the room smell.
- Sound Control: to emit relaxing sounds.
- Remote Control Socket: for controlling AC 100 [V] socket (on-off control).

A snapshot of the SmartBox is shown in Fig. 3.

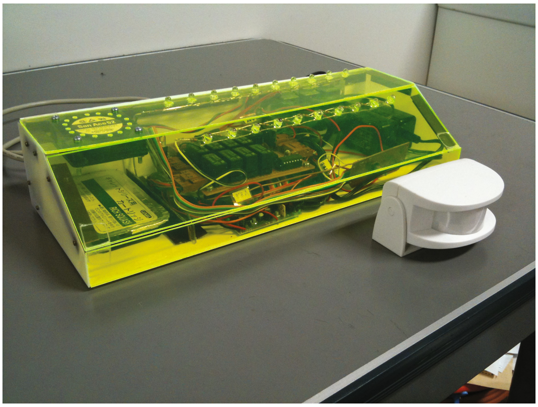


Fig. 3. A snapshot of SmartBox.

5 Experiment and Simulation Results

We carried out an experiment and then used the mean-shift clustering algorithm to cluster the experimental data. We have collected theta waves data by Fast Fourier Transform (FFT) [21]. The FFT can derive the power of EEG data. These data are collected by using the scikit-learn, which is a general purpose machine learning library for the Python [22].

Table 2. Simulation parameters.

Parameters	Values
Number of clusters	4
Initial centroids	Designation
Precomputed distance	True

The simulation parameters of mean-shift clustering algorithm are shown in Table 2. We carried out an experiment with a student of our laboratory for theta type of brain waves. We used MWM to get the data and considered four situations: sleeping, relaxing, active and moving. The experimental data are shown in Fig. 4.

Then, we used mean-shift clustering algorithm to cluster the data as shown Fig. 5. The + mark represents the center of gravity of each cluster. It can be seen from this figure that the number of points of sleep and relax is large. The number of points of active and moving is small. Therefore, this state of student seems to be relatively calm.

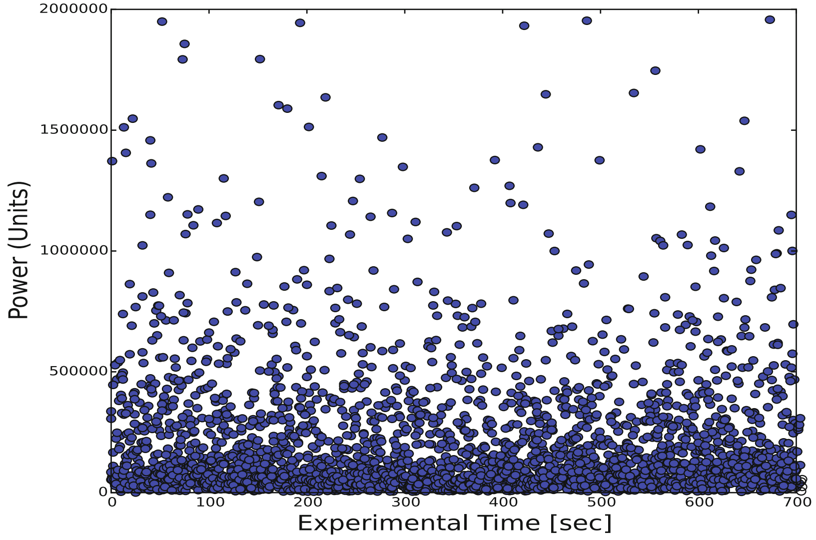


Fig. 4. Measural data for theta wave.

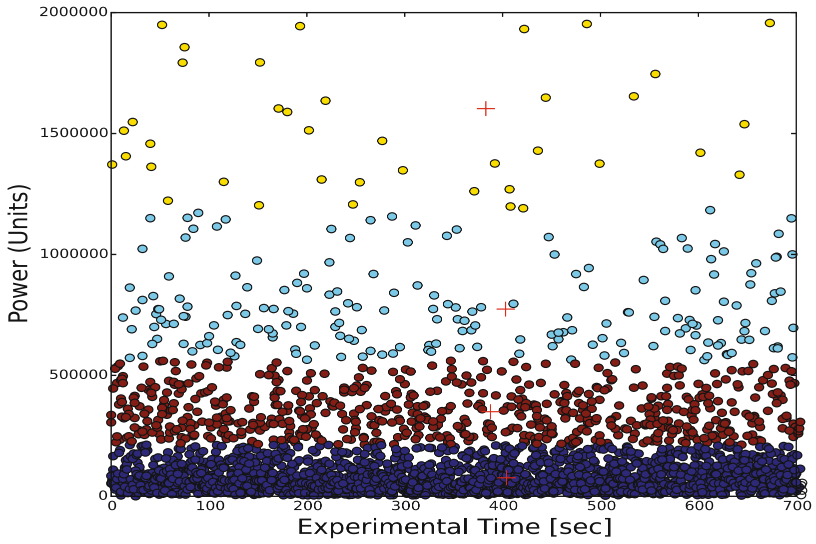


Fig. 5. Clustering of theta wave data.

6 Conclusions and Future Work

In this paper, we presented an IoT-based e-learning testbed. We evaluated its performance by carrying out an experiment and clustering the data.

We carried out the experiment with a student of our laboratory for theta type of brain waves. We used MWM to get the data and considered four situations: sleeping, relaxing, active and moving. Then, we used mean-shift clustering algorithm to cluster the data. The evaluation results show that our tesbed can judge the human situation by using theta waves.

In the future, we will carry out many experiments using the implemented testbed.

References

1. Matsuo, K., Barolli, L., Xhafa, F., Kolici, V., Koyama, A., Durresi, A., Miho, R.: Implementation of an E-Learning system using P2P, Web and Sensor Technologies. In: Proceedings of IEEE Advanced Information Networking and Applications, pp. 800–807 (2009)
2. Matsuo, K., Barolli, L., Arnedo-Moreno, J., Xhafa, F., Koyama, A., Durresi, A.: Experimental results and evaluation of SmartBox stimulation device in a P2P E-Learning system. In: Proceedings of Network-Based Information Systems, pp. 37–44 (2009)
3. Domingo, M.G., Forner, J.A.M.: Expanding the learning environment: combining physicality and virtuality - the internet of things for eLearning. In: Proceedings of 10-th IEEE International Conference on Advanced Learning Technologies, pp. 730–731 (2010)
4. Gasparini, I., Eyharabide, V., Schiaffino, S., Pimenta, M.S., Amandi, A., de Oliveira, J.P.M.: Improving user profiling for a richer personalization: modeling context in E-Learning. In: Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers, Chap. 12, pp. 182–197 (2012)
5. de Freitas, V., Marcal, V.P., Gasparini, I., Amaral, M.A., Proenca Jr., M.L., Brunetto, M.A.C., Pimenta, M.S., Ribeiro, C.H.F.P., de Lima, J.V., de Oliveira, J.P.M.: AdaptWeb: an adaptive web-based courseware. In: Proceedings of International Conference on Information and Communication Technologies in Education, pp. 131–134 (2002)
6. Schiaffino, S., Garcia, P., Amandi, A.: eTeacher: providing personalized assistance to e-Learning students. Comput. Educ. **51**(4), 1744–1754 (2008)
7. Zanella, A., Bui, N., Castellani, A., Vangelista, L.: Internet of things for smart cities. IEEE Internet Things J. **1**(1), 22–32 (2014)
8. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
9. Bellavista, P., Cardone, G., Corradi, A., Foschini, L.: Convergence of MANET and WSN in IoT urban scenarios. IEEE Sens. J. **13**(10), 3558–3567 (2013)
10. Derpanis, K.G.: Mean Shift Clustering. http://www.cse.yorku.ca/~kosta/CompVis-Notes/mean-shift.pdf Accessed 14 Sept. 2016
11. Comaniciu, D.: Variable bandwidth density-based fusion. In: Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR-2003), vol. 1, pp. 59–66 (2003)
12. Tuzel, O., Porikli, F., Meer, P.: Kernel methods for weakly supervised mean shift clustering. In: Proceedings of 12-th IEEE International Conference on Computer Vision, pp. 48–55 (2009)
13. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
14. Raspberry Pi Foundation. http://www.raspberrypi.org/

15. Oda, T., Barolli, A., Sakamoto, S., Barolli, L., Ikeda, M., Uchida, K.: Implementation and experimental results of a WMN testbed in indoor environment considering LoS scenario. In: Proceedings of 29-th IEEE International Conference on Advanced Information Networking and Applications, pp. 37–42 (2015)
16. NeuroSky to Release MindWave Mobile. http://mindwavemobile.neurosky.com
17. Knyazev, G., et al.: EEG delta oscillations as a correlate of basic homeostatic and motivational processes. In: Neuroscience Biobehavioral Reviews, vol. 36, issue no. 1, pp. 677–695. Elsevier (2012). doi:10.1016/j.neubiorev.2011.10.002
18. Klimesch, W., et al.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews, vol. 29, issue no. (2–3), pp. 169–195. Elsevier (1999)
19. Teplan, M., et al.: Fundamentals of EGG measurement. In: Measurement Science Review, vol. 2, Section 2, pp. 1–11 (2002)
20. Vialatte, F.B., Bakardjian, H., Prasad, R., Cichocki, A.: EEG paroxysmal gamma waves during Bhramari Pranayama: a Yoga breathing technique. In: Consciousness and Cognition, vol. 18, issue no. 4, pp. 977–988. Elesevier (2009). doi:10.1016/j.concog.2008.01.004
21. Akin, M.: Comparison of wavelet transform and FFT methods in the analysis of EEG signals. J. Med. Syst. **26**(3), 241–247 (2002)
22. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**(10), 2825–2830 (2011)

Implementation of an Actor Node for an Ambient Intelligence Testbed Considering Bed Temperature and Room Lighting: Its Effects on Human Sleeping Condition

Ryoichiro Obukata[1(✉)], Miralda Cuka[1], Donald Elmazi[1], Tetsuya Oda[2], Keita Matsuo[3], and Leonard Barolli[3]

[1] Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
obukenkyuu@gmail.com, mcuka91@gmail.com, donald.elmazi@gmail.com

[2] Department of Information and Computer Engineering, Okayama University of Science (OUS), 1-1 Ridaicho, Kita-ku, Okayama 700–0005, Japan
oda.tetsuya.fit@gmail.com

[3] Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan
{kt-matsuo,barolli}@fit.ac.jp

Abstract. Ambient intelligence (AmI) deals with a new world of ubiquitous computing devices, where physical environments interact intelligently and unobtrusively with people. AmI environments can be diverse, such as homes, offices, meeting rooms, schools, hospitals, control centers, vehicles, tourist attractions, stores, sports facilities, and music devices. In this paper, we present the implementation and evaluation of actor node for AmI testbed. The actor node is eqquiped with a Reidan Shiki PAD and a fluorescent lamp. For evaluation, we considered respiratory rate and heart rate metrics. We carried out an experiment and clustered sensed data by k-means clustering algorithm. From experimental results, we found that the actor node of the implemented AmI testbed gives a good effect to human during sleeping.

1 Introduction

Ambient Intelligence (AmI) is the vision that technology will become invisible, embedded in our natural surroundings, present whenever we need it, enabled by simple and effortless interactions, attuned to all our senses, adaptive to users and context and autonomously acting [1]. High quality information and content must be available to any user, anywhere, at any time, and on any device.

In order that AmI becomes a reality, it should completely envelope humans, without constraining them. Distributed embedded systems for AmI are going to change the way we design embedded systems, in general, as well as the way we think about such systems. But, more importantly, they will have a great

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_7

impact on the way we live. Applications ranging from safe driving systems, smart buildings and home security, smart fabrics or e-textiles, to manufacturing systems and rescue and recovery operations in hostile environments, are poised to become part of society and human lives.

There are a lot of works done for AmI. In [2], the authors present a simulation environment that offers a library of Networked Control Systems (NCS) blocks. Thus, the constraints can be considered and integrated in the design process. They describe a real process, an inverted pendulum, which is automated based on Mica nodes. These nodes were designed especially for AmI purposes. This real NCS serves as a challenging benchmark for proving the AmI suitability of the controllers.

In [3], the authors present the development of an adaptive embedded agent, based on a hybrid PCA-NFS scheme, able to perform true real-time control of AmI environments in the long term. The proposed architecture is a single-chip HW/SW architecture. It consists of a soft processor core (SW partition), a set of NFS cores (HW partition), the HW/SW interface, and input/output (I/O) peripherals. An application example based on data obtained in an ubiquitous computing environment has been successfully implemented using an FPGA of Xilinx's Virtex 5 family [4].

In [5], the authors describe a framework to Context Acquisition Services and Reasoning Algorithms (CASanDRA) to be directly consumed by any type of application needing to handle context information. CASanDRA decouples the acquisition and inference tasks from the application development by offering a set of interfaces for information retrieval. The framework design is based on a data fusion-oriented architecture. CASanDRA has been designed to be easily scalable; it simplifies the integration of both new sensor access interfaces and fusion algorithms deployment, as it also aims at serving as a testbed for research.

In this work, we implement an actor node for an AmI testbed. We investigate the effects of the bed temperature and room lighting on human sleeping condition by using the k-means clustering algorithm. As evaluation metrics, we considered respiratory rate and heart rate.

The structure of the paper is as follows. In Sect. 2, we present a short description of AmI. In Sect. 3, we give a brief introduction of k-means clustering algorithm. In Sect. 4, we show the description and design of the testbed. In Sect. 5, we discuss the experimental results. Finally, conclusions and future work are given in Sect. 6.

2 Ambient Intelligence (AmI)

In the future, small devices will monitor the health status in a continuous manner, diagnose any possible health conditions, have conversation with people to persuade them to change the lifestyle for maintaining better health, and communicates with the doctor, if needed [6]. The device might even be embedded into the regular clothing fibers in the form of very tiny sensors and it might communicate with other devices including the variety of sensors embedded into the home to monitor the lifestyle. For example, people might be alarmed about the lack of a healthy diet based on the items present in the fridge and based on what they are eating outside regularly.

The AmI paradigm represents the future vision of intelligent computing where environments support the people inhabiting them [7–9]. In this new computing paradigm, the conventional input and output media no longer exist, rather the sensors and processors will be integrated into everyday objects, working together in harmony in order to support the inhabitants [10]. By relying on various artificial intelligence techniques, AmI promises the successful interpretation of the wealth of contextual information obtained from such embedded sensors, and will adapt the environment to the user needs in a transparent and anticipatory manner.

3 The *k*-means Algorithm

Here, we briefly describes the standard k-means algorithm [11]. The k-means is a typical clustering algorithm in data mining and which is widely used for clustering large set of data. The k-means algorithm is one of the most simple, non-supervised learning algorithms, which was applied to solve the problem of the well-known cluster [12]. It is a partitioning clustering algorithm, this method is to classify the given date objects into k different clusters through the iterative, converging to a local minimum. So the results of generated clusters are compact and independent. The algorithm consists of two separate phases. The first phase selects k centers randomly, where the value k is fixed in advance. The next phase is to take each data object to the nearest center. Euclidean distance is generally considered to determine the distance between each data object and the cluster centers. When all the data objects are included in some clusters, the first step is completed and an early grouping is done. Recalculating the average of the early formed clusters. This iterative process continues repeatedly until the criterion function becomes the minimum. Supposing that the target object is x, x_i indicates the average of cluster C_i, criterion function is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - x_i|^2 , \tag{1}$$

where E is the sum of the squared error of all objects in database. The distance of criterion function is Euclidean distance, which is used for determining the nearest distance between each data object and cluster center. The Euclidean distance between one vector $x = (x_1, x_2, \ldots, x_n)$ and another vector $y = (y_1, y_2, \ldots, y_n)$, The Euclidean distance $d(x_i, y_i)$ can be obtained as follow:

$$d(x_i, y_i) = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{\frac{1}{2}} . \tag{2}$$

The process of k-means algorithm is shown in Algorithm 1. The k-means clustering algorithm always converges to local minimum. Before the k-means

Algorithm 1. The process of k-means algorithm.

1: **Input**: Number of desired clusters, k, and a database $D = d_1, d_2, \ldots, d_n$ containing n data objects;
2: **Output**: A set of k clusters;
3: Randomly select k data objects from dataset D as initial cluster centers;
4: Calculate the distance between each data object d_i $(1 \leq i \leq n)$ and all k cluster centers c_j $(1 \leq j \leq k)$ and assign data object d_i to the nearest cluster;
5: For each cluster j $(1 \leq j \leq k)$, recalculate the cluster center;
6: Until no changing in the center of clusters;

algorithm converges, calculations of distance and cluster centers are done while loops are executed a number of times, where the positive integer t is known as the number of k-means iterations. The precise value of t varies depending on the initial starting cluster centers [13]. The distribution of data points has a relationship with the new clustering center, so the computational time complexity of the k-means algorithm is $O(nkt)$. The n is the number of all data objects, k is the number of clusters, t is the iterations of algorithm. Usually requiring $k \ll n$ and $t \ll n$.

4 Testbed Description

In Fig. 1 is shown the structure of AmI testbed. Our testbed is composed of five Raspberry Pi 3 Model B [14–17]. The Raspberry Pi is a credit card-sized single-board computer developed by the Raspberry Pi Foundation [18]. The operating systems mounted on these machines are Raspbian version Debian 7.8 with kernel 3.18.11 [19].

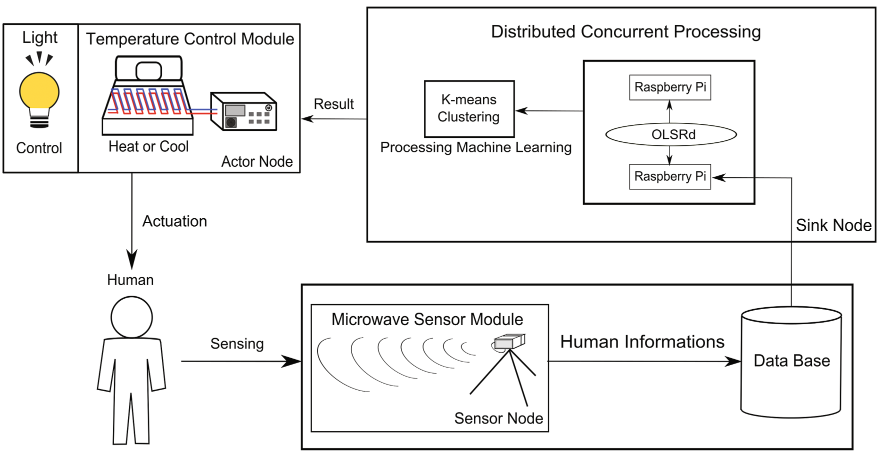


Fig. 1. Structure of AmI testbed.

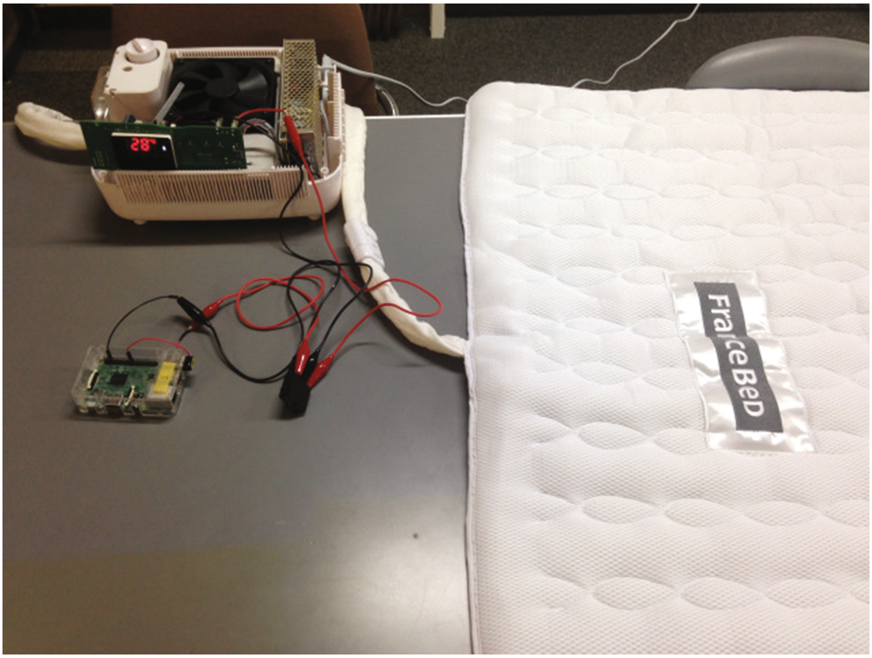


Fig. 2. Snapshot of AmI testbed actor node.

We use Microwave Sensor Module (MSM) called DC6M4JN3000, which emits microwaves in the direction of a human or animal subject [20]. These microwaves reflect back off the surface of the subject and change slightly in accordance with movements of the subject's heart and lungs. From these changes, the DC6M4JN3000 measures biological information such as heart and respiratory rates.

The DC6M4JN3000 is capable of measuring heart rate within a margin of error of $\pm 10[\%]$ when placed roughly three meters away from the target subject. The unit uses microwaves, so it can detect targets located behind obstacles such as mattresses, doors, and walls. This makes it possible to measure biological information even when the target is asleep or in situations where the targets privacy must be maintained (such as in the washroom or bathroom), thereby enabling this sensor module to boost the level of service given in elderly care or nursing care.

For actor node, we use Reidan Shiki PAD (see Fig. 2) and a fluorescent lamp. The Reidan Shiki PAD can be used for cooling and heating the bed [21]. It can adjust the temperature between 15 and 48 [°C].

As shown in Fig. 1, by using the MSM the system get the respiratory rate and heart rate data, which are sent to sink node equipped with five Raspberry Pi 3 Model B computers [22, 23]. In the sink is carried out the distributed concurrent processing for k-means algorithm. Then, the results is sent to temperature module and lighting module to control the Reidan Shiki PAD and fluorescent lamp.

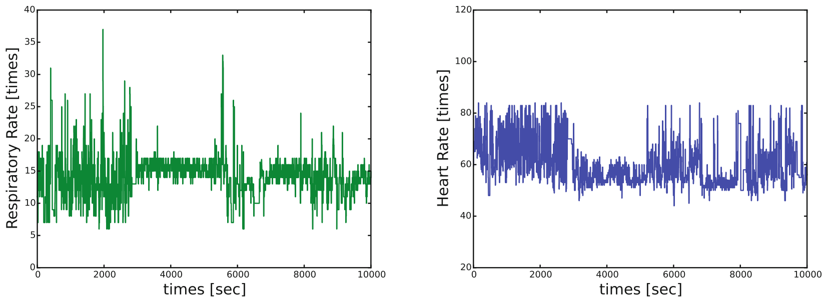
5 Experimental Results

We carried out an experiment with a student of our laboratory by using the implemented testbed. The experimental parameters are shown in Table 1. We collected data for respiratory rate and heart rate.

Table 1. Simulation parameters.

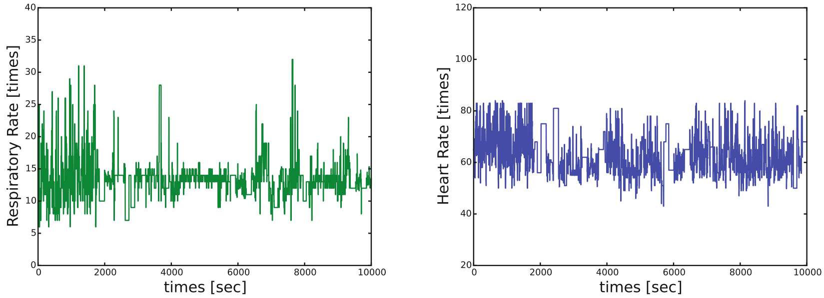
Parameters	Values
Number of clusters	3
Initial centroids	Random
Precompute distance	True

In Figs. 3 and 4, we show the respiratory rate and heart rate using AmI tesbed for two cases: room lighting and the dark room, respectively. In Fig. 5, we present the result of clustered data using k-means algorithm for these two cases. We can see 3 regions of clustering: Rapid Eye Movement (REM) sleep, light non-REM sleep, deep non-REM sleep. The red cluster shows REM sleep, the blue cluster shows light non-REM sleep and the green cluster shows deep non-REM sleep. The "+" mark shows the center of gravity of each cluster. Comparing Fig. 5(a) and (b), we can see that the implemented actor node equipped with fluorescent lamp gives a good effect to human during sleeping, because in the case of dark room there are more green cluster points compared with case of lighting the room. This shows that by using AmI testbed and room lighting control the human has a better sleeping condition.



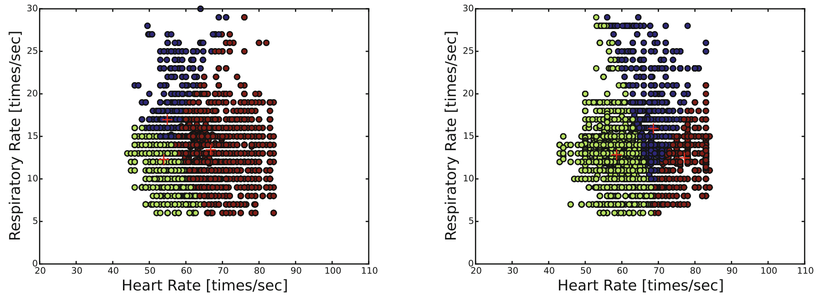
(a) Respiratory rate (b) Heart rate

Fig. 3. Sensing data during sleeping (using AmI testbed and lighting room).



(a) Respiratory rate (b) Heart rate

Fig. 4. Sensing data during sleeping (using AmI testbed and dark room).



(a) Clustering of using testbed and light no control. (b) Clustering of using testbed and light control.

Fig. 5. Experimental results using k-`means` clustering algorithm.

6 Conclusions

In this paper, we presented the implementation and evaluation of an actor node for AmI testbed. We carried out an experiment and clustered sensed data by k-means clustering algorithm. From experimental results, we found that the implemented AmI testbed and room lighting control gives a good effect to human during sleeping.

In the future, we would like to make extensive experiments using the implemented AmI testbed.

References

1. Lindwer, M., Marculescu, D., Basten, T., Zimmermann, R., Marculescu, R., Jung, S., Cantatore, E.: Ambient intelligence visions and achievements: linking abstract ideas to real-world concepts. In: Design, Automation and Test in Europe Conference and Exhibition, pp. 10–15 (2003)

2. Gabel, O., Litz, L., Reif, M.: NCS Testbed for Ambient Intelligence. IEEE Int. Conf. Syst. Man Cybernet. **1**, 115–120 (2005)
3. del Campo, I., Martinez, M.V., Echanobe, J., Basterretxea, K.: A Hardware/Software embedded agent for RealTime control of ambient-intelligence environments. In: IEEE International Conference on Fuzzy Systems, pp. 1–8 (2012)
4. Virtex 5 Family Overview. Xilinx Inc., San Jose (2009)
5. Bernardos, A.M., Tarrio, P., Casar, J.R.: CASanDRA: a framework to provide context acquisition Services ANd reasoning algorithms for ambient intelligence applications. In: International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 372–377 (2009)
6. Acampora, G., Cook, D., Rashidi, P., Vasilakos, A.V.: A survey on ambient intelligence in health care. Proc. IEEE **101**(12), 2470–2494 (2013)
7. Aarts, E., Wichert, R.: Ambient intelligence. In: Technology Guide, pp. 244–249 (2009)
8. Aarts, E., de Ruyter, B.: New research perspectives on ambient intelligence. J. Ambient Intell. Smart Environ. **1**(1), 5–14 (2009)
9. Vasilakos, A., Pedrycz, W.: Ambient Intelligence, Wireless Networking, and Ubiquitous Computing. Artech House Inc., Norwood (2006)
10. Sadri, F.: Ambient intelligence: a survey, ACM Computing Surveys, Article 36, vol. 43, issue no. 4, 66 pages (2011)
11. Na, S., Xumin, L., Yong, G.: Research on k-means clustering algorithm: an improved k-means clustering algorithm. In: Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 63–67 (2010)
12. Jigui, S., Jie, L., Lianyu, Z.: Clustering algorithms research. J. Softw. **19**(1), 48–61 (2008)
13. Abdul Nazeer, K.A., Sebastian, M.P.: Improving the accuracy and efficiency of the k-means clustering algorithm. In: Proceedings of the World Congress on Engineering, vol. 1, pp. 1–3 (2009)
14. Oda, T., Barolli, A., Sakamoto, S., Barolli, L., Ikeda, M., Uchida, K.: Implementation and experimental results of a WMN testbed in indoor environment considering LoS Scenario. In: The 29-th IEEE International Conference on Advanced Information Networking and Applications, pp. 37–42 (2015)
15. Oda, T., Barolli, L.: Experimental results of a Raspberry Pi based WMN testbed considering CPU frequency. In: The 30-th IEEE International Conference on Advanced Information Networking and Applications, pp. 981–986 (2016)
16. Oda, T., Elmazi, D., Yamada, M., Obukata, R., Barolli, L., Takizawa, M.: Experimental results of a Raspberry Pi based WMN testbed in indoor environment: a comparison study of LoS and NLoS scenarios. In: The 19-th International Conference on Network-Based Information Systems, pp. 9–14 (2016)
17. Obukata, R., Oda, T., Barolli, L.: Design of an ambient intelligence Testbed for improving quality of life. In: The 9-th International Symposium on Mining and Web, pp. 714–719 (2016)
18. Raspberry Pi Foundation. http://www.raspberrypi.org/
19. Raspbian: FrontPage. https://www.raspbian.org/
20. Sharp to Release Microwave Sensor Module. http://www.sharp-world.com/corporate/news/150625.html/
21. Corporate History - Francebed. http://www.francebed.co.jp/eng/about/history/index.html

22. Oda, T., Elmazi, D., Ishitaki, T., Barolli, A., Matsuo, K., Barolli, L.: Experimental results of a Raspberry Pi based WMN Testbed for multiple flows and distributed concurrent processing. In: The 10-th International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 201–206 (2015)
23. Obukata, R., Oda, T., Elmazi, D., Ikeda, M., Matsuo, K., Barolli, L.: Performance evaluation of an AmI Testbed for improving QoL: evaluation using clustering approach considering parallel processing, lecture notes on data engineering and communications technologies. In: Proceedings of the 11-th International Conference on Broad-Band Wireless Computing, Communication and Applications, vol. 2, pp. 623–630 (2016)

Sentiment Analysis of Specific Product's Features Using Product Tree for Application in New Product Development

Monireh Alsadat Mirtalaie[1(✉)], Omar Khadeer Hussain[1], Elizabeth Chang[1], and Farookh Khadeer Hussain[2]

[1] University of New South Wales, Canberra, Australia
monireh.mirtalaie@student.adfa.edu.au
[2] School of Software, University of Technology, Sydney, Australia

Abstract. New Product Development (NPD) is a multi-step process by which novel products are introduced in the market. Sentiment analysis, which ascertains the popularity of each new feature added to the product, is one of the key steps in this process. In this paper we present an approach by which product designers analyze users' reviews from social media platforms to determine the popularity of a specific product's feature in order to make a decision about adding it to the product's next generation. Our proposed approach utilizes a product tree generated from a product specification document to facilitate forming an efficient link between features mentioned in the users' reviews and those of the product designer's interest. Furthermore, it captures the links/interactions between a feature of interest and its other related features in a product to ascertain its polarity.

Keywords: New Product Development · Sentiment analysis · Product feature · Product tree

1 Introduction

With the constant adoption of Web 2.0 technologies, different social media channels increasingly are used by customers to share their experiences. For user-focused companies, these experiences are a vital source of information in determining their operational strategy and plans for the future. Past examples of where companies have analyzed user experiences and utilized them in their future operations are identification of health-care related issues, predicting market trend etc. [1]. In this paper our focus is on *New Product Development (NPD)*. User experiences in NPD are online reviews and/or opinions that provide invaluable information such as users' purchasing decisions, their likes, dislikes and such to product designers in developing the next iteration of a product. Given its importance, it is not surprising to see a significant increase in approaches that analyze users' opinion for producing new products [2–5].

In our previous work, we proposed a systemic framework named FEATURE for assisting product designers in the ideation phase of NPD [6, 7]. FEATURE as shown in Fig. 1 consists of three modules. The first, *New Feature Finder (NFF)* recommends to

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_8

product designers the novel features from other products that can be incorporated into the next iteration of a product [7]. By product feature, we mean characteristics of a product that describe its capabilities, components or appearance. Owing to the importance of users’ satisfaction in the success of a new product, *Feature Sentiment Analyzer (FSA)*, the second module, aims to analyze the popularity of NFF’s proposed features. *Feature Recommender (FR)* is the third module and assists product designers in choosing the features to be considered in the next iteration of a product from the many recommended according to different decision criteria.

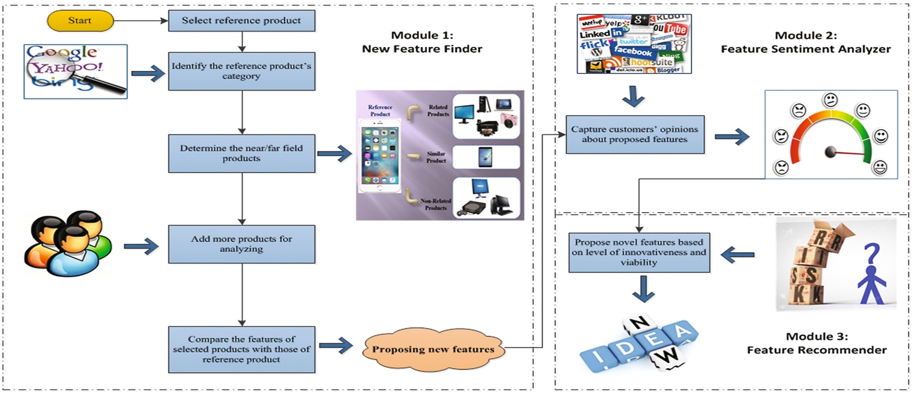


Fig. 1. - Overview of FEATURE

Our focus in this paper is on the *Feature Sentiment Analyzer (FSA)* module that analyzes user reviews to ascertain the popularity of each recommended feature from NFF. This process is widely known as sentiment analysis. In the literature, while there are many studies and approaches which attempt to extract useful knowledge from users’ reviews and analyze their semantic orientation, they cannot be directly applied to FSA as its requirements are different in the following ways:

1. FSA’s aim is to determine the popularity of specific features recommended by NFF after analyzing the product reviews. To the best of our knowledge, while there are many approaches that do feature-based sentiment analysis, they do not focus on determining the polarity of specific features using social media.
2. Once the features mentioned in the reviews are identified, they need to be linked to the features proposed by NFF. In most cases this is not a simple one-to-one match between these two features and various ambiguities at different level/s need to be addressed. While hierarchical relations between features of a product have been considered in the literature [4, 8, 9] to simplify the process of linking, their objective is to determine the overall polarity of a review, and not of a particular feature as required in FSA.

In this paper, to address these issues we propose an approach that is capable of determining the polarity of particular features using product reviews and with respect to the hierarchical relationships in the product's features. The rest of the paper is organized as follows: Sect. 2 reviews the related works; Sect. 3 represents our proposed approach; in Sect. 4 we conclude the paper by highlighting the contributions of our proposed model to the field of sentiment analysis and our future work.

2 Related Works

Sentiment analysis is the study of analyzing users' opinions about different products, news, services etc. which can be leveraged to make better decisions in various areas [10]. Feature-based sentiment analysis tries to detect the polarity of a document with respect to its discussed product (service) features [9]. In general, three tasks can be defined when performing a feature-based sentiment analysis: *feature- opinion identification*, *sentiment classification* and *feature category detection (aggregation)* [11]. In practice, not every study focuses on all of these tasks. Feature-opinion identification is concerned with detecting the feature-opinion pairs in a document. Several methods exist in the literature for fulfilling this task. Distinct linguist rules have been developed to extract various features that the users have mentioned as well as their associations with opinion words [12–14]. Identifying frequent nouns [15], information gain [16], Gini index [17], Conditional Random Fields (CRFs) [18, 19], topic modeling [20] and Bag-of-words [21] are other techniques which have been used in previous studies to handle this task. The main goal of the sentiment classification task is to determine the polarity or semantic orientation of extracted opinion words. This task can be done using supervised machine learnings such as support vector machine [17, 18] and Naïve Bayesian algorithm [15], opinion lexicons [13], Pointwise mutual information (PMI) [14], users' ratings [22], cosine similarity [21] and such like. The final task, feature category detection, aggregates the sentiment values for each product's feature in order to provide a concise overview. To do so, the extracted features need to be mapped into the main features of the product. While some studies used topic modeling techniques to complete this task [23], there are other works that have clustered the extracted features using pre-defined seed words [18, 22], word frequencies [13], hypernyms [24] and K-means technique [25].

However, as mentioned earlier, a challenge in this step is to link the customers' sentiments to the product features that they are describing. Recent research suggests that the results of feature-based sentiment analysis can be improved using an ontological approach [26]. In other words, to complete the aggregation task, some studies argue that using the hierarchical relationships which exist among various features of a product can be leveraged to improve the result of sentiment analysis [9, 10, 26, 27]. An online learning algorithm is proposed to do sentiment analysis of product reviews considering the parent-child relationship of various features [8]. In another study, an ontology-based sentiment analysis approach is developed using topic modeling techniques on product reviews [27]. ConceptNet, a large semantic network including common sense knowledge, is exploited to construct product ontology tree and further incorporating the importance of each feature into consideration while computing the overall sentiment score of a review [9, 10].

Although ConceptNet or review features can be utilized to develop an ontology tree, since they do not contain all features of a product the generated tree might be incomplete. Hence, in this study, we use product specification document as our source for detecting the hierarchical associations between various features of a product. While in previous works the sentiment score of parent features dominate those of their children in calculating the overall polarity of a review, since our focus in this study is to determine the popularity of the proposed feature we assign the highest weight to the feature itself and lower it down accordingly as we approach the root node in the product tree. In other words, since the proposed features mostly are from the lowest level (leaf nodes) of the product tree, allocating more weights to the nodes at higher levels will result in missing the real opinions about the proposed features themselves. Therefore, considering the useful knowledge that incorporating parent-child relations between features can bring into sentiment analysis studies, we propose a formula to not only take into account the information of a particular feature, but also the weighted information of its parents and ancestors in determining its polarity.

3 Proposed Approach

Before explaining our approach to address the above-mentioned challenges, we define some key terms:

Proposed features are the novel features that can be incorporated to the next iteration of the main product.
Source product is the product from which the proposed features are extracted.
Review features are those that are discussed by users in online reviews.
Product specification document is provided by the manufacturers introducing the various features of a product, mostly in the form of a table.
Feature description document is the white paper or objective review that is mostly written by an expert user describing the performance of a feature in a product.

Social media is largely unstructured and hence user reviews are broadly of two types, *pro/con* and *summary*. While pro/con-based reviews only list the positive and negative aspects of a product, summary type reviews are more elaborative and include a paragraph or two on users' opinions or general information on the performance of different aspects of products. As the objective of this study is to find the popularity of specific features, we concentrate on considering pro/con reviews to capture user opinions on different features of a product. Figure 2 represents a sample of two pro-type reviews about a camera. Our objective is to analyze these reviews and further link them to the corresponding proposed features from NFF. To achieve this, the following challenges arise.

Pro: display, shutter, viewfinder, AE/AF control
Pro: perfect especial effects, good monitor

Fig. 2. Representation of pro review

Step 1: Identifying the features that the reviews are discussing: As shown in Fig. 2, a user may express his/her feeling about different features of a product in his/her review. Therefore, the first step is to detect and identify product features that are being mentioned in the reviews. To achieve this goal we leverage SenticNet [28] which is explained further in Sect. 3.1.1

Step 2: Determining the sentiment orientation of extracted features: Once the review features are extracted, we need to find out the polarity of the expressed opinion words, if any exists. Since there is no fixed pattern for users to express their feelings in the pro/con reviews, they may either mention a product feature as a positive or negative point or use more expressive clauses to state their opinions. For example from the first pro review of Fig. 2, while the users are satisfied with the features mentioned, their level of satisfaction is much higher in the second review as words like 'perfect' and 'good' and are used to express their opinion. Hence, such varying levels of satisfaction need to be captured when determining the polarity of the features in the second review. We achieve this by using *Deeply moving software* [29]. After determining the polarity of the review features, the next challenge is to link them to the standard notions or product's components. For example from Fig. 2, while we see that both users are happy with the display of the camera, they have used different terms to describe it (such as display and monitor). Our first task is to identify those features that have similar meanings but are represented by different terminologies. To tackle this issue in FSA, we propose ontology rules to group similar terms. Steps 1 & 2 are shown as Phase 1 in Fig. 3. Detailed process is explored in Sect. 3.1.

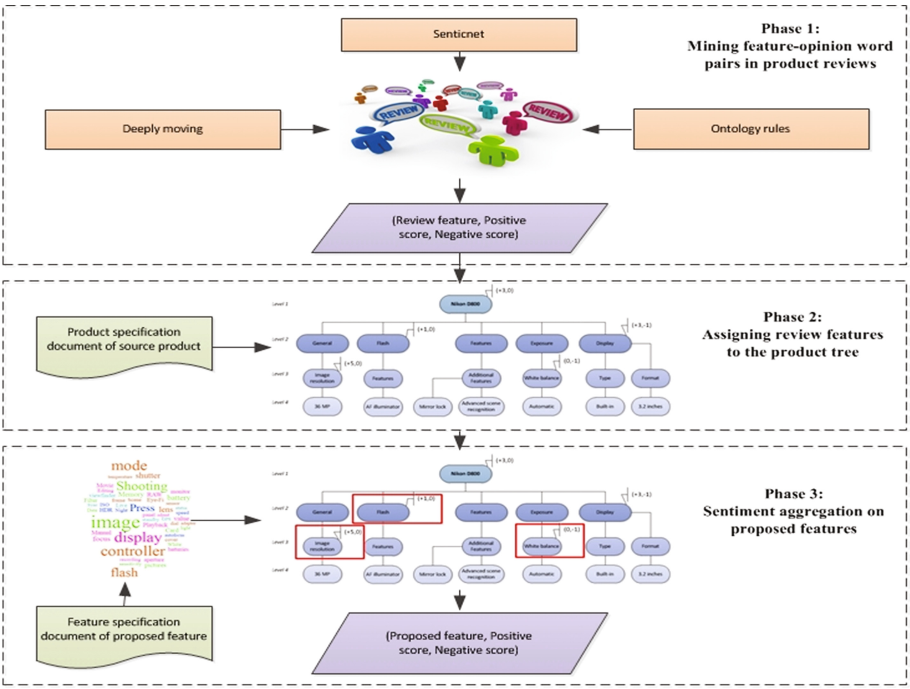


Fig. 3. An overview of the high level working of FSA

Step 3: Linking the review features to the product's components: Next task is to link the review features with the product's standard notions. As mentioned in Sect. 2, existing works use product trees that depicts various features of a product in parent-child relations to achieve this step. However, these works use review features or existing knowledge resources to generate the product tree, which may result in having an incomplete representation. In our approach, we develop a product tree by utilizing the product specification document, which will have a full list of product features. Using a product tree is helpful not only in linking the review features to their related product's components but also in having more accurate sentiment analysis, as explained in next steps. Once the product tree is generated, the review features are linked with their corresponding product features along with their polarity values. Step 3 is shown as Phase 2 in Fig. 3. Further exploration of this step is in Sect. 3.2.

Step 4: Matching the proposed features from NFF to the review features: As our main goal is to find the popularity of the proposed features from NFF, we need to link those features with the ones users mention in their product reviews. As the proposed features from NFF might be in professional language, they are unlikely to be found in their exact terms within user reviews. Therefore, to match them we first need to replace the proposed features from NFF with their most frequently used terms or aspects describing them. We achieve this by generating a word cloud for each proposed feature using its feature specification document which gives the various different terms of them. A word cloud is a graphical visualization of keywords in a text which have been selected based on their frequency rate in the context [7]. Besides, frequent terms of a word cloud reveals the links or interactions between different features of a product. For example, if we consider *advanced scene recognition* as a proposed feature, in a camera it is responsible for recognizing different characteristics related to a subject, like position, color and tones while taking a photo. It also controls other features such as autofocus, flash, white balance (Fig. 4). Using a word cloud enables us to capture such links. Hence, if according to the user's review, *flash* has a positive polarity and based on word cloud it is linked to *advanced scene recognition*, we can infer that the user is indirectly pleased with the performance of *advanced scene recognition*, as well. The exploration of this step is in Sect. 3.3.1.

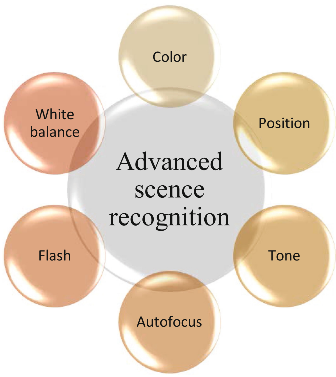


Fig. 4. Aspects (interactions) of *advanced scene recognition* in a camera

Step 5: Determining the polarity of each proposed feature: In the last step, we need to aggregate the sentiment value of various proposed features' aspects with respect to the hierarchical relationships amongst product features. Using parent-child relations here can help in determining the polarity of the proposed features that have not been mentioned in product reviews. For instance, if *3D tracking AF* is a proposed feature, there may exist no review about this specific feature because it is a technical term. However, according to Fig. 2, the user is happy with the performance of *AE/AF control* feature in the camera. Since *3D tracking AF* is a child of *AE/AF control* in the camera's product tree we argue that, to some extent, the user is happy with the performance of the child node - *3D tracking AF* and utilize the parent's sentiment value to compute that of *3D tracking AF*. Detailed working of this phase is explained in Sect. 3.3.2. Steps 4 & 5 are shown as Phase 3 in Fig. 3.

The detailed working of each phase of FSA is explained below:

3.1 Phase 1: Mining Feature-Opinion Word Pairs in Product Reviews

This phase aims to extract the features of source products and their associated opinion words from social media. The working steps are explained in following:

3.1.1 Feature Extraction and Initial Sentiment Determination

A review or a given sentence may discuss more than one product feature and the user's opinion about them. Therefore, it is essential to find the polarity of a review with respect to each of its discussed features. As the first point in this step, the pro/con reviews discussing the source products are collected. Since there is no specific pattern for writing pro/con reviews, for each identified feature the algorithm assigns the initial polarity of (+1) and (−1) if it is mentioned as a pro or con review, respectively.

This is helpful in increasing the model's accuracy for the following reasons:

- Some users might mention just product features in pro/con reviews like *display*, *battery*. While other users may prefer to express their feelings in the form of adverbs, adjectives or even full sentences about a product feature in pro/con reviews such as *lovely display*, *poor battery*. Assigning an initial polarity score to the features mentioned as just names in the reviews will help us to assign them a basic polarity value. For those features that are described with adverbs and adjectives we adjust the polarity further in the next steps of this phase.
- Opinion words may have different polarity depending on the context or features that they are describing. For example, *large size* might be mentioned as a positive point for a mobile phone while for a camera it can be considered as a con. Hence, employing initial polarity can help us to consider broadly the polarity of ambiguous context-based opinion words.
- There might be cases where the sentiment analysis lexicon is unable to find the polarity of the extracted opinion words and hence mark them as *neutral.* Having initial polarity will result in classifying these reviews more properly.

In the next step, as not all users have the same level of knowledge and experience in relation to a product, by relying on the websites' metrics, we assign a higher weight to the reviews written by expert users. Since each review may contain information about different features of a product, as a pre-processing step, the algorithm breaks the reviews to its different clauses using a comma separator. In order to extract feature-opinion word pairs, the algorithm follows an approach which needs no labeled data for training. A rule-based software, *SenticNet*, is employed to capture the association between product feature and the expressions of opinions that come along it in the reviews to describe the user's feeling about that feature [28]. SenticNet is a freely available concept-level sentiment analysis tool, which exploits the natural language processing techniques to do tasks such as polarity detection, concept identification, aspect mining and multimodal sentiment analysis. Our objective in this step is to find the pairs of (product feature, opinion word) within online reviews.

3.1.2 Feature Sentiment Classification

In order to analyze more accurately, our next goal is to classify the extracted opinion words based on their polarity strength. Numerous sentiment lexicons are publicly available which contain the term/words with their polarity value. In our approach of FSA, we classify the opinion words with the help of *Deeply moving software* [29] on the scale of −5 to 5 (strongly negative to strongly positive respectively). However, for some words, the sentiment lexicon algorithm may incorrectly classify the polarity of words mentioned in a pro as negative and vice versa. This drawback is addressed by retaining the initial polarity scores of those features from Sect. 3.1.1. For the rest of opinion words, the initial polarity will be updated to −5, −3, +3 and +5 for those which have been labeled as *very negative*, *negative*, *positive* and *very positive*, respectively. Next step is to aggregate the sentiment scores for each product feature over collected reviews. We define ontology rules to group those features which have the same meaning but different terminologies such as *display* and *screen*. The overall positive (negative) sentiment score is computed by determining if the review is from an expert or ordinary user using Eq. 1. The results will be stored as the tuple of (feature, positive score, negative score) in the database.

$$PS^{+} = \frac{\sum_{i=1}^{k} p_i e + \sum_{i=1}^{m} p_i o}{k+m} \tag{1}$$

where,

$e, o > o, e + o = 1$, and,

e = weight assigned to the comments posted by expert users
o = weight assigned to the comments posted by ordinary users
p = polarity strength
k = Number of comments which have positive attitudes and are mentioned by expert users
m = Number of comments which have positive attitudes and are mentioned by ordinary users
PS^{+}= *average positive polarity strength for each feature*

An example of working steps of this phase for online reviews about a camera is presented in Fig. 5. For more simplicity, we have considered $e = o = 1$ in this example.

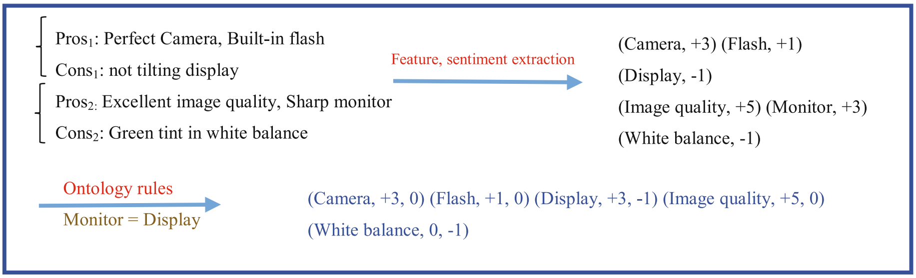


Fig. 5. Overview of feature-sentiment extraction steps

3.2 Phase 2: Assigning Review Features to the Product Tree

As discussed earlier, arranging product's aspects in a hierarchical topology can result in not losing sentiment scores of interrelated product features. In this phase of FSA the algorithm first generates a product-tree from the source product specification document. Next, it finds the match between nodes in the product tree and the features which have been extracted from online reviews. The working steps of this phase are described in more detail as following:

3.2.1 Generating Product-Tree

The hierarchical relationship between various features of a product can be represented in form of a tree, where the nodes depict the features, and the edges present parent-child relations between them. In order to generate a product tree for each source product, the algorithm collects the product specification documents and stores them in the form of a table. Each cell of product table is annotated by *Node ID* and *Parent ID*. *Node ID* is a unique number for each tree node and works as an identity for each feature in the product tree. *Parent ID* represents the node id of the immediate parent of each node which helps tracking the higher-level features of a particular node. The tree is generated by placing the source product as the root node (level 1) and proceeds to the lower-levels, or leaf nodes, according to the provided parent ids. Finally, the tree nodes are labelled with their synonyms based on the pre-defined ontology rules.

3.2.2 Assigning Review Features to the Product Tree

After generating the product tree, next step is to assign the review features to their corresponding nodes in the product tree in order to transfer their sentiment values to the tree nodes. The searching procedure begins by investigating the first level node of the product tree for the review feature. The algorithm will stop examining the lower levels once it finds a review feature at a level, otherwise, it proceeds forward to the leaf nodes. If the algorithm is not able to associate the review feature to any tree node, depending on whether the feature is a multiword or not, it follows different approaches. It asks an expert to determine the best corresponding node for that particular feature, while in case of a multiword like *ISO performance*, it breaks the feature into its subparts (*ISO* and *performance*) and then starts searching the tree for each subpart from level 1. Once the assigning procedure is finished, the sentiment scores of review features are transferred to their related tree nodes. Therefore, the output of this phase is a product tree along with sentiment scores. A snapshot of a generated product tree for a camera (Nikon D800) along with sentiment scores of its different features using information of previous example is represented in Fig. 6.

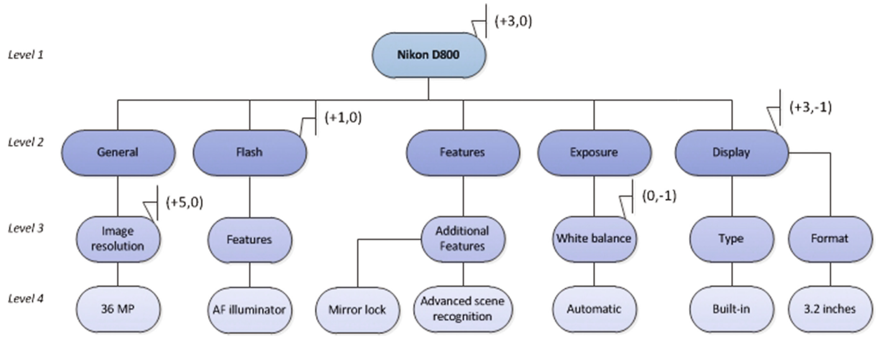


Fig. 6. A snapshot of camera product tree (Nikon D800) labelled with sentiment values

3.3 Phase 3: Sentiment Aggregation on Proposed Features

The objective of final phase of FSA is to find polarity of the proposed features from NFF based on user reviews. To do so, the algorithm first identifies the main aspects of each proposed features by extracting the most frequent nouns from their word cloud. Once different aspects of the proposed features are determined, by following the searching process as discussed in Sect. 3.2.2, the aspects are assigned to their corresponding product tree-node. Finally, the sentiment score of each proposed feature is computed through aggregation of polarity scores of its related tree nodes weighted by their importance (frequency). Different steps of this phase are as following:

3.3.1 Identifying the Most Frequent Aspects of Proposed Features from NFF

In this step, the algorithm generates a word cloud for each proposed feature recommended from NFF using feature's specification document. While the generated word cloud contains the most commonly occurring aspects of each proposed feature, it may also represent terms which are irrelevant in our further analysis, such as "allows", "keep", etc. Therefore, the next step after the generation of the word cloud is to prune the extracted terms. The pruning process involves asking an expert to remove those terms which do not represent an aspect of proposed features. Doing so, the most common aspects of each proposed feature are determined amongst frequent nouns in their word cloud. The extracted aspects are then weighted against their frequency rate as an indicator of their association with the proposed features. Next step is to match the identified aspects to the product tree using the process in Sect. 3.2.2.

3.3.2 Computing the Sentiment Score of Proposed Features

The objective of this step is to compute the overall sentiment score of each proposed feature. Considering s_i as the sentiment score of an aspect, p_i the polarity of the node in level i of the product tree, and j as an indicator of product tree levels, the following equation is used to compute the polarity score of each aspect of proposed features.

$$s_i = \begin{cases} p_i & \text{if } i = 1 \\ 0.8p_i + \sum_{j=0}^{i-2} \frac{0.2}{i^2-3i+3}(j+1)p_{(j+1)} & \text{if } i \neq 1 \end{cases} \tag{2}$$

$i = 1$ *indicates the root of the tree which contains the feeling expressions about the product as a whole*

Equation 2 assigns the whole polarity score to the aspect, if it is associated with the root node of the product tree. Otherwise, it considers the polarity score of the corresponding node in the tree, and also those of its parent node and, if applicable, the ancestor nodes but with lower weights. Positive and negative polarity score are computed separately. If the aspect cannot be associated with any nodes in the tree, the expert will determine its corresponding node. Once the polarity scores of the given aspects of a proposed feature are computed, the algorithm aggregates them to calculate the positive (negative) sentiment score of the proposed feature using Eq. 3.

$$TPSS = \sum_{k=1}^{K} w_k s_k \tag{3}$$

Where,

TPSS: *total positive sentiment score of each proposed feature*
w_k: *frequency rate of each aspect in the proposed feature's specification documents*
s_k: *sentiment score of the aspect*
K : *maximum number of proposed feature's aspects*

Considering *Advanced scene recognition* as a proposed feature, the working steps of this phase are summarized in Fig. 7.

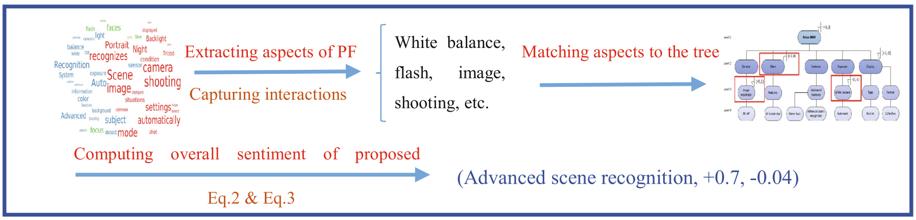


Fig. 7. Overview of different steps in finding popularity of proposed features

4 Conclusion

This paper presents the working steps of FEATURE's second module. The objective of FSA is to discover the users' opinions towards the proposed features using online reviews. The proposed algorithm first collects pro/con reviews discussing source products. The reviews are then analyzed for extracting the features as well as opinion words which they have discussed. The polarity of opinion words are identified using a dictionary. Next, a product tree is developed for each source product utilizing its product specification document. Once the tree is generated, the algorithm links the computed sentiment values to their related features in the product tree. Since the proposed features might be in professional languages, for each of them a word cloud is generated to determine their most frequent aspects. After that, the algorithm maps the extracted aspects to their related features in the product tree. Finally, the popularity of proposed features is computed by aggregating the sentiment scores of their different aspects. The uniqueness of FSA is that it can identify the sentiment values of particular features with respect to the hierarchical relations which exist in a product. It also captures the interactions between different features of a product using a word cloud. In our future work the applicability of FSA will be tested and demonstrated by using the case scenario of a mobile phone.

References

1. Tuarob, S., Tucker, C.S.: A product feature inference model for mining implicit customer preferences within large scale social media networks. In: ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Boston, Massachusetts (2015)
2. Tucker, C., Kim, H.: predicting emerging product design trend by mining publicly available customer review data. In: 18th International Conference on Engineering Design Technical University of Denmark (2011). https://www.designsociety.org/publication/30612/predicting_emerging_product_design_trend_by_mining_publicly_available_customer_review_data
3. Li, S., Nahar, K., Fung, B.C.M.: Product customization of tablet computers based on the information of online reviews by customers. J. Intell. Manuf. **26**, 97–110 (2015). doi:10.1007/s10845-013-0765-7
4. Li, Y.M., Chen, H.M., Liou, J.H., Lin, L.F.: Creating social intelligence for product portfolio design. Decis. Support Syst. **66**, 123–134 (2014). doi:10.1016/j.dss.2014.06.013
5. Wang, H., Wang, W.: Product weakness finder: an opinion-aware system through sentiment analysis. In: Industrial Management & Data Systems, vol. 114, pp. 1301–1320, (2014). http://www.emeraldinsight.com/doi/pdfplus/10.1108/IMDS-02-2014-0069
6. Mirtalaie, M.A., Hussain, O.K., Chang, E.: FEATURE : new product development using feature- drift based framework for unique aspect recommendation. In: IEEE International Conference on e-Business Engineering, Macau, China (2016). doi:10.1109/ICEBE.2016.43
7. Mirtalaie, M.A., Hussain, O.K., Chang, E., Hussain, F.K.: A decision support framework for identifying novel ideas. In: New Product Development from Cross-Domain Analysis, Information Systems, vol. 69, pp. 59–80, (2017). doi:https://doi.org/10.1016/j.is.2017.04.003

8. Wei, W., Gulla, J.A.: Sentiment learning on product reviews via sentiment ontology tree. In: ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics. Proceedings of the Conference, pp. 404–413 (2010). http://www.scopus.com/inward/record.url?eid=2-s2.0-84859962524&partnerID=tZOtx3y1
9. Mukherjee, S., Joshi, S.: Sentiment aggregation using ConceptNet ontology. In: Proceedings of Sixth International Joint Conference on Natural Language Processing, pp. 570–578 (2013) http://www.aclweb.org/anthology/I13-1065
10. Agarwal, B., Mittal, N., Bansal, P., Garg, S.: Sentiment analysis using common-sense and context information. In: Computational Intelligence and Neuroscience, (2015). doi:10.1155/2015/715730
11. Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Trans. Knowl. Data Eng. **28**, 813–830 (2016). doi:10.1109/TKDE.2015.2485209
12. Liu, Q., Gao, Z., Liu, B., Zhang, Y.: Automated rule selection for aspect extraction in opinion mining. In: Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 1291–1297 (2015)
13. Gu, X., Kim, S.: What parts of your apps are loved by users? In: 2015 30th IEEE/ACM International Conference Automation Software Engineering IEEE Computer Society, Washington, pp. 760–770 (2015) doi:10.1109/ASE.2015.57
14. Bagheri, A., Saraee, M., De Jong, F.: Care more about customers: unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. Knowledge-Based Syst. **52**, 201–213 (2013). doi:10.1016/j.knosys.2013.08.011
15. Jeyapriya, A., Selvi, C.S.K.: Extracting aspects and mining opinions in product reviews using supervised learning algorithm. In: 2nd International Conference on Communication and Electronics Systems, ICECS, pp. 548–552. IEEE, Coimbatore (2015). doi:10.1109/ECS.2015.7124967
16. Chen, Y., Perozzi, B., Skiena, S.: Vector-based similarity measurements for historical figures. In: Lecture Notes in Computer Science (Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 179–190 (2015). doi:10.1007/978-3-319-25087-8_17
17. Manek, A.S., Shenoy, P.D., Mohan, M.C., Venugopal, K.R.: Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier, World Wide Web (2016). doi:10.1007/s11280-015-0381-x.
18. Alghunaim, A., Mohtarami, M., Cyphers, S., Glass, J.: A Vector Space Approach for Aspect Based Sentiment Analysis, pp. 116–122 (2015)
19. Samha, A.K., Li, Y., Zhang, J.: Aspect - based opinion mining from product reviews using conditional random fields. In: AusDM 2015: the 13th Australasian Data Mining Conference, University of Technology, Sydney, Australia (2015)
20. Bagheri, A., Saraee, M., de Jong, F.: ADM-LDA: an aspect detection model based on topic modelling using the structure of review sentences. J. Inf. Sci. **40**, 621–636 (2014). doi:10.1177/0165551514538744
21. Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S.K., Roy, S.: Sentiment analysis using cosine similarity measure. In: IEEE 2nd International Conference on Recent Trends in Information System, pp. 27–32. doi:10.1109/ReTIS.2015.7232847
22. Huang, J., Etzioni, O., Zettlemoyer, L., Clark, K., Lee, C.: Revminer: an extractive interface for navigating reviews on a smartphone. In: Proceedings of 25th Annual ACM Symposium on User Interface Software and Technology, pp. 3–12 (2012). doi:10.1145/2380116.2380120
23. Zhai, Z., Liu, B., Xu, H.: Constrained LDA for grouping product features in opinion mining. Adv. Knowl. Discov. Data Min. **6634**, 448–459 (2011). doi:10.1007/978-3-642-20841-6_37

24. Suleman, K., Vechtomova, O.: Discovering aspects of online consumer reviews. J. Inf. Sci. **42**, 492–506 (2015). doi:10.1177/0165551515595742
25. Ye, K., Li, L., Guo, M., Qian, Y., Yuan, H.B.: Summarizing product aspects from massive online review with word representation. Knowl. Sci. Eng. Manag. **9403,** 318–323 (2015). doi:10.1007/978-3-319-25159-2
26. Zhou, L., Chaovalit, P.: Ontology-supported polarity mining. J. Am. Soc. Inf. Sci. Technol. **59**, 98–110 (2008). doi:10.1002/asi
27. Lau, R.Y.K., Li, C., Liao, S.S.Y.: Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis. Decis. Support Syst. **65**, 80–94 (2014). doi:10.1016/j.dss.2014.05.005
28. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet 4: a semantic resource for sentiment analysis based on conceptual primitives. In: Proceedings of COLING 2016, 26th International Conference Computational Linguistics Technical Papers, Osaka, Japan, pp. 2666–2677 (2016). http://sentic.net/computing/
29. Deeply Moving, (n.d.). https://nlp.stanford.edu/sentiment/ Accessed 8 May 2017

A Heuristic Machine Learning Based Approach for Utilizing Scarce Data in Estimating Fuel Consumption of Heavy Duty Trucks

Atefe Zakeri(✉), Morteza Saberi, Omar Khadeer Hussain, and Elizabeth Chang

School of Business, University of New South Wales, Canberra, BC 2610, Australia
Atefe.Zakeri@student.adfa.edu.au, m.saberi.ie@gmail.com, {o.hussain,E.Chang}@adfa.edu.au

Abstract. Although we live in an information overwhelmed era, in many applications it is still difficult to collect meaningful data due to data scarcity issues, time constraints and the cost in getting the data available. In such scenarios, we need to make better use of the scarce data available so that it can be utilized for performing further analysis. Existing approaches use available data for performing data analytics only if the estimation accuracy of the whole dataset satisfies a defined threshold. However, this approach is not beneficial when the data is scarce and the overall estimation accuracy is below the given threshold. To address this issue, we develop a heuristic approach for getting the most benefit out of the available data. We classify the existing data into classes of different errors and identify the usable data from the available data so it can be used by decision makers for performing further data analytics.

1 Introduction

In the current information driven era, we are overwhelmed with huge amounts of data. This ever-increasing expansion of data is a result of the computerization of society as well as the development of powerful online storage tools [1, 2]. However, just the capture and collection of data is of no use until that data is put into the right place and utilized by organizations to further assist them in their operations. Prediction by using data is one of the activities that is utilized by organizations to determine their future operations. However, prediction is beneficial only if the prediction accuracy of an estimation of output by machine learning methods exceeds the defined estimation accuracy threshold as shown in Fig. 1.

However, in the case of estimation accuracy not meeting the given threshold, the model configuration can be iteratively revised by changing the learning algorithms and model parameters to see if an acceptable result is achieved. In the case of estimation accuracy still not being within the threshold in spite of all possible permutations, then the dataset is totally rejected and not utilized by the current approaches. This may not

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_9

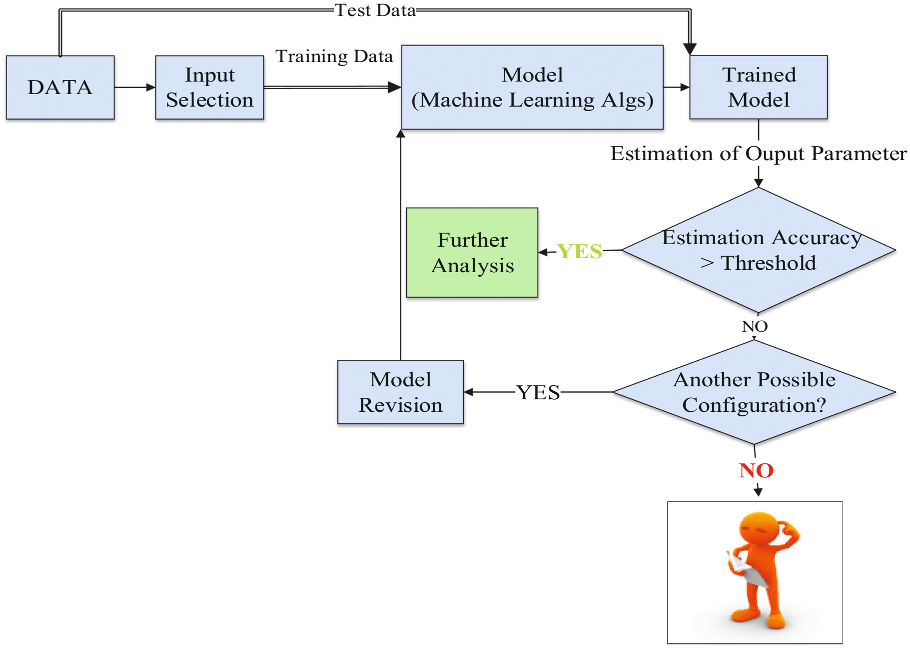


Fig. 1. Representation of the existing gap in the literature

be a problem where collecting more data is a simple, cost-reasonable task; but in those areas in which data collection involves spending lots of effort, time and money, a smarter approach of dealing with data for performing predictive analytics is needed. This is because while for some records of data the accuracy is lower than the expectations (threshold), it is very possible that for some other records of data the accuracy is high. So, for getting the most possible benefits out of the available dataset, in this research it is argued that we can dissect the dataset for which the estimation accuracy is not satisfactory to consider each record of data individually and see if it leads to either a “Green”, “Yellow” or a “Red” record as explained later in the paper.

In other words, when the estimation accuracy is not satisfactory, rather than taking a blanket approach of discarding the whole dataset from the predictive analytics process, we propose a heuristic approach for getting the most out of the available data by utilizing that part of data which leads to high accuracy results of the output estimation. This possibility of considering each record in the data as opposed to the whole dataset is not considered in the current literature. This approach is of great importance in situations as follows:

- In some applications such as automotive engineering, high cost or time is associated with data collection. For example, the collection of a small set of data for a diesel engine takes up to several weeks [3]. Also, sometimes all effective input parameters on the output are not available. This can be because of the complexity in identifying all effective input parameters which are scattered around the organization in heterogeneous sources and formats [4] or limitations in access to data due to confidential reasons. In such cases, it is very important to intelligently get the most use out of the existing collected data due to its scarcity and to the high cost in gaining it.

- Also, in organizations a knowledge engineer uses his/her domain knowledge and interprets information for making systems decisions [5]. This person is not necessarily an expert in artificial intelligence and machine learning who might develop customized algorithms which result in the right accuracy for estimation. By using the proposed heuristic approach, organizations can make use of the current available data for data analytics and get the most value out of these data to support their organizational decisions.

The rest of the paper is organized as follows: in Sect. 2, the model is represented in a scenario based manner. Section 3 validates the proposed model by applying it to a real world example and finally Sect. 4 concludes the paper with a discussion on the future work.

2 Model Framework

In this section the proposed heuristic framework for getting the most benefit out of the available data is explained. Algorithms 1–3 show the simplified representation of the procedure used in this framework which is explained as follows. For estimation of output parameter using the machine learning algorithms, as shown in Algorithm 1, data should be first divided into a training set and a testing set. Training data is used for model training while the testing data is used to measure the performance of a model on unseen data. For estimation of the output, several regression algorithms may be used. Performance of these algorithms varies due to different approaches they apply for estimation of the output.

To find the estimator with the highest estimation accuracy on the data, *Estimation Module* is used, as described in Algorithm 2. This module selects the best estimation model among the k available set of models (e.g. Decision tree). This is done by comparing the value of the Mean Absolute Percentage Error (MAPE) as shown in Eq. 1:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}. \tag{1}$$

In this equation Y_i is the output value for the ith record while $\hat{Y}_i$ is the value being estimated.

After comparing the value of MAPE on the testing set for all the estimation models, the model with the smallest MAPE on the testing set will be selected. As a result, the output of the Estimation Module will determine the model that gives the best estimation

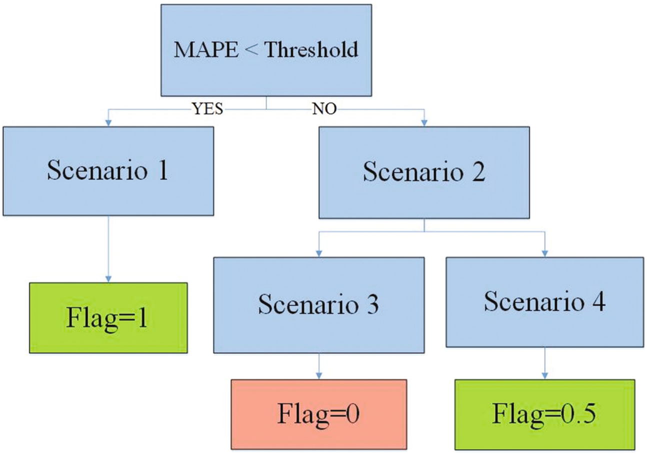


Fig. 2. Hierarchical structure of scenarios

results, namely the *Best_Estimator*. Once the best estimator is specified, it is then applied to calculate the MAPE on the testing set as well as for all the dataset as shown in Algorithm 1. Four scenarios can then be defined during the whole procedure with the hierarchical structure as shown in Fig. 2:

Scenario 1: MAPE is Less than a Given Threshold (∂)
In this case, the value for parameter *flag* as shown in line 24 of Algorithm 1 will be set to 1, indicating that the estimation values given by the *Best_Estimator* is acceptable and the dataset can be used for further analysis as required.

Scenario 2: MAPE is Greater than a Given Threshold (∂)
If the average accuracy does not meet the given threshold, instead of discarding the whole data as done by the current approaches, we propose that the focus should turn to each individual data record. This is because for some records of data, the accuracy of prediction may greatly surpass the defined threshold, while for some others it might be below the threshold. So, we need to separately analyze each data point. To do so, we utilize the "Error classification approach" as explained below.

1. Error Classification Approach
Based on the value of MAPE, each data record is divided into one of three classes as follows which are also described in lines 5–14 of Algorithm 1:

- Class *A* (Green records): Records with errors less than 5%. These records are very valuable for estimation purposes since the output value for these types of records can be estimated with high accuracy.

Algorithm 1. Main (DATA)

Input: A set of data with size n* m : **DATA**
Output: flag, Best_Estimator, Best_Classifier

$[TRAIN, TEST] \leftarrow DATA$
$Best\_Estimator \leftarrow Estimation\ Module\ (DATA)$
$MAPE \leftarrow Best\_Estimator\_MAPE(TEST)$
$MAPE\_T \leftarrow Best\_Estimator\_MAPE(DATA)$
if $(MAPE > \partial)$ do
  for i = 1: |DATA| **do**
    if $(MAPE\_T(i) < 0.05)$ **then**
      $DATANEW(i, m \leftarrow A)$
    elseif $(0.15 < MAPE\_T(i) < 0.05)$
      $DATANEW(i, m \leftarrow B)$
    else
      $DATANEW(i, m \leftarrow C)$
    end
  end
  $[TRAIN, TEST] \leftarrow DATANEW$
  $Best\_Classifier \leftarrow Classification\ Module\ (DATA)$
  $F1 \leftarrow Best\_Classifier(TEST)$
  if $F_1 > \gamma$ **then**
    $flag \leftarrow 0.5$
  else
    $flag \leftarrow 0$
  end
else
  $flag \leftarrow 1$
end
Return flag, Best_Estimator, Best_Classifier

Algorithm 2. $Estimation\ Module$ (DATA, models)

Input: A set of data with size n* m : **DATA,**
Set of estimation models with size k*1 : **models**
Output: $Best\_Estimator$

$[TRAIN, TEST] \leftarrow DATA$
for $i = 1{:}k$ **do**
  $model_i^* \leftarrow model_i(TRAIN)$
  $MAPE_{model_i^*} \leftarrow model_i^*(TEST)$
end
$j \leftarrow ArgMin\ (MAPE_{model^*})$
$Best\_Estimator \leftarrow model_j^*$
Return $Best\_Estimator$

Algorithm 3. $Classification\ Module$ (DATA, models)

Input: A set of data with size r* m : **DATANEW**, set of classification models with size l*1 : **models**

Output: $Best\_Classifier$

$[TRAIN, TEST] \leftarrow DATANEW$

for $i = 1{:}\,l$ **do**

$\quad model_i^{\,*} \leftarrow model_i(TRAIN)$

$\quad Accuracy\_Percentage_i \leftarrow model_i^{\,*}(TEST)$

end

$j \leftarrow ArgMax\ (Accuracy\_Percentage_i)$

$Best\_Classifier \leftarrow model_j^{\,*}$

Return $Best\_Classifier$

- Class B (Yellow records): Records with errors between 5% and 15%. Although the errors for these records are not as low as the ones in class A, they are still acceptable to be considered for further analysis. So, from the data analytics perspective, these records of data are also worthy of utilizing them in the estimation procedure.
- Class C (Red records): Records with errors more than 15%. Identifying records of this type is very important since they can seriously decrease the performance of estimation if they are used in the process. Thus, we need to find and remove these records from the whole dataset in order to get a better average accuracy of estimation. This can be done manually, but to automate it we need to learn about the common patterns and then make sure whether they exist in the data which allows these data points to be classified in one class. We explain that further in the next sub-section.

2. Developing a Classifier to Categorize Each Data Record into the Type of Error They Belong To

In this step, regardless of what specifically hidden patterns might be, we try to validate the classification of data in an automated way in order to get each of the three classes introduced before. If there is a classifier (like *C4.5*) which can classify the data as mentioned above, then we can be assured that the records in each individual class would probably share common features. So, we develop a classifier to take each data record and classify it in the error class it belongs to in an automated way. We apply a classification problem in which the class parameter is “Error of estimation” and there are three predefined error classes over the whole dataset as stated above. The best classifier is selected among a set of available classification models by using the *Classification Module* as shown in Algorithm 3. Now, based on the classification accuracy two scenarios, namely scenarios 3 and 4, may occur as shown in Fig. 2.

Scenario 3: The Accuracy of Error Classification is Less than a Given Threshold (γ)
In this case, the best classifier developed in scenario 2 fails to accurately classify the input data to the error class they belong to. So, as shown in Algorithm 1, the value for *flag* will be set to 0 indicating the classifier's failure in dealing with such data.

Scenario 4: The Accuracy of Error Classification is Greater than a Given Threshold (γ)
If the best classifier succeeds in classifying the input records according to their expected error class, then the value for *flag* will be set to 0.5, and with the termination of Algorithm 1, three output values, namely *flag*, *Best_Estimator* and *Best_Classifier* will result. In this situation, the *Best_Classifier* is successful in classifying the errors with an acceptable reliability rate either in class A, B or C. If an individual data record is labelled as A or B, it will be used by the *Best_Estimator* to get the estimation of output value and to be used for further analysis as required. But if the data record is labelled as C, it means that the estimation for this record contains a large error and is not acceptable. So, this record of data is removed from the whole dataset to get a more accurate output estimation.

In summary, the proposed framework, instead of ignoring all of the dataset because its error value is above the threshold, dissects each individual record in it and considers and utilizes records that fall in error classes *A* and *B* for further analysis. To show the applicability and validity of the proposed framework using real world data, in the next section, a case study from the domain of green logistics on fuel consumption estimation is discussed.

3 Validation - Case Study: Fuel Consumption Estimation of Heavy Duty Trucks

Due to the increasing contribution of freight vehicles as the largest global agent to fuel consumption and GHG emissions [6], various initiatives have been made by transportation companies to move their operations to green logistics. One of ways this being done is by transportation companies utilizing sources of fuel other than diesel which are also better on the environment. In other words, transportation companies are in the process of converting their trucks from operating on fossil fuels such as diesel to work on more green sources of fuel such as natural gas [7]. To make an informed decision which guarantees the profitability of this conversion, they need to have an accurate estimation of fuel consumption for their trucks under both diesel and natural gas sources.

As a result, estimation and prediction of fuel consumption for these heavy duty trucks have recently been studied widely. However, the collection of data for this purpose is very difficult due to the following reasons:

1. Delivery trucks are mostly working in very different and variable duty cycles including different types of roadway (freeways, surface streets, rural and urban roads), differing vehicle's speed and variable payloads. These variable working conditions will lead to a huge collection of parameters that affect the fuel consumption and make the estimation of fuel consumption quite complicated [6, 8].

2. Apart from different operating conditions, the fuel consumption is also affected by many other input parameters which are directly dependent to each vehicle's specific configuration including the truck's working life, the condition of tyres [9, 10].
3. In addition to the problem caused by the variety of input parameters, another constraint is related to the difficulty in acquiring the data and hence, data scarcity. Data collection for heavy duty trucks is not only very costly but also time-consuming. Sometimes the collection of a small set of data for a diesel engine takes up to several weeks [3].

In light of these existing difficulties in collecting the data for heavy duty trucks, it is very important to get the maximum benefit out of the data that is currently available. However, due to the above described parameter of variability, data collected comes from several external and heterogeneous data sources including both those dependent on the vehicle's specific configuration and also those related to ambient conditions. So, it is very possible that the estimation for the fuel consumption given by different machine learning algorithms was not very accurate, which results in a high estimation error. In traditional approaches as previously discussed, if the accuracy given by different machine learning algorithms does not meet the defined threshold, there are no further steps. However, given the scarcity of such data and the difficulty in collecting them, in this paper we apply our framework to increase the slice of information that we can consider to perform further analysis.

3.1 Inputs, Outputs and Preprocessing

According to the existing works already done by [11, 12], vehicle speed and roadway grade are among the most widely used input parameters [13] to estimate the fuel consumption since they are easy to measure and collect. However, for vehicles that are especially responsible for moving goods, another effective parameter, namely vehicle weight is recommended to be considered for fuel consumption calculations by a number of researchers in this area [6, 14, 15]. For trucks which frequently pick up and drop off the payload, vehicle weight is frequently changing which can be measured through on-board weight sensors.

Unfortunately, the logistics organization involved in this research does not have access to instantaneous weight measurements since no sensors were present to measure the vehicle weight. So, one of the most effective parameters on fuel consumption is missing from the analysis. However, after investigating the collected parameters, we realized that the "Engine Load" which is easily available from the truck Electronic Control Unit (ECU) has a strong positive correlation with the output parameter, which is fuel consumption. In addition, engine load is directly affected by vehicle weight. As a result, to overcome the shortcoming of the lack of access to vehicle weight data, engine load parameter can be used to indirectly reflect the impact of variable payload on fuel consumption.

Consequently, input parameters are Vehicle Speed, Roadway Grade and Engine Load while the output is Instantaneous Fuel Consumption. Vehicle speed can be directly obtained from the ECU, whereas spatial information such as latitude and longitude are acquired from a GPS device. So using a GPS software, road grade information is extracted for each pair of latitude and longitude that corresponds to each data point.

Data are collected at each instant of time with the frequency of 1 Hz (1 reading per second). Some preprocessing steps are also done on data to prepare it for running on the estimation algorithm. This included removing records from the dataset that had the data with very low speed as an indication that the vehicle is near to stop or encountered congestion while traveling. Since congestion has a high impact on fuel consumption [16], these specific periods of time are excluded from the dataset.

After preprocessing of data, a dataset containing of 4000 data records is selected for model validation which is broken down into 3000 records for training and the remaining 1000 records for the testing set. Estimation and classification thresholds (∂ and γ) are both considered as 0.85.

3.2 Implementation and Results

According to the proposed heuristic model, the first task is to find the best estimator. Since the output parameter which is fuel consumption is continuous, we should choose a suitable estimation algorithm capable of working with numeric values. Using Weka interface [17], seven well-known algorithms are trained and tested. Their associated MAPE values are reported in Table 1.

Table 1. Overall MAPE on the testing set for estimation algorithms.

Algorithm	ANN	M5P	SMO Reg	Linear Reg	M5Rules	ZeroR	IBk
MAPE	37.69	22.53	81.91	86.17	26.61	324.61	19.93

The best estimator is *IBk* with the MAPE value of almost 20%. Under the accepted protocol in the machine learning community, since the error of estimation is high and the estimation accuracy is less than the threshold (∂ = 85%), this accuracy is not accepted and none of the 4000 input data records will be used for further analysis.

However, we should investigate whether the reported unsatisfying estimation can be improved by recognizing the unacceptable part of the data. To achieve this, based on the proposed model, we go deep through the dataset to identify parts which we can use further. Thus, using the best estimator (*IBk*) on the data, the value of MAPE for each individual data record is calculated and then by comparing the corresponding MAPE for each record with ∂ threshold, we manually classify the error corresponding to each data record in one of the three classes of *A*, *B* and *C* as discussed in the Algorithm 1.

As the estimation ability of *IBk* is of importance, and also to avoid the overfitting problem, we can just rely on *IBk*'s output on the testing data with a size of 1000. Thus, this portion of data is further divided into a training set (800 records) and a testing set (200 records). Moreover, to be able to apply a wide range of classification methods provided by *Weka* interface, nominal class attribute (*A*, *B* or *C*) is converted to numeric class as required. Also, for algorithms capable of working with either nominal or numeric class values, both situations are considered.

Various classification algorithms were applied on the data and well-known accuracy estimation criteria, namely F_1, are used for performance comparison as shown in Eq. 2:

$$F_1 = \frac{2 \times precision \times recall}{precision + recall}. \tag{2}$$

Table 2 shows F_1 score together with the normal accuracy, precision and recall for the classification algorithms used.

Table 2. Comparison of classification algorithms

		Evaluation criteria			
		Accuracy	Precision	Recall	F1
Classification algorithm	*ANN (Nom[a])*	*0.86*	*0.91*	*0.81*	*0.85*
	ANN (Num[b])	*0.85*	*0.85*	*1.00*	*0.92*
	IBk (Nom)	*0.88*	*0.99*	*0.80*	*0.88*
	IBk (Num)	*0.43*	*1.00*	*0.33*	*0.50*
	J48 (C4.5)	*0.84*	*0.93*	*0.82*	*0.88*
	Lib SVM	***0.91***	***0.92***	***0.96***	***0.94***
	M5P	*0.86*	*0.85*	*1.00*	*0.92*
	Naïve Bays	*0.79*	*0.91*	*0.65*	*0.76*

[a]Nominal
[b]Numeric

Examination of this table shows the value for F_1 exceeds the accuracy threshold γ for most of the algorithms except *IBk* and *Naïve Bays*. *Lib SVM* outperforms the other algorithms since it has the largest value for both F_1 and Accuracy. Thus, it is considered as the best classifier on this dataset which also means that with an accuracy of 94% we are able to classify each record of data into one of the error classes (*A*, *B* or *C*) to which they belong. Now based on Line 19 of Algorithm 1, the value for *flag* is set to 0.5.

While this successful classification shows that there is some relation between data and error, we do not aim to extract this pattern due to the complexity of this relationship. Rather, we simply benefit from the power of this classification. In fact, if the developed classifier labels a record as *A* or *B*, we proceed with the best estimator; otherwise, we do not deal with the unsatisfying estimated part of data (Class *C*).

By applying the best classifier (*Lib SVM*) on the primary dataset, 452 records out of 4000 records (near 11%) are classified as *C*. Therefore, by discarding 11% of records we are able to make use of the remaining 89% of data for data analytic purposes. In addition, by removing records with unsatisfying estimation from the whole dataset, as expected, the average estimation accuracy which already was 80.07% (with the MAPE of 19.93%) increased to almost 99% which is quite promising.

4 Conclusion and Future Research

In this research a heuristic framework is developed to maximize the benefit out of the available data for output estimation. This framework is especially helpful in those cases in which data collection is tedious and difficult which in turn leads to data scarcity issues. In cases where the estimation accuracy is not within the accepted given threshold, rather than discarding the whole dataset, this approach can be used to make use of the available data for performing further data analytics on it. In our future work, we will look to investigate further in three directions: the first one is to extend the proposed framework to estimate the error for automated classification in case of continuous incoming stream of data; the second is to adjust the output to an input record with error value of the class it matches with; the third is to identify the size of training and testing sets to be used when we have incoming stream of data addressing scalability issues that may arise.

References

1. Witten, I.H., Frank, E., Hall, M.A.: Data Mining Practical Machine Learning Tools and Techniques, 3rd edn. Elsevier, Amsterdam (2011)
2. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco (2011)
3. Wong, P.K., Vong, C.M., Cheung, C.S., Wong, K.I.: Diesel engine modelling using extreme learning machine under scarce and exponential data sets. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **21**(Suppl. 2), 87–98 (2013)
4. Venter, M.I.: Business Intelligence (BI) initiatives: failures versus success. Interim Interdiscip. J. **4**(1), 149–163 (2005)
5. Aylett, R., Doniat, C.: Supporting the domain expert in planning domain construction. In: AIPS-2002 Workshop on Knowledge Engineering Tools and Techniques for AI Planning (2002)
6. Sharpe, B.E.N., Muncrief, R.: Literature Review: Real World Fuel Consumption of Heavy-Duty Vehicles in The United States, China, and The European Union, January 2015
7. Zakeri, A., Chang, E., Hussain, O.K.: TUNE: tree mining-based approach for dual-fuel engine performance management of heavy duty trucks. In: 2016 IEEE International Conference on e-Business Engineering (2016)
8. Department of Resources Energy and Tourism Australia: Analyses of Diesel Use for Mine Haul and Transport Operations (2011)
9. Noda, A., Tsukamoto, Y., Sato, T., Yagi, H.: Evaluation method for HDV fuel economy performance with PC Simulation and mapping procedure. SAE Tech. Pap., vol. 2003-01-20 (2003)
10. Capps, G., Franzese, O., Knee, B., Lascurain, M.B., Otaduy, P.: Class-8 Heavy Truck Duty Cycle Project Final Report (2008)
11. Ribeiro, V., Rodrigues, J., Aguiar, A.: Mining geographic data for fuel consumption estimation. In: Proceedings of IEEE 16th International Annual Conference on Intelligent Transportation Systems, pp. 124–129, October 2013
12. Pelkmans, L., Debal, P., Hood, T., Hauser, G.: Development of a simulation tool to calculate fuel consumption and emissions of vehicles operating in dynamic conditions. SAE Tech. Pap., vol. 2004-01-18 (2004)

13. Vilaça Bastos Silva, A.: Estimating Fuel Consumption From GPS Data (2014)
14. Hileman, J.I., Katz, J.B., Mantilla, J.G., Fleming, G.: Payload fuel energy efficiency as a metric for aviation environmental performance. In: 26th International Congress of the Aeronautical Sciences (2008)
15. Transportation Research Board and National Research Council: Technologies and Approaches to Reducing the Fuel Consumption of Medium- and Heavy-Duty Vehicles, The National Academies Press, Washington, DC (2010). https://doi.org/10.17226/12845
16. Franzese, O.: Effect of Weight and Roadway Grade on the Fuel Economy of Class-8 Freight Trucks, October 2011
17. The University of Waikato (2016)

A Rudimentary Approach to Unmanned Aerial Vehicle Guided Improvised Explosive Device Shrapnel Dispersal Simulation

Christopher C.K. Chan(✉), Alexander Ferworn, and David Tran

Department of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON, Canada
{christopher.c.chan,aferworn,d27tran}@ryerson.ca

Abstract. This paper proposes a methodology to compute, model and simulate a Directionally Focused Charge (DFC) explosive, delivered and deployed on an Unmanned Aerial Vehicle (UAV), with simple particle game engine physics heuristics, for estimating shrapnel trajectories and areas of impact on an urban terrain. As a preliminary study, we model a simple DFC explosive, also known as a directionally focused fragmentary charge, which is composed of a flat top and fixed sized metal canister containing nuts, bolts and ball bearings. The simulation models a small UAV capable of delivering a maximum payload of 10 kg within a flight distance of 5 km. The simulated UAV is modeled after a commonly available heavy lift commercial drone. The terrain dataset is obtained through Google Earth Engine's public data catalog – a standard Earth science raster dataset. We assert that this methodology can provide response and counter-IED teams involved in explosive threat detection with relevant information pertaining to the estimates of the risk associated with significant shrapnel impact in urban areas.

1 Introduction and Background

An improvised explosive device (IED) is a bomb constructed from military or other explosive material and deployed in unconventional ways - potentially resulting in property damage, injury and/or death [1].

The terms Unmanned Aerial Vehicle (UAV) and Unmanned Aerial System (UAS) specifically refer to any flight-capable vehicle that does not have a pilot on board and can be reused for subsequent flights [2]. Commercially available precision-guided UAVs have given rise to unique and new potential threats to populated areas from extremist activity [3]. UAVs that are equipped with explosives meant to inflict damage and harm have become a worthwhile investment for terrorist groups given the UAV's expendable nature, low size, and safe stand-off range from potential targets [4, 5]. The use of this technology for the purposes of inflicting damage to people and property has created threats to public safety and therefore more attention is warranted. A potentially fruitful direction for counter threat assessment research for UAV-carried IEDs, is the detailed reasoning and situational awareness obtained from accurate system simulations – which may be useful tools for pre and post attack analysis and for preventative planning.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_10

Since the very nature of real-world, UAV-carried, IED explosions will be relatively unique for any design and will vary in explosive power, it is important to study the potential impact that shrapnel will have on target surfaces. One method of obtaining this form of situational awareness is to employ simulations to analyze shrapnel impact points and trajectories in order to aid in counter-IED planning processes (Fig. 1).

Fig. 1. UAV Guided IED Explosion Simulation, the simulation uses a DFC model where shrapnel is considered as an explosive dispersal of solid particles.

Simulations that provide reasoning and information regarding UAV guided explosions will need to be accurate and effective given the nature of the situation. They involve a multitude of factors such as the type of explosive payload, the design of the explosive, height of the explosion, the nature of target surfaces, the composition of shrapnel and consideration of other forces at play.

In order for our work to relate to reality and lead to an accurate simulation, we start with an assumption and focus on one type of liquid-based explosive, a DFC, a controlled explosive dispersal algorithm of solid particles [6], and a non-specific UAV to deliver the payload to a certain height and point above the 3D terrain. Our simulation is developed in the Unity game engine [7], and relies on the additional assumption of simple aerial heuristics. Our focus is limited to the initial forces at play – computed with game engine particle physics. The secondary effects of the explosion and shrapnel impact such as collateral damage from sympathetic secondary explosions are not considered [8].

We assert that our rudimentary method to computing, modeling and simulating the impact points and trajectory of shrapnel is a step towards a high-fidelity UAV-guided IED splatter analysis simulation.

DFC is a variation of an IED, with characteristics such as a flat top plate - as opposed to concave top plates, commonly seen in explosively formed penetrator/projectiles (EFPs) [9]. DFCs have canisters that are commonly designed from cast copper or cut metal, equipped with shrapnel as nuts, bolts and ball bearings [10] (Fig. 2).

These characteristics allow for digital models to determine trajectory with a higher degree of accuracy than DFC's counterpart – EFPs, in addition, the light-weight thin canister produces minimal effect on the trajectory of shrapnel, and the design of a flat

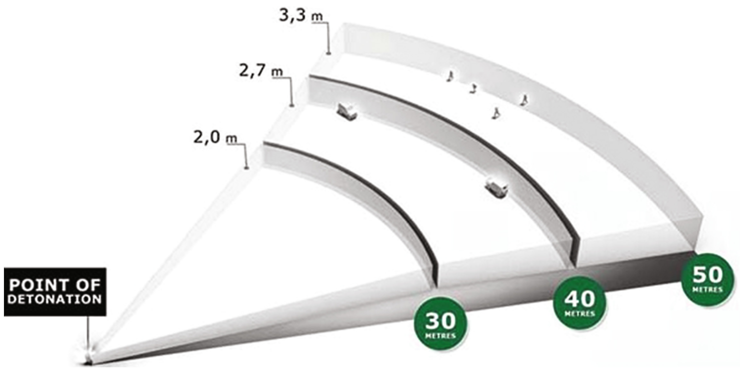


Fig. 2. Impact zone for a Directed Fragmentation Charge (DFC) – 30 m distance with overall effective range of 50 m [11].

top panel tends to produce a lower half spherical dispersal of shrapnel when subject to forces of a controlled liquid explosion [6].

2 Related Work

In the computer graphics community, many publications address modelling and simulating explosions based on physically based approaches [12–14], these works concentrate on the shockwave effects and modelling the propagation of a simple explosion through the air using computational fluid dynamic models in a closed system. These techniques are useful for generating and rendering special visual effects such as dust clouds, and fireballs, but may not be accurate or useful in determining impact and trajectories of debris and shrapnel.

For real time interactive simulations, particle systems and imaged-based techniques are commonly used, where the focus is to generate and render fire, explosions, and clouds in a more accurate dispersal of particles manner [15–17].

Simulations that compute, model and simulate explosions utilize closed system techniques to hone in on a viable predictive simulation for a very specific use case, which is often using a more computationally friendly approach and simplified method of mimicking the physical interactions involved in a real life explosion.

The closed system equations commonly used relate to the field of fluid dynamics [18], material point [19], and vortex particle method [20] – all of which are focused studies meant for a specific use case and are computationally viable for that specific purpose, such as reproducing a realistic visual model and special effect of a gas explosion resulting in dispersal of gas particles (Fig. 3).

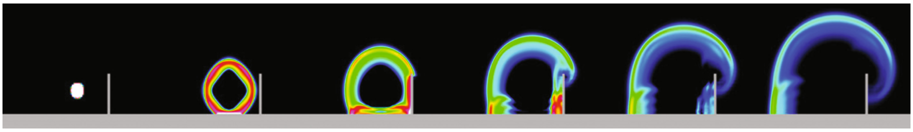


Fig. 3. A computer graphics implementation of animated cross sections of a blast wave for simulating an explosion near a barrier and the changes in blast wave at timed intervals [13].

Fluid dynamics applied to explosions allow for easier computation of forces - elements act as if they are enclosed in a viscous particle environment rather than a culmination of other forces that are inevitably at play in an open air, open system real explosion.

The most closely related work is a simulation which models terrain deformations and fractures from the impact of explosions [21]. The theory behind this work is heading towards the full scientific simulation of the real physical processes associated with explosions.

High fidelity simulations would be complex and computationally expensive, and are generally neglected in the field of computer graphics, in which the focus is mainly on the visualization of the explosion, blast waves, and visual impact on surrounding objects (Fig. 4).

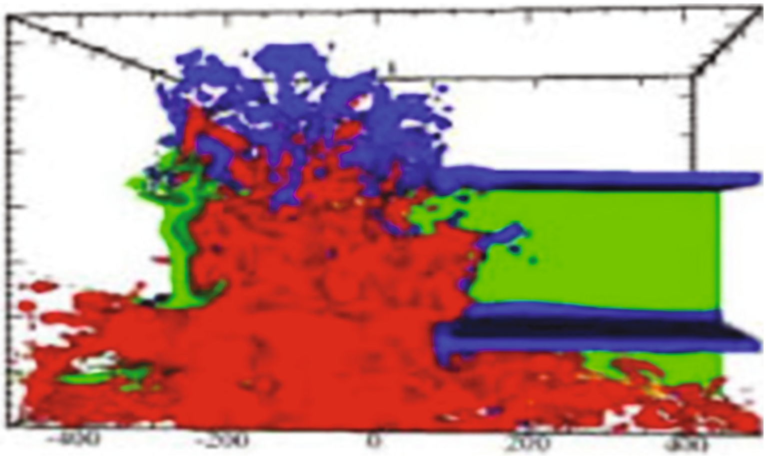


Fig. 4. Simulation of the detonation of an explosive device in a structure (closed system fluid dynamics simulation) [22].

Simulations for military applications such as modeling high-precision guided missiles and their impacts are abundant [23–25], but little work has been done with the primary focus of UAV guided IED explosions and impact on target surfaces.

3 Methodology and Experiments

3.1 Explosive Dispersal of Solid Particles

In our game simulation, we consider 100 equal game objects as shrapnel (nuts, bolts, and ball bearings). For simplicity and computation sake, these game objects representing shrapnel are equal in sizes. These game objects are scaled to 1/100 the size of the UAV model in the game. The objects are attached to the UAV and move with it when it is in flight. Each individual game object is associated with the Zhang et al. implementation of explosive dispersal of solid particles [6]. The important feature of this dispersal process is the initial geometry of dispersion. Based on our DFC model, we apply a spherical geometric dispersal to each shrapnel game object – and by applying concentration profiles, and velocity information, the shrapnel disperses as a solid particle cloud as a function of time (Fig. 5).

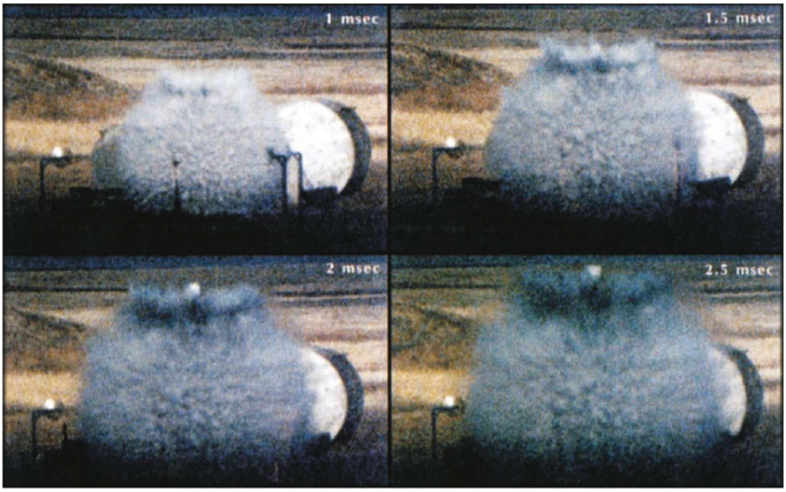


Fig. 5. Explosive dispersal of solid particles, modelled using Zhang et al. gas-solid flow model, which incorporates material density and pressure [6].

Zhang et al. uses flow topology to account for the shockwave sling-shot effect, and propagation of waves. There are other factors at play, such as uncertainties in pressure, sound speed and random inelastic collisions, which inherently add too much complexity to the equations. For the purposes of our rudimentary simulation, we do not implement these factors. However, the addition of such complexities should be implemented as future work.

In our simulation, we use the basic flow topology equation, which takes into account material density, pressure and initial forces with a geometric spherical explosion under the assumption of a closed system. This heuristic model also assumes no external forces are applied other than material density, pressure, force and direction.

3.2 Game Engine Heuristics

We assume a closed system environment in our simulation, so the UAV and trajectories of shrapnel are not subject to turbulence and/or wind gusts. A 3D 10 km by 10 km 3D swath of terrain is taken from the public dataset provided by Google Earth. The size of the UAV model, DFC, and shrapnel are adjusted to scale - proportionate to the size of the terrain. Manual adjustment of this process is required.

3D Google Earth Terrain

Google Earth [26] provides a 3D virtual map and geographical information created with a catalog of satellite imagery and geospatial datasets for research purposes. Our simulation imports any 10 km by 10 km 3D Google Earth terrain, as shown in Fig. 6, and the geospatial data allows for spatial analysis of significant points of impact from debris and shrapnel.

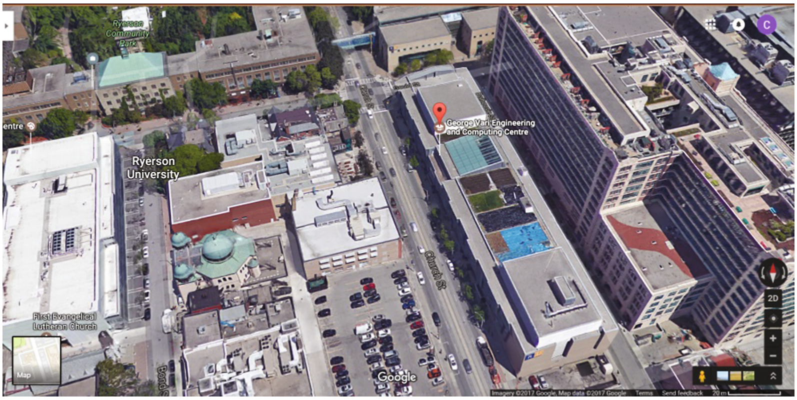


Fig. 6. 3D virtual geographical map taken from Google Earth of the area around Ryerson University, Toronto, Canada [26].

The scale of this model is imported with a unit 1, and all game objects (UAV model and shrapnel model) are scaled proportionately in reference to this terrain. It should be noted that special care should be taken when scaling all models accordingly, as the implementation of explosive dispersal of solid particles is directly affected by the size of the shrapnel model.

UAV Model

The simulated UAV is modeled after the Freefly ATLA 8, capable of delivering a 10 kg payload and providing a standoff distance of 5 km (Fig. 7).

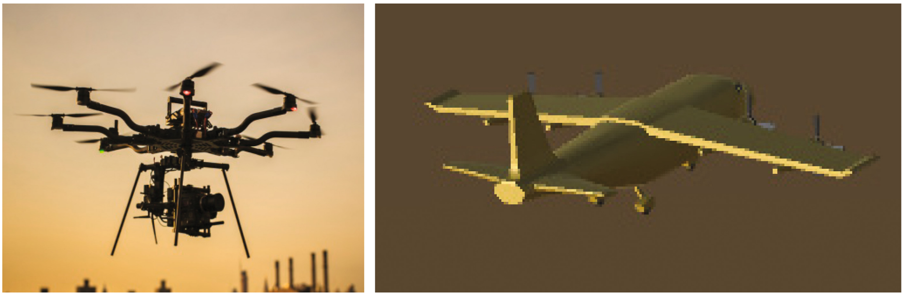


Fig. 7. Left: Freefly ATLA version 8, with 6 rotors [27], represented by the image on the right: a standard mini airplane model to fly to a location and deliver a payload.

One UAV model is imported into the game and simulates simple flight capability - maintaining its relative altitude at all times. The UAV model operates only to deliver the payload to a desired height and location above the 3D terrain. The UAV model's current capabilities are sufficient for the purposes of our simulation, however additional features, such as the ability to evade detection, may be added later as future work.

Shrapnel Models

In our simulation, shrapnel is represented as a spherical game object, equipped with mesh tight bounding boxes. We assume that if any of the 100 shrapnel game objects collides with the 3D terrain, then it is considered a significant point of impact. The spheres moves in the simulation according to user-defined volume, pressure, force and directional values [6] and are configured to detect collisions to the mesh of the 3D terrain.

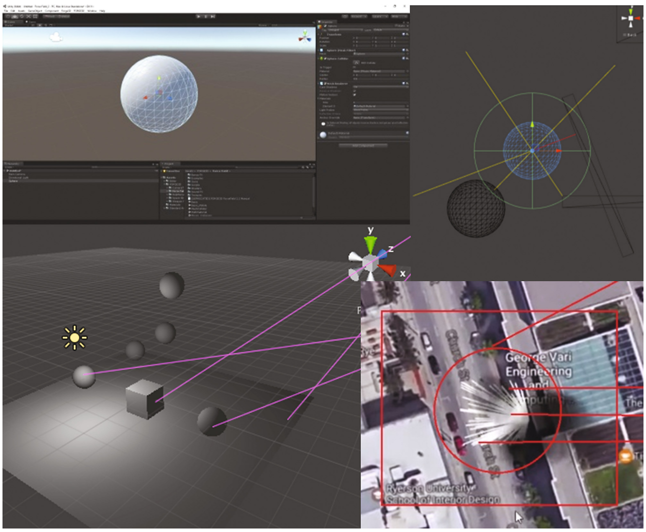


Fig. 8. Left: Freefly ATLA version 8, with 6 rotors [27], represented by the image on the right: a standard mini airplane model to fly to a location and deliver a payload.

We assert that the point of collision with the 3D terrain suggest a likely point of impact in a real-world environment. An additional bounding box with a radius of 0.5 km is marked after the explosion occurs, and is provided as a potential visual indication of shrapnel impact. Caution should be used in the interpretation of the marked bounding box, as it may not necessarily encompass all shrapnel surface impacts – this actual real-world evaluation requires further empirical study that is beyond the scope of this paper (Fig. 8).

DFC Implementation

Shrapnel game objects are grouped together representing the canister of a DFC. No actual model is in place for the charge, the explosion is simulated by the algorithm attached to each shrapnel game object. The entirety of these objects is placed under the UAV model which move in the simulation as a unit (Fig. 9).

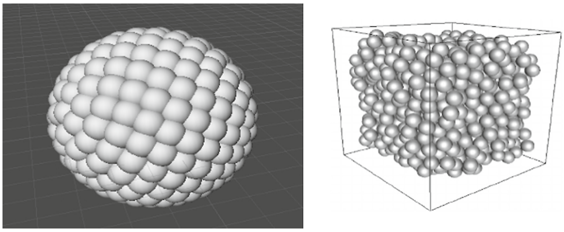


Fig. 9. Left: Multiple spherical unity game objects representing shrapnel, grouped together as a sphere. Right: Shrapnel game objects enclosed in another game object representing the DFC's 'canister'.

3.3 Gameplay Modes

Pre-explosion Flight Mode

The user controls and directs the UAV model to any point on the 3D terrain, real-time geospatial information is provided on a "debug panel" such as longitude, latitude, distance above ground, and geographic coordinates (e.g. Degrees, minutes, and seconds (DMS): 41°24′12.2″N) for the user to accurately place the UAV at a specific point above the 3D terrain. The user can activate the explosion and the subsequent explosive dispersal of shrapnel game objects which will automatically compute and simulate their trajectories and possible collision points with the mesh of the 3D terrain.

Post-explosion Exploration Mode

After the explosion occurs, users are able to move the in-game camera anywhere above the 3D terrain to analyze and view the impact points more closely. All the impact points are recorded as a list of geographical coordinates with their corresponding elevation above ground that the game object collided with. For example, a typical collision point may be somewhere along a wall of a high-rise building in an urban area (Fig. 10).

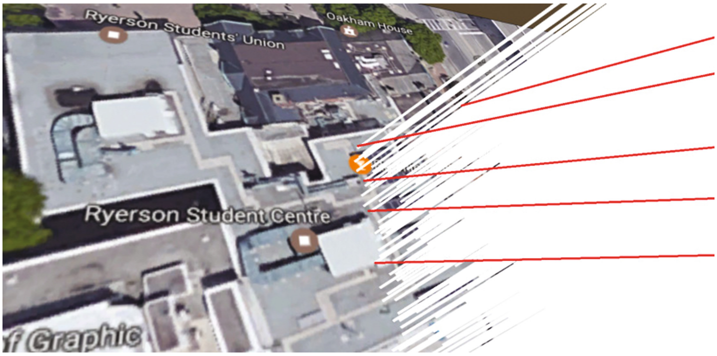


Fig. 10. Exploration of the possible area and points of impact of shrapnel after an explosion occurs.

4 Results

The table suggests a possible list of geographical coordinates of shrapnel impact points when a UAV and its explosive package was at 43°39′28.4″N 79°22′40.2″W, Longitude: 43.657890, latitude: 79.377825 and hovering at 0.039 km above the ground (Table 1).

Table 1. Suggested short list of recorded shrapnel game objects that collided with certain geographical coordinates.

Shrapnel object ID	Geographical coordinates	Distance above ground (km)
1	43°39′29.0″N 79°22′40.8″W 43.658060, −79.377991	0.005
2	43°39′28.6″N 79°22′40.8″W 43.657943, −79.377990	0.004
3	43°39′27.9″N 79°22′40.5″W 43.657762, −79.377902	0.010
4	43°39′29.2″N 79°22′39.8″W 43.658100, −79.377709	0.003
5	43°39′28.4″N 79°22′39.2″W 43.657899, −79.377555	0.011
6	43°39′27.9″N 79°22′39.4″W 43.657748, −79.377619	0
7	43°39′27.7″N 79°22′40.0″W 43.657682, −79.377791	0.004
8	43°39′27.1″N 79°22′40.2″W 43.657534, −79.377823	0.003
9	43°39′27.2″N 79°22′40.7″W 43.657545, −79.377973	0.002
10	43°39′27.9″N 79°22′40.9″W 43.657758, −79.378040	0.001

Based on our simulation for this particular scenario, we have estimated possible geographical coordinates of shrapnel impact points of a given UAV guided DFC explosion. The validity of the simulation will require further empirical analysis, but should be sufficient for a proof-of-concept theoretical application of IED shrapnel dispersal and impact on 3D terrain.

5 Conclusion and Future Work

We have proposed a simulation that suggests possible reasoning for a UAV guided DFC IED attack on any target given an arbitrary 3D terrain surface. We assert that this simulation is an approach towards a high-fidelity UAV guided IED explosive dispersal of shrapnel analysis simulation, which has yet to be fully conceived, given the complexity of the forces and many unknown elements at play. The simulation provides ample room for additional factors, which are relevant to calculating a more accurate shrapnel trajectory and impact onto 3D terrain, to be implemented along with the current features.

Future work may include verification of the use of a closed system fluid dynamics flow topology algorithm, and an empirical study with a real drone, a simple reproducible controlled explosive, and a payload of trackable inert and other debris.

References

1. Chan, C.C., Ferworn, A., Young, C.M.: Serious gaming for improvised explosive device neutralization training. In: MATEC Web of Conferences, vol. 68, p. 02006 (2016)
2. Sullivan, J.M.: Revolution or evolution? The rise of the UAVs, in technology and society. In: Proceedings of 2005 International Symposium on Weapons and Wires: Prevention and Safety in a Time of Fear, ISTAS 2005, pp. 94–101 (2005)
3. Miasnikov, E.: Threat of terrorism using unmanned aerial vehicles: technical aspects, Moscow, Russia: Center for Arms Control, Energy, and Environmental Studies, Moscow Institute of Physics and Technology (2005)
4. Mandelbaum, J., Ralston, J., Gutmanis, I., Hull, A., Martin, C.: Terrorist use of improvised or commercially available precision-guided UAVs at stand-off ranges: an approach for formulating mitigation considerations, DTIC Document (2005)
5. Vish, J.A.: Guided Standoff Weapons: A Threat to Expeditionary Air Power, DTIC Document (2006)
6. Zhang, F., Frost, D., Thibault, P., Murray, S.: Explosive dispersal of solid particles. Shock Waves **10**(6), 431–443 (2001)
7. U. G. Engine: Unity Game Engine-Official Site, pp. 1534–4320. http://unity3d.com. Accessed 9 Oct 2008
8. Mays, G., Smith, P.D.: Blast Effects on Buildings: Design of Buildings to Optimize Resistance to Blast Loading. Thomas Telford, London (1995)
9. Directional charge explosive device: Google Patents (1969)
10. Directional warhead: Google Patents (1971)
11. Campion, T.: Family of Multi-Purpose Infantry Munition Systems MPIMS. Small Arms Defense J. (2015)
12. Nealen, A., Müller, M., Keiser, R., Boxerman, E., Carlson, M.: Physically based deformable models in computer graphics. Comput. Graph. Forum **25**(4), 809–836 (2006)

13. Yngve, G.D., O'Brien, J.F., Hodgins, J.K.: Animating explosions. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 29–36 (2000)
14. Neff, M., Fiume, E.: A visual model for blast waves and fracture. Graph. Interface **99**, 193–202 (1999)
15. Reeves, W.T.: Particle systems—a technique for modeling a class of fuzzy objects. ACM Trans. Graph. (TOG) **2**(2), 91–108 (1983)
16. Sims, K.: Particle animation and rendering using data parallel computation. Newslett. ACM SIGGRAPH Comput. Graph. **24**(4), 405–413 (1990). ACM
17. Wan, T.R., Tang, W.: Explosive impact on real-time. In: Vince, J., Earnshaw, R. (eds.) Advances in Modelling, Animation and Rendering, pp. 247–265. Springer, New York (2002)
18. Popat, N., Catlin, C., Arntzen, B., Lindstedt, R., Hjertager, B.H., Solberg, T., Saeter, O., Van den Berg, A.: Investigations to improve and assess the accuracy of computational fluid dynamic based explosion models. J. Hazard. Mater. **45**(1), 1–25 (1996)
19. Wang, Y., Beom, H., Sun, M., Lin, S.: Numerical simulation of explosive welding using the material point method. Int. J. Impact Eng. **38**(1), 51–60 (2011)
20. Selle, A., Rasmussen, N., Fedkiw, R.: A vortex particle method for smoke, water and explosions. ACM Trans. Graph. (TOG) **24**(3), 910–914 (2005)
21. Rami, N., Proctor, M.D.: Real time physically-based modeling and simulation of cratering and fragmentation of terrain. Simulation **83**(12), 830–841 (2007)
22. Melosh, H., Ryan, E., Asphaug, E.: Dynamic fragmentation in impacts: hydrocode simulation of laboratory impacts. J. Geophys. Res. Planets **97**(E9), 14735–14759 (1992)
23. Cioppa, T.M., Lucas, T.W., Sanchez, S.M.: Military applications of agent-based simulations. In: Proceedings of the 36th Conference on Winter simulation, pp. 171–180 (2004)
24. Smith, R.D.: Essential techniques for military modeling and simulation. In: Proceedings of the 30th Conference on Winter Simulation, pp. 805–812 (1998)
25. Schmidt, R.M., Housen, K.R.: Some recent advances in the scaling of impact and explosion cratering. Int. J. Impact Eng. **5**(1–4), 543–560 (1987)
26. Lisle, R.J.: Google Earth: a new geological resource. Geol. Today **22**(1), 29–32 (2006)
27. Brouillette, M.: Heavy lifting drones fill a niche. Mech. Eng. **139**(2), 23 (2017)

Demand Side Management Using Bacterial Foraging and Crow Search Algorithm Optimization Techniques

Almas Tabssam, Komal Pervaz, Arje Saba, Zain ul Abdeen, Mashab Farooqi, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. Energy is the most valuable resource in every day life. However, energy demand is going high day by day. The high consumption of energy causes series of energy crisis. This problem can be handled with many optimization techniques by integrating demand side management with traditional grid. The main purpose of demand side management is to reduce the peak load and smart grid targets reduce the electric cost and load management by shifting the load from on peak hours to off peak hours. In this work, I adopt the Bacterial Foraging Algorithm (BFA) and Crow Search Algorithm (CSA). Simulation results show that our propose techniques reduce the total cost and peak average ratio by scheduling the load for 24 h. Results show that BFA is perform better than CSA and archived the objectives.

1 Introduction

Electricity bring revolution in human life and used in almost every field. Every domestic and industrial machine or electric appliance needs electricity to perform their functionality. There are many natural and other sources to generate electricity like hydroelectric power, wind power and solar power. Smart Grid (SG) in Fig. 1 supply the electricity in domestic and industrial areas in smart and intelligent way by using different communication technologies.

In paper [1], usage of electricity is increased in domestic and industrial areas. For this utility educates their customers for electricity usages and about electricity wastage. To control and manage the demand of electricity, utility charge the customers high rate of electricity in peak hours.

The demand of electricity is increasing due to rapid increase of human population. To tackle this problem utility need to generate more power that is very costly and increase the emission of co2. So for this demand side management (DSM) [2] play an important role to manage the electricity usages. In [3], Utility adopt different strategies to manage the load in peak hours by charging high electricity rates or offer some incentives to use electricity in off peak hours. There are following objective of DSM

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_11

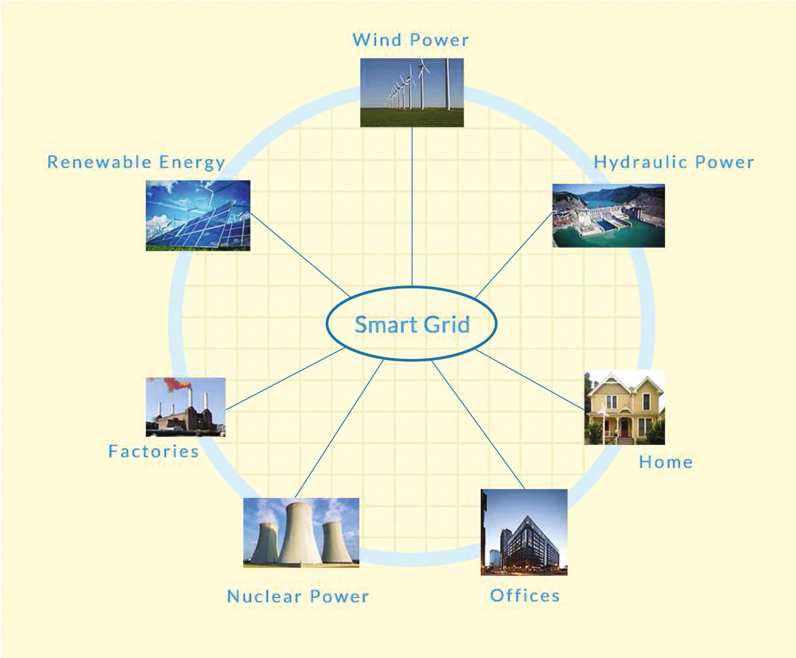


Fig. 1. Smart grid

(a) Reduce peak curve, directly control the curve by load balancing.
(b) Reduce cost, by scheduling the load cost is reduced.
(c) Increase user comfort, reduce the waiting time for each appliance.
(d) Reduce peak average ratio (PAR), by shifting load from peak hours to off peak hours.

Utility introduce different price signals which are vary from country to country and season to season. There are following price signals

(a) Time of Use (TOU)
(b) Real Time Pricing (RTP)
(c) Critical Peak Pricing(CPP)
(d) Day ahead pricing (DAP)
(e) Incline Block Rate (IBR)

TOU and RTP are most the popular pricing signal used in domestic areas. We have used the RTP signal to calculate the cost of electricity for 24 h. Many techniques in Artificial Intelligence (AI) are proposed to schedule the load. We get an optimal solution by scheduling appliances and reduces the electricity cost and PAR.

This paper presents the HOME Energy Management System (HEMS) in Fig. 2 that manage the load in customer side. HEM System consist of following components

(a) Smart Appliances
(b) Smart Meter
(c) Data Collection Point (DCP)
(d) Utility

By using HEM System we send the our daily appliances load to utility by using smart meter. In [4] utility provides the electricity according to load information that is received from smart meter. In [5–7], integer linear programming, non integer linear programming and dynamic programming techniques are used to reduce the electric cost and energy consumption. However these techniques are not handle the large number of appliances. We used two techniques BFA and CSA to overcome the deficiencies of previous techniques and full fill the objectives. Rest of the paper divided in following sections: Sect. 2 represents related work, Sect. 3 represents proposed system, Sect. 4 represents simulation and Sect. 5 represent the conclusion.

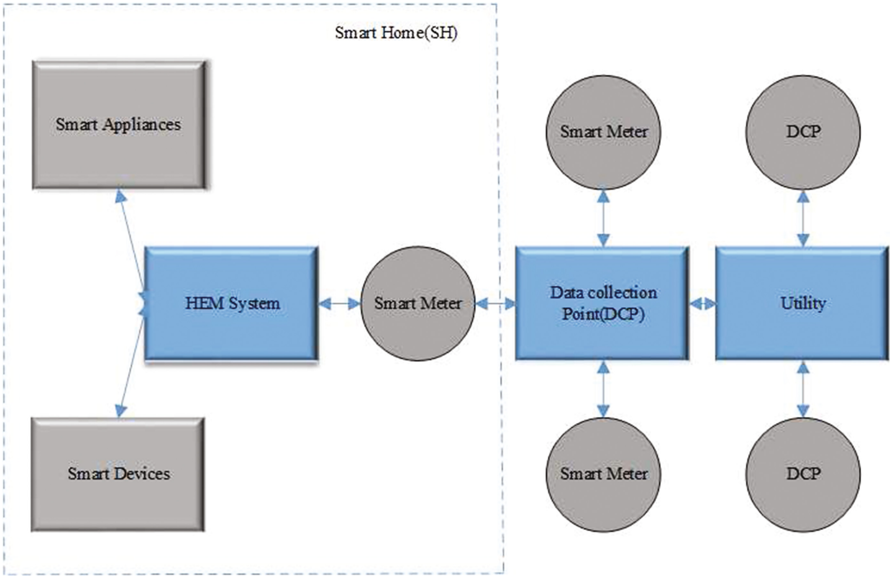


Fig. 2. Home Energy Management System

2 Related Work

In past years many of the researchers have worked on HEMS and proposed different optimization techniques to get remarkable improvement in smart grid. In this section, existing literature and different optimization techniques are discussed.

In paper [6], author addressed the problem of peak load in peak hours. In on peak hours if the load is high then user pay the maximum cost in those specific hours according to tariff rate provided by supplier company. So the author objective is to decrease the load in on peak hours and schedule the load in such a way that load is shifted to off peak hours. For this purpose author used integer linear programming (ILP) based optimization technique that schedule the load for whole day and balance the load for 24 h. This technique schedules the load and appliances according to 24 h format and try to achieve the user comfort. Proposed technique achieved the target to balance the load for 24 h and minimize the cost. For achieving the goal author ignores the user comfort/waiting time that is the limitation of this paper.

In [7], author addressed the problem of electricity cost and PAR in home area network (HAN). For this purpose, author used the technique genetic algorithm (GA) and introduces the energy management controller (EMC) for energy management of a home. By applying the GA technique author schedule the load in such a way that loads is shifted from on to off peak hour. Author achieved the goals by adopting the combined pricing model of RTP and IBR and schedule power effectively that decrease electricity cost and PAR. For achieving the goal author ignores the user comfort/waiting time and renewable energy resources (RES) which are very important factor for cost reducing and power scheduling.

In [8], author addressed the problem of PAR, Electricity cost and user comfort. There are following techniques that author used in this paper: BFA, Wind Driven Optimization (WDO), Binary Particle Swarm Optimization (BPSO) and Hybrid Wind Driven Optimization Algorithm (GWDA). Author used these techniques to schedule the load and minimize the peak average ratio and increase the user comfort. In this paper, author used the RTP signal for calculate total cost of electricity. For achieving the goal author can not integrate the RES that is the limitation.

In paper [9], author addressed the problem of Electricity Payment and User Discomfort. For this Author schedule the load for 24 h by categorizing the appliances in two types: (a) Fixed appliances (b) Shiftable appliances. The optimization problem is a convex optimization problem and can be solved independently for each appliance. So in that way author schedule the appliances one by one. Author used the "Day Ahead" price signal to minimize the electricity cost. It used three mode of operation for different appliances and maximizes the user comfort. For achieving the goal author ignore the PAR that is the limitation of this paper.

In [10], author addressed the problem of PAR, Electricity cost and user comfort by integrating the renewable energy resources (RES). In this paper, the techniques used are GA, ACO and BPSO. IBR and ToU are the pricing signal used to calculate the cost. Author divides the appliances in three categories: (a) Fixed Appliances (b) Shiftable Appliances (c) Elastic Appliances also mentioned the LOT for shiftable and elastic appliances. To achieve the objectives the author schedule the load in such a way that maximizes user comfort and minimize cost, PAR. Author ignores security and privacy issues between end user and utility that is the limitation of this paper.

In [11], author addressed the problem of electricity cost and power trading by integrating the RES. Dynamic programming technique is used to schedule the load and then integrate the RES to minimize the electricity cost. Also the author objective is trading the electricity with utility and sells the electricity to earn revenue. So in that way every user gains some market share to sell their extra electricity that is generated by RES. Extra power that is generated from new resources are sell to other on user desire price. By achieving the goal author ignore the PAR, user comfort/waiting Time that is the deficiency of this paper.

DR is very important in energy management, to fulfill the user electricity needs according to appliances load. For this utility give the incentive to end user, that user cannot use the appliances in on peak hours. Utility also give some other social pattern incentive and compel the user to reduce the electricity in on peak hours. RES in [12] is integrated to fulfill the user electricity needs. For this author used the dynamic programming technique for solve the problem of DR. ToU price signal is used in paper to calculate the electricity cost of every hour of day. According to this pricing signal author schedule the appliances to reduce the load from peak hours. The limitation of this paper author ignore the waiting time/user comfort and RES installation cost.

In this section, DSM is very important feature in smart grid. In [13] author addressed the problem to handle the load uncertainty and reduce the electricity payments. Optimal solution is used to schedule the load to reduce the load uncertainty. In this way, when load is schedule electricity cost is minimized and comfort is increased. In this paper IBR+RTP pricing signal is used to reduce the PAR. The solutions that author provided encourage the user to reduce electricity cost in peak hours to solve the problem of DSM. Author ignores the user comfort and RES that are the limitations of this paper.

In this section, paper [14], Author addresses the problem of electricity cost by measuring it with a day ahead pricing scheme and reduces the PAR. Author schedule the load for a day and shift the load from peak hour to other hours and reduce the electricity cost. For this utility communicate the user and inform them according to day by day price signal. So utility offer some incentive to user to shift their load to peak load to off peak load. Author also gives some strategy to utility to reduce the installation cost. Author ignores the waiting time and RES installation Cost in this paper.

Author addressed the problem of Demand Response (DR) in [15] by integrating the RES to minimize the electricity cost. In this paper author monitor the load of appliances and check the electricity tariff of the utility and then schedule the appliances. For reducing electric cost and shift the load author give the incentive to user according to available pricing scheme. Author also integrates the RES to meet the DR of user and also supply to other user. When RES integrated author made decision when to store electricity in off peak hour. In this paper author ignore the user comfort and RES installation Cost, which are the limitation of this paper.

Author addressed the problem of DSM by integrating the new electric resources in [16]. For this author calculate the future need of electricity of user and try to meet the demand of the user. Load is different according to region because the weather is not same in all regions. So user electricity demand is changed according to region and loads vary between regions to region. To meet the user demand of electricity author used the RES. RES generated power is used to compensate the user demand in peak hour. In that way author mange the appliances and used RES energy to reduce the cost. RES installation cost is ignored in this paper that is the deficiency of this paper.

In paper [17], author addressed the problem of the DSM and its future requirement and also discussed the RES. In this paper author optimize the electricity generated distribution and distribute the power to HEMS. It provides the synchronization from generator to grid supply and maintains the voltage. For this Bus system is used in grid for distribution of power to house hold user in three page and single phase. For this reactive power is generated to achieve the single and multi objective goals. So this strategy is very useful for utility and planner to distribute the power to end user. So in this way author manage demand side load management to meet the load demands of user. So in this paper, author addresses the DSM problem by ignoring these factors user comfort, PAR and electricity cost.

3 Problem Statement

SG provide the communication channel between utility and customer. Customer used the electricity according to their needs and utility provide the electricity as per demand of users. DSM play an important role between utility and customer to manage the load in peak hours by giving some incentives or educating customers. DSM make energy efficient, dynamic response and dynamic demand depends upon the customers.

DSM handle the load in such way to reduce the load in peak hours. Different appliances of high voltage and low voltage are used in a single home. Refrigerator, motor pump are high voltage appliances where fan, light and computer are low voltage appliances. DSM schedule these appliances to reduce the load in peak hours. DSM also control electricity bill and overcome the extensive use of energy. Some of schedulers are try to minimize the cost in peak hours and neglect the PAR by shifting the load in off peak hours. Other schedulers are also left one or more time slot empty and no appliances scheduled in these time slots. In such cases, system can not be validated and these type of schedulers make the burden on utility that lead to starvation.

4 Proposed System Model

We minimize the total cost as explained in Eq. 1 by shifting load from on-peak hours to-off peak hours. We used the RTP signal for calculating electricity price and power consumption. Equation 2 is used to calculate the overall load. Equation 3 is used to calculate the load of every time slot of hours. Table 1 represents the appliances load and there power consumption.

Table 1. Detail description of appliances

Appliances types	Name	Power in kilowatts
Non interruptible appliances	Washing_Machine	0.78
	Dish_Washer	0.36
	Cloth_Dyer	4.40
Brust load interruptible appliance	Space_Heater	1.50
	Air_Conditioner	1.44
	Refrigerator	0.73
	Water_Heater	4.45
Fixed appliances	Lighting	0.6
	Fans	0.75
	Cloths_Iron	1.5
	Microwave_Oven	1.18
	Toaster	0.5
	Coffee_Maker	0.8

$$\text{Cost} = \sum_{hour=1}^{24} (E_{rate}^{hour} \times P_{Rate}^{App}) \tag{1}$$

$$Ff = \min \begin{cases} l_{od}^{i \in N_p} \geq mean(L_{od}^{Us}) \quad E_{Rate}^{Hour} \leq mean(E_{Rate}) \\ l_{od}^{i \in N_p} \geq std(L_{od}^{Us}) \\ l_{od}^{i \in N_p} \leq mean(L_{od}^{Us}) \quad E_{Rate}^{Hour} > mean(E_{Rate}) \end{cases}$$

$$L_{od}^{S} = \sum\nolimits_{hour=1}^{24} l_{od}^{hour} \tag{2}$$

$$l_{od} = P_{rate}^{app} \times App \tag{3}$$

$$G1 = min(Cost) \tag{4}$$

$$G2 = min(PAR) \tag{5}$$

$$G3 = min(L^s od) \tag{6}$$

$$PAR = \frac{\max(L^s od)}{Average(L^s od)} \tag{7}$$

There are many optimization techniques, which are used to find the best optimal solution in swarm optimization. Some of the AI techniques are used to solve the optimization problems, in which BFA and CSA are the most popular techniques. These are inspired from food foraging and animals behavior for searching the food. BFA is nature inspired algorithm, which is based on food foraging of real bacteria. CSA is also nature inspired algorithm based on crow food foraging behavior.

4.1 BFA

From its natural behavior of food foraging, some of the animals have poor strategies to find the food and other one have strong strategies to find the food. In multiple population generation, animal who have poor strategies are replaced with strong one. There are three steps to achieve this (1) Chemotaxis its represent the length of life time of bacteria by calculating the chemotatic steps. Two basic operations are performed by real bacteria: i) Swim ii) Tumble. Cost of fitness is calculated by one bacteria position to new position and measured chemotatic steps from one position to new one position where food is found. (2) Reproduction on the base of best fitness cost, bacteria is contributed in the next generation. (3) Elimination and dispersal: bacteria that has low fitness cost is discarded and new bacteria is produced randomly in new population.

4.2 CSA

Crows are considered very intelligent and clever bird and they have large brain according to their body. Crows are remembering the other bird faces and warn them if they act unfriendly. From its natural behavior, crow food foraging technique is very popular for swarm optimization. Steps perform during food foraging are (1) Crows are initialize memory position find the food. (2) Evaluate the fitness and cost of position. (3) Generate new position. (4) Check the feasibility of new position. (5) Evaluate the fitness of new position. (6) Update memory

5 Simulation and Results

In this section, we evaluate the performance of our purposed system. Main objective of this simulation is to reduce the electric cost and PAR by load blanching on DSM with RTP signals. Figure 3 show the schedule and unscheduled load on grid for single home for 24 h. We used RTP tariff that has high price during peak hours and vary according to demand. We clearly see in Fig. 3 peak load of schedule that are minimum as compare to unscheduled load and reduce the cost. Total cost is demonstrated in Fig. 4 that represent the effectiveness of our proposed system. BFA and CSA reduced the total cost as compare to unscheduled load. Figure 5 shows the cost per hour of a day where unscheduled load create the max peak curve as compare to BFA and CSA schedule load and has the max price signal. BFA technique reduced the maximum cost which is best

Algorithm 1. BFA for Smart Grid Scheduling

Initialization (PoP,Np,Ne,Nr,Nc,Ns,Ci)
Evaluate the initial PoP
$Jlast \leftarrow Ji$
for $l = 1 \rightarrow Ne$ **do**
  for $k = 1 \rightarrow Nr$ **do**
    for $j = 1 \rightarrow Nc$ **do**
      for $i = 1 \rightarrow Np$ **do**
        Find the new position $\Theta_i[j, k, l]$ for PoP
        Compute the fitness $Ji[j, k, l]$ for PoP
        for $s = 1 \rightarrow Ns$ **do**
          if $Ji < Jlast$ **then**
            Jlast←Ji
            geto step4
          end if
        end for
      end for
    end for
    Find the fitness of Θ_i
    Select the gbest
  end for
  Elimination and dispersal step
end for

Algorithm 2. CSA for SG scheduling

Randomly initialize the position of a flock of N crows in the search space
Evaluate the position of the crows
Initialize the memory of each crow
while $iter < iter_{max}$ **do**
  for $i = 1$:$N(allNcrowsoftheflock)$ **do**
    Randomly choose one of the crows to follow (for example j)
    Define an awareness probability
    if $r_j \geq AP^{j,iter}$ **then**
      $x^{i,iter+1} = x^{i,iter} + r_i \times fl^{i,iter} \times (m^{i,iter} - x^{i,iter})$
      else
      $x^{i,iter+1}$ = a random position of search space
    end if
  end for
  Check the feasibility of new positions
  Evaluate the new position of the crows
  Update the memory of crows
end while

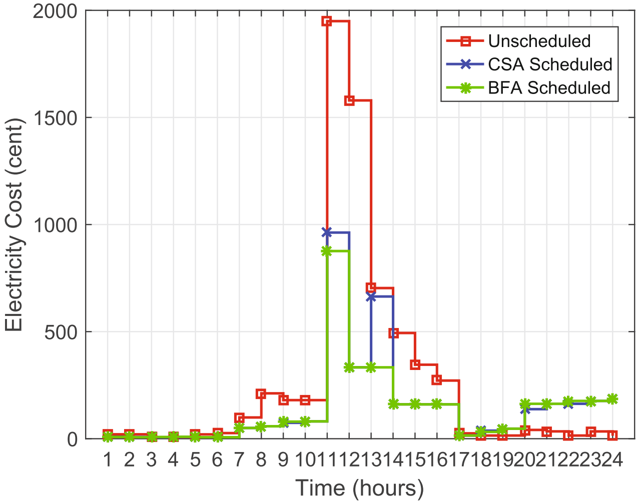


Fig. 3. Electricity cost during 24 h

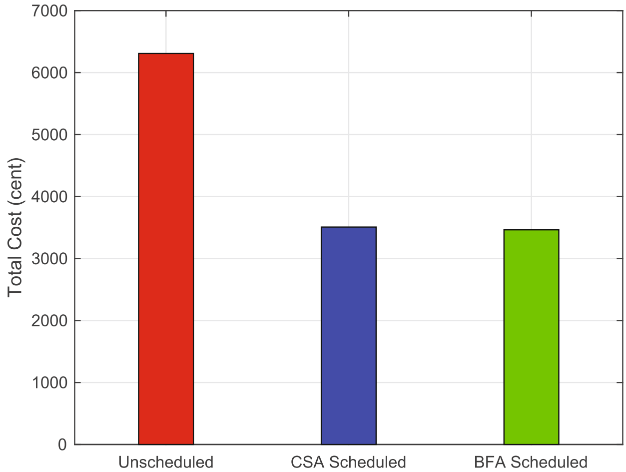


Fig. 4. Total cost per day

from CSA. Figure 6 shows the PAR results, PAR of CSA is low but cost is high as compare to BFA so there is trade off between cost and PAR. PAR of BFA is low from both CSA and unscheduled which show the effectiveness of BFA. Figure 7 shows the waiting time of two techniques BFA and CSA, Where the waiting time of BFA is greater than the waiting time of CSA. Electric cost is reduced if the waiting time is high because user wait for specific time where no peak curve is generated so it bear the low cost in off peak hours. So there is a trade off between waiting time and electric cost. If the waiting time of appliance

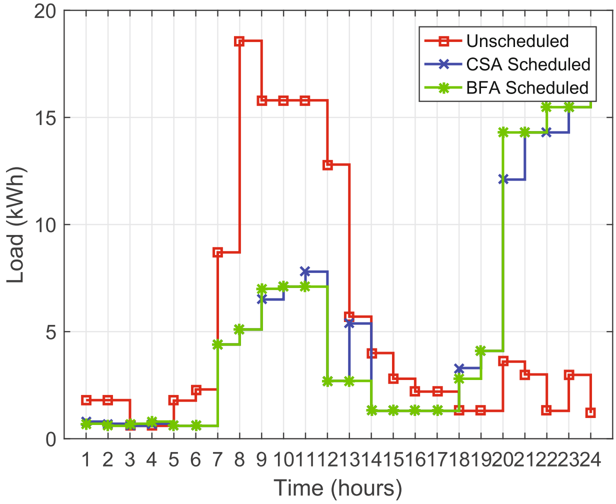


Fig. 5. Load of a home during 24 h

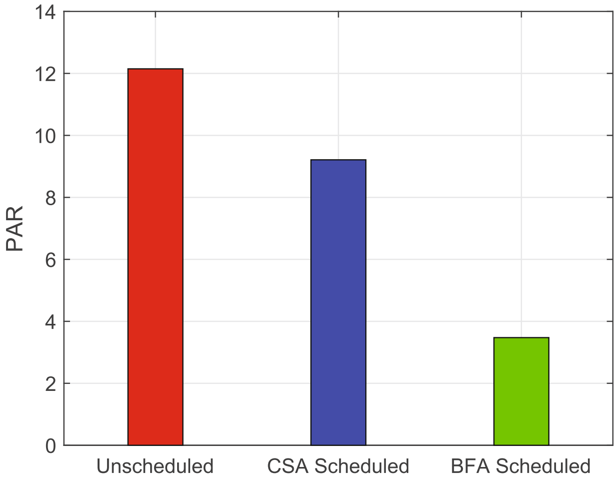


Fig. 6. Peak average ratios

is low and users switch on the device with short delay then they charge high electric cost. These results are shown that our proposed techniques archived the objectives and BFA perform outstanding than CSA. BFA show the influence on CSA and reduce the high cost and high PAR.

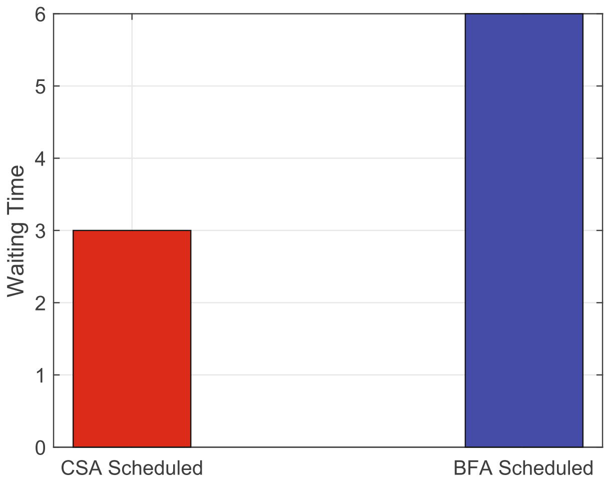


Fig. 7. Average waiting time of two group

6 Conclusion

DSM is very effective to schedule the load in HEM system and blanche the load. Most of the user shifted the load on renewable energy resources that are not enough to decrease the cost. Many of the AI techniques are used to blanche the load in peak hours. For this purpose, We used two techniques BFA and CSA to schedule the load. When we schedule the load and shift the load in off peak hours, Create peak in off peak hours and charge high RTP. Our proposed techniques perform well and results show that BFA perform outstanding than CSA. However, there is trade-off between cost and PAR. BFA results show that it gives optimal solution and has fast convergence rate then CSA. In future, we work on hybrid technique and also integrate the RES to increase the user comfort.

References

1. Gul, M.S., Patidar, S.: Understanding the energy consumption and occupancy of a multi-purpose academic building. Energy Build. **87**, 155–165 (2015)
2. Palensky, P., Dietrich, D.: Demand side management: demand response, intelligent energy systems, and smart loads. IEEE Trans. Ind. Inf. **7**(3), 381–388 (2011)
3. Rabiee, A., Sadeghi, M., Aghaeic, J., Heidari, A.: Optimal operation of microgrids through simultaneous scheduling of electrical vehicles and responsive loads considering wind and PV units uncertainties. Renew. Sustain. Energy Rev. **57**, 721–739 (2016)
4. Chiu, W.-Y., Sun, H., Poor, H.V.: Energy imbalance management using a robust pricing scheme. IEEE Trans. Smart Grid **4**(2), 896–904 (2013)
5. Chavali, P., Yang, P., Nehorai, A.: A distributed algorithm of appliance scheduling for home energy management system. IEEE Trans. Smart Grid **5**(1), 282–290 (2014)

6. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES, Innovative Smart Grid Technologies (ISGT), p. 15. IEEE (2012)
7. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
8. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
9. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
10. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
11. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
12. Ozturk, Y., Senthilkumar, D., Kumar, S., Lee, G.: An intelligent home energy management system to improve demand response. IEEE Trans. Smart Grid **4**(2), 694–701 (2012)
13. Samadi, P., Mohsenian-Rad, H., Wong, V.W.S., Schober, R.: Tackling the load uncertainty challenges for energy consumption scheduling in smart grid. IEEE Trans. Smart Grid **4**(2), 1007–1016 (2013)
14. Joe-Wong, C., Sen, S., Ha, S., Chiang, M.: Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility. IEEE J. Sel. Areas. Commun. **30**(6), 1075–1085 (2012)
15. Aghaei, J., Alizadeh, M.I.: Demand response in smart electricity grids equipped with renewable energy sources: a review. Renew. Sustain. Energy Rev. **18**, 64–72 (2013)
16. Pina, A., Silva, C., Ferrao, P.: The impact of demand side management strategies in the penetration of renewable electricity. Energy **41**(1), 128–137 (2012)
17. Alonso, M., Amaris, H., Alvarez-Ortega, C.: Integration of renewable energy sources in smart grids by means of evolutionary optimization algorithms. Expert Syst. Appl. **39**(5), 5513–5522 (2012)

Scheduling of Appliances in Home Energy Management System Using Elephant Herding Optimization and Enhanced Differential Evolution

Muhammad Azeem Sarwar, Basit Amin, Nasir Ayub, Syed Hassnain Faraz, Sajawal Ur Rehman Khan, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. In this study, problem of scheduling of appliances in Home Energy Management System (HEMS) is analyzed and a solution is proposed. Although there are many heuristic algorithms for solving the scheduling problem however we considered a swarm based heuristic algorithm Elephant Herding Optimisation (EHO). EHO uses the herding behaviour of elephants to handle the problem. To validate our research work, we simulate the single home with 12 appliances and scheduling is performed using EHO. We divided the appliances into two categories Interruptible and non-interruptible. Time of Use (TOU) pricing signal is used. Simulation results show that EHO is efficient as compare to Enhanced Differential Evolution (EDE) and unscheduled case. EHO technique is efficient in scheduling the appliances and reducing the waiting time.

1 Introduction

The demand of electricity is increasing day by day creating a shortfall. As the population is growing electricity needs are getting higher in residential area. The U.S. Dept. of Energy states that, since 1982 the growth in peak electricity usage is increasing every year by 25% [1]. Efficient utilization of energy is necessary to reduce the cost and demand of electricity in peak hours. Peak hours refers to period of energy consumption with high demand. These peak hours time slots are critical as the demand is high. Therefore, it is important to present a solution to this problem of high energy usage in specific peak hours. An intelligent system is required to transform the electricity usage in such a way that the user demand is fulfilled without any challenge. Smart grids are power grid systems that consists of smart meters with the capability of sensing and measuring the power consumptions of the costumers. In smart grid the end user can adjust their load by scheduling the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_12

appliances in response to the given pricing scheme and peak loads. Figure 1 demonstrates smart grid. To meet the increasing demand of electricity lots of work is done by different researchers in HEMS. This management of electricity is referred as the Demand Side Management (DSM). DSM is the program which is being installed at end user side by utility companies to tackle the consumption of electricity. End user can manage their electricity demand by scheduling the appliances in such a way that they meet their basic demand of electricity with minimum cost or maximum user comfort. To achieve these scheduling goals different heuristics techniques are used to schedule the appliances in residential area. The heuristic algorithms such as Genetic Algorithm (GA), Harmony search Algorithm (HSA), EDE, Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) are few of them.

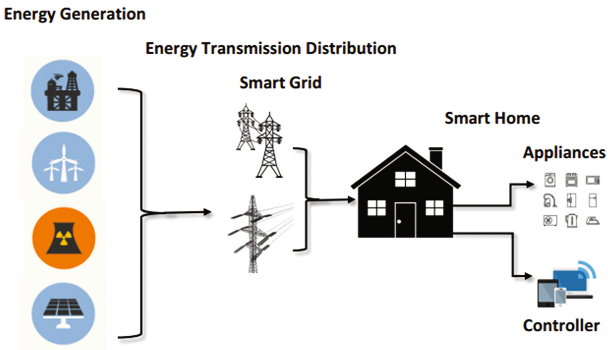


Fig. 1. Energy flow in smart grid.

Different types of pricing schemes are used by the utility companies for the electricity cost and billing. Real Time Pricing (RTP), Critical Peak Pricing (CPP) and Time of Use (TOU) are few pricing schemes that are being used. By using this scheme costumer can minimize their electricity cost with certain level of user comfort. In this research work we consider the time of use pricing scheme. TOU pricing varies based on the time of day and week customers use electricity. The rest of the paper is organized as follows. Section 2 contains study of related work in same domain. In Sect. 3 proposed system model of the paper is presented. Section 4 discuses the simulations and results of the proposed model. At the end in Sect. 5 this research work is concluded.

2 Related Work

Many researchers around the globe worked in smart grid and Energy Management System (EMS), to optimally schedule the smart appliances in recent years. Related work is categorized on the basis of techniques: (i) Heuristic algorithms (ii) Linear Programming (iii) Game model and other optimization algorithms. In this regard few research papers are discussed below.

A DSM controller for residential area in smart grid using RTP scheme and heuristic algorithms is designed [2]. In designed controller, different heuristic algorithms are applied. Authors proposed hybrid GWD. The objective of the author is to reduce the cost and maximize the User Comfort (UC). The results shows electricity cost is minimized and the author achieved the desired UC using proposed approach. Energy Management Controller (EMC) circumvents the peak formation while considering the electric utilities, electricity bill reduces and also maintains the acceptable user comfort [3]. The author also integrated RES in HEMS. Author used unique bill calculation method with two pricing schemes TOU and IBR. The results shows that proposed system reduced the electricity bill and minimized the PAR. As there is a tradeoff between cost and UC, UC is compromised. In [4], authors presented an approach for scheduling the electricity usage in home for reducing the electricity cost and PAR. The authors used GA and combined pricing strategy consisting RTP and IBR. Author divided one hour into five time slots with total 120 slots each day. The results satisfy that electricity cost and PAR is reduced. Authors proposed an approach for scheduling the residential user's appliances in [12]. A new pricing strategy Customer Incentive Pricing (CIP) is proposed. Authors uses heuristic optimization technique GA to find optimal solution. Flexible and non-interruptible appliances are considers in the paper. Authors uses 15 min intervals meaning 96 slots. A version of GA called genitor is used. Overall peak is reduced and cost of electricity is also reduced by the framework however UC is compromised.

The author used the linear programming technique [5]. The main objective of the research was to minimize the peak load. Liner programming language is used. Results verifies that author optimized peak load. UC and electricity cost is neglected in paper. In [9], optimal load control in industrial sector is investigated. The paper address the industrial load control under different pricing schemes like TOU, DAP, PP, IBR and CPP. Tractable mixed-integer linear programming optimization is used. Different case studies on steel mill are presented which shows that pricing signal effects the optimal demand response of industrial load. In [13], author designed the problem of selecting the customer for DR as Integer Liner Programming (ILP). Author considers that the DR scheduling algorithm can minimize the curtailment error. Results shows that ILP provides better solution with low error cases. Cost factor is neglected in this paper by the authors. Scheduling technique for the residential area users to obtain a trade-off in the power cost and comfort is presented [6]. Day ahead pricing signal is used in the paper. The trade-off between price and discomfort is achieved. Taguchi loss function is used for modeling the discomfort cost. Simulation results were compared with Flat pricing scheme and day ahead pricing scheme. The result shows that the authors achieved low electricity price and reduced the PAR and achieved the limited UC because of the trade-off between cost and UC.

Authors focus on scheduling the load and power trading with high penetration of RERs [7]. Game theory model approach is considered. Results shows that presented algorithm reduced the energy expenditure of the users. In [8], a bi-level programming approach is presented. Authors develop a bi-level model for peak minimization. Result verifies that an optimal trade-off in revenue generated and user cost can be obtained. Authors address efficient computation of sparse load shifting in the DSM and the scheduling problem of appliances in [10]. Firstly regulations are applied to re-model the DSM. Second, author develops a bi-directional framework for achieving the Nash equilibrium in DSM. Third, a fast gradient method is designed. The results shows that the presented algorithm by finding the Nash equilibrium reduces the PAR. Author neglected the cost of electricity in achieving user comfort. Energy management problem is addressed in [11]. The objective of the research is to reduce the electricity cost. Authors proposed an RTP based DR algorithm of achieving optimal load control. A Stackelberg game approach is used. Where there is one leader and many followers. EMC behaves like leader and each device as a follower. Author verifies that the proposed Stackelberg game-based DR (SGDR) successfully reduce the energy consumption in high pricing hours. Importance is not given to consumers in user comfort.

3 Proposed System Model

The demand for electricity is increasing day by day in residential area. Many techniques are implemented to meet the increasing demand of electricity however most of them neglected the UC. An efficient load management system is proposed for reducing the cost and increasing UC. As there is a trade off between cost and UC, a balance is achieved between both. This section emphasizes on the proposed approach for scheduling the power consumption of all appliances of home. The proposed approach is based on the TOU pricing scheme. The single home is considered in this approach to get the optimal solution. The appliances are categorized in two different types i.e. Interruptible appliances and Non-Interruptible appliances. Interruptible appliances are the appliances which can be shifted to any time slot of the day but non-interruptible appliances are the once that cannot be shifted to other time slots and have fixed start time and Length of operation Time (LOT). Table 1 shows the interruptible appliances and Table 2 shows the non interruptible appliances of single home along with their power ratings. In this approach, one day is divided into 24 equal slots. One slot has 60 min. Figure 2 shows peak hours in 24 hour time slot which are 11 to 19.

3.1 Enhanced Differential Evolution

EDE is heuristic algorithm that is used to solve the general optimization problem. EDE technique is the enhancement of the DE. In DE only one type of vector is calculated and in EDE five types of vectors are calculated to find the fitness value. The trial vector with minimum cost is the fitness function of EDE. The cross over rate in EDE is 0.3, 0.6 and 0.9. The fitness function calculated from trial vectors is further used to calculate the electricity cost of the appliances. The steps of EDE are;

Step 1: Initialization of population.
Step 2: Mutation.

Table 1. Parameters of interruptible appliances.

Appliances	Power (KWh)
Space heater	1
Heat pump	0.11
Portable heater	1.0
Water heater	4.50
Cloth washer	0.51
Cloth dryer	5.00
Dishwasher	1.20
First-refrigerator	0.50

Table 2. Parameters of non-interruptible appliances.

Appliances	Power (KWh)
Fan	0.5
Furnace fan	0.38
Central AC	2.80
Room AC	0.90

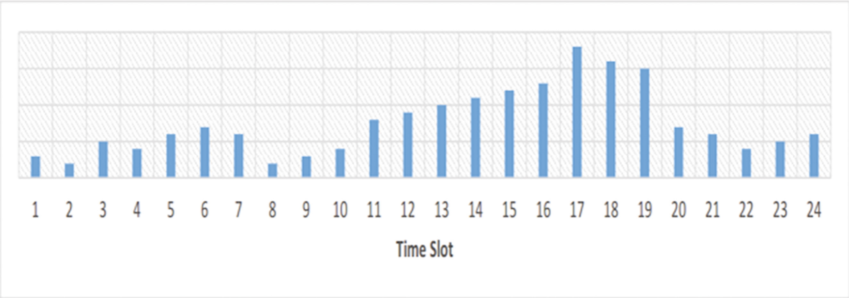


Fig. 2. 24 Hour time slot.

Step 3: Cross over.
Step 4: Calculate Trail vectors with their fitness function.
Step 5: Picking trial vector with minimum cost.
Step 6: Find the greatest individual in the population.
Step 7: End.

EMC is used in smart homes that communicates with the utility. The utility gets the information about energy consumption and user gets to know about the pricing scheme according to the hours. Two way communication between appliances and EMC is shown in Fig. 3.

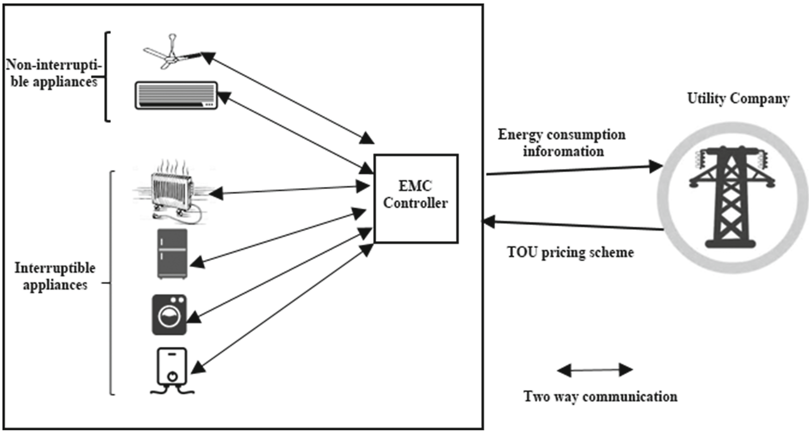


Fig. 3. EMC controller.

3.2 Elephant Herding Optimisation

EHO is a swarm based heuristic algorithm for finding the optimal solution in the search space. Herding behavior of elephants is considered as two operators [14]; first one is clan updating operator and second is separating operator. In this paper clan is considered as categories of the appliances where each elephant is represented as appliance. Two types of appliances interruptible and non-interruptible are considers as clans.

3.2.1 Clan Updating Operator

After separating the worst values from the population. The fittest values are updated using clan updating operator and worst values are discarded. The algorithm of updating operator is presented.

Algorithm 1. Updating Clan Operator

*[]

1: **for** $ci = 1 \rightarrow nClan$ **do**
2: **for** $j = nci \rightarrow nClan$ **do**
3: update xci,j and generate Xnew,ci,j
4: **if** xci,j=Xbest,ci **then**
5: update Xci,j and generate Xnew,ci,j
6: **end if**
7: **end for**j
8: **end for**ci

3.2.2 Separating Operator

The separating operator is used to separate the worst and fittest value. The worst elephant is separated from clan by using separating operator of EHO. The algorithm for separating operator is presented in Algorithm 2.

Algorithm 2. Separating Operator

*[h!]

1: **for** $ci = 1 \rightarrow nClan$ **do**

2: Replace the worst elephant in clan Ci

3: **end for**ci

4 Simulations and Results

In this section, the simulation results are discussed. The performance of EHO in home appliance scheduling is compared with EDE and unscheduled. TOU pricing scheme is used to implement the two schemes. The effect of EDE and EHO on electricity cost, PAR, load and user comfort is described below.

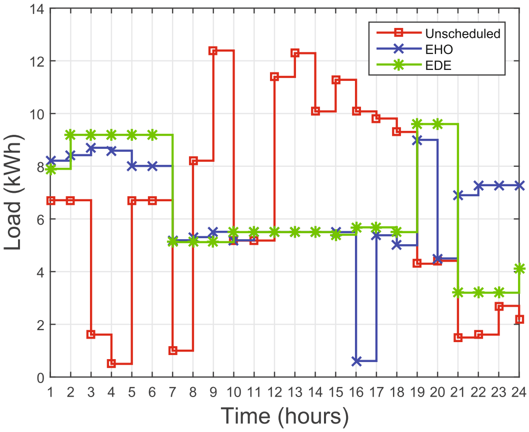


Fig. 4. Hourly electricity load.

Energy consumption of EHO with respect to EDE and unschedule is shown in Fig. 4. Load is managed by shifting the load in to off peak hours that reduce the energy cost. TOU pricing scheme is used. Users are charged according to the consumption of electricity in last week, last day or last hour. The figure shows the maximum energy consumption values are 12.4 kWh, 9.7 kWh and 9 kWh for unscheduled case, EDE and EHO respectively. The energy consumption in EDE is 78.2% and in EHO energy consumption is 72.5%. In case of EHO, Clan updater and separator operator is used and in EDE these result are achieved using trial vectors (Fig. 5).

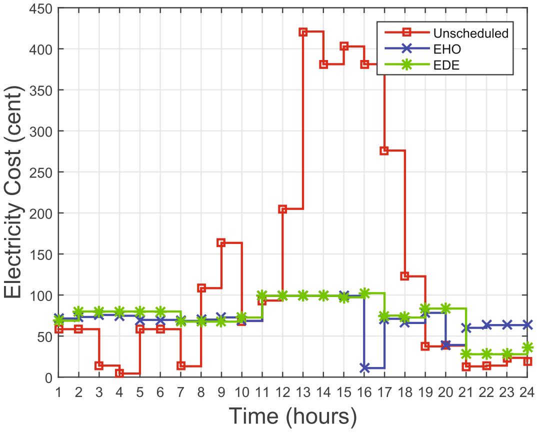


Fig. 5. Total electricity cost.

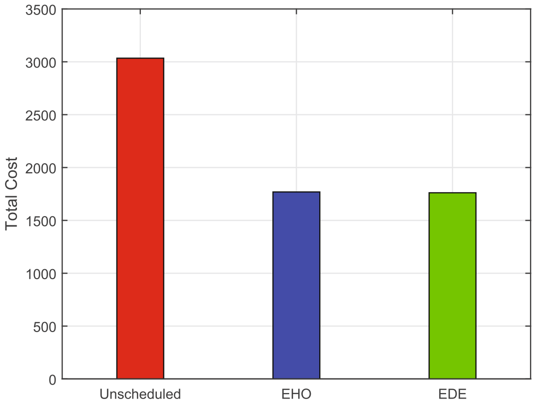


Fig. 6. Hourly electricity cost.

The amount of electricity cost of unscheduled case is 410 cents as shown in Fig. 6. Electricity bill reduces from 410 to 100 cents and 90 cents respectively in case of EDE and EHO. Electricity bill reduces to 24.39% in case of EDE and 21.95% in case of EHO. During peak hours, cost reduction is achieved by shifting the load from peak hours to off peak hours using EDE and EHO. EHO performs better than EDE in terms of electricity cost reduction because of updating the single best solution having maximum values and discarding all other solutions. In terms of EHO only fittest individuals are kept.

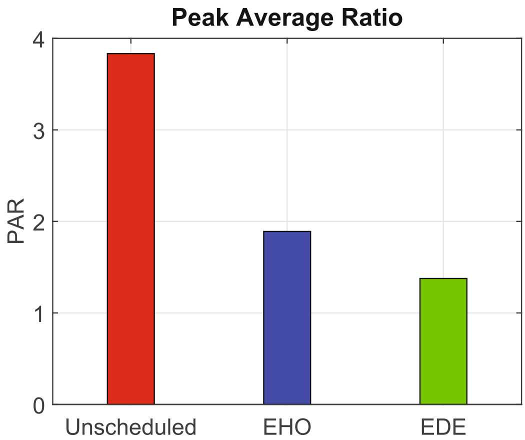


Fig. 7. Peak to average ratio.

Figure 7 shows the PAR performance of EHO, EDE and Unscheduled. As compared to unscheduled case, PAR of EDE and EHO is reduced. The PAR in EHO is 50% and PAR in EDE is 35%. PAR is increased in unscheduled case because the maximum appliances operates in high demanded electricity hours which increases the price during these hours and a peak is formed. By using EHO, EDE and TOU pricing scheme appliances are not operated during peak price most of appliances operated in start hours of day so PAR is not increased during peak hour. PAR graph for EDE and EHO displays that during on peak hours and off peak hours of the day power consumption of appliances is distributed without creating Peak.

The waiting time of EHO and EDE is shown in the Fig. 8. The figure shows that the waiting time of the EHO algorithm is less than the EDE algorithm. The waiting time of the EHO is 3 whereas the waiting time of the EDE algorithm

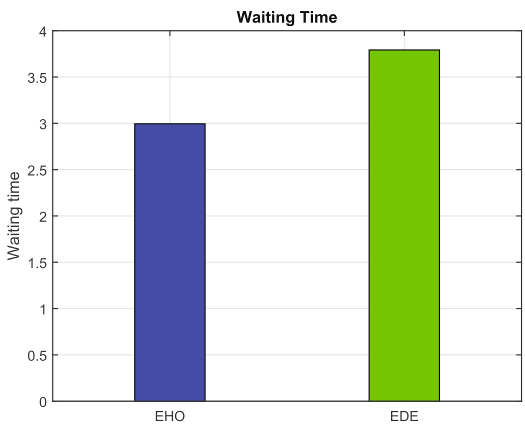


Fig. 8. Waiting time.

is 3.9. Waiting time is reduced to 3 min as in EDE it was almost 4 min. The waiting time of the EHO algorithm is less because we reduce the PAR of EDE by shifting the load of on peak hours to off peak hours. PAR of EHO is sacrificed when we reduce the electricity cost and waiting time of appliances. There is a trade off between waiting time with cost and PAR.

5 Conclusion

In this paper, an efficient load management system based on EHO optimisation algorithm is proposed. The herding behaviour of elephants i.e. clan updating and separating operator is used to generate the best schedule for appliances. For single home, two types of appliances are considered interruptible and non interruptible and TOU pricing signal is used. The simulations were performed and results are obtained showing EHO technique is performing better in reducing the waiting time and cost considering the trade off between both. Results of EHO technique is compared with unscheduled and EDE. EHO performs better in reducing the cost and waiting time. There is a trade off between PAR, cost and waiting time so PAR of EHO technique is greater than EDE. So surely proposed technique can be a reliable solution for scheduling, reducing cost and waiting time in HEMS.

References

1. Litos Strategic Communication: The Smart Grid: An Introduction. U.S. Dept. Energy, Washington, DC, USA, Technical report 560.01.04 (2009). http://energy.gov/oe/downloads/smart-grid-introduction-0
2. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**, 319 (2017)
3. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
4. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013). doi:10.1109/TSG.2013.2251018
5. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: IEEE PES Innovative Smart Grid Technologies (ISGT), Washington, DC, pp. 1–5 (2012). doi:10.1109/ISGT.2012.6175785
6. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electric. Power Energy Syst. **78**, 320–325 (2016). ISSN 0142-0615. http://doi.org/10.1016/j.ijepes.2015.11.099
7. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016). doi:10.1109/TSG.2015.2435708
8. Afar, S., Brotcorne, L., Marcotte, P., Savard, G.: Achieving an optimal trade-off between revenue and energy peak within a smart grid environment. Renew. Energy **91**, 293–301 (2016). ISSN 0960-1481. http://doi.org/10.1016/j.renene.2016.01.055

9. Gholian, A., Mohsenian-Rad, H., Hua, Y.: Optimal industrial load control in smart grid. IEEE Trans. Smart Grid **7**(5), 2305–2316 (2016). doi:10.1109/TSG.2015.2468577
10. Li, C., Yu, X., Yu, W., Chen, G., Wang, J.: Efficient computation for sparse load shifting in demand side management. IEEE Trans. Smart Grid **8**(1), 250–261 (2017). doi:10.1109/TSG.2016.2521377
11. Yu, M., Hong, S.H.: A real-time demand-response algorithm for smart grids: a Stackelberg game approach. IEEE Trans. Smart Grid **7**(2), 879–888 (2016). doi:10.1109/TSG.2015.2413813
12. Hansen, T.M., Roche, R., Suryanarayanan, S., Maciejewski, A.A., Siegel, H.J.: Heuristic optimization for an aggregator-based resource allocation in the smart grid. IEEE Trans. Smart Grid **6**(4), 1785–1794 (2015). doi:10.1109/TSG.2015.2399359
13. Kuppannagari, S.R., Kannan, R., Chelmis, C., Tehrani, A.S., Prasanna, V.K.: Optimal customer targeting for sustainable demand response in smart grids. Procedia Comput. Sci. **80**, 324–334 (2016). ISSN 1877-0509. http://dx.doi.org/10.1016/j.procs.2016.05.324
14. Wang, G.G., Deb, S., Coelho, L.D.S.: Elephant herding optimization. In: 3rd International Symposium on Computational and Business Intelligence (ISCBI 2015), Bali, pp. 1–5 (2015). doi:10.1109/ISCBI.2015.8

A New Meta-heuristic Optimization Algorithm Inspired from Strawberry Plant for Demand Side Management in Smart Grid

Muhammad Sufyan Khan, C.H. Anwar ul Hassan, Hazrat Abubakar Sadiq, Ishtiaq Ali, Asad Rauf, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. In recent years, different Demand Side Management (DSM) techniques have been proposed to involve users in decision making process of Smart Grid (SG). Power consumption pattern of shiftable home appliances is schedule to achieve desired benefits of high User Comfort (UC) and low energy consumption. In this paper, an Energy Management Controller (EMC) is designed by using two meta-heuristic algorithms: Strawberry Algorithm (SBA) and Enhanced Differential Evolution (EDE). The main objectives are electricity bill minimization, reduction in Peak to Average Ratio (PAR) and maximization of UC. However, there always exist a trade-off between cost minimization and UC maximization. Simulation results verify that, SBA perform better then EDE in terms of cost reduction while EDE perform far better than SBA in terms of UC maximization.

Keywords: Demand side management · Smart grid · Meta-heuristic techniques · Strawberry algorithm · Enhanced differential evolution · Real time pricing scheme

1 Introduction

Today, energy is the most valuable resource in term of human needs. Over the past decades, efficient use of energy have attracted extraordinary attention due to energy crisis. Therefore, new techniques and methods are being explored to fulfill the demand and supply gap of energy. However, energy demand is increasing day by day which causes serious problems. Electricity is one of the most valuable source of energy. Electrical energy is generated by converting different forms of energy such as, solar energy, nuclear energy etc. to electrical energy, which is then send to power grids to distribute it among the users. Traditional power grid is inadequate to fulfill challenges of power grid such as stability, reliability and

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_13

robustness [1]. Therefor new infrastructure is needed to handle these demands as shown in Fig. 1.

In this regards, Smart Grid (SG) is introduced. SG integrates sensors, communication technologies and control system with existing power grid to enable bi-directional communication between user and supplier. SG also include smart appliances, smart meter (SM) and other energy resources etc. Smart meter is used to exchange information of power consumption between utility and consumer. Energy optimization is based on this information. One of the most important aspect of SG is Demand Side Management (DSM) for energy optimization. DSM provide the best solution to maintain balance between supply and demand of the consumer. DSM has two main functions: Demand Response (DR) and load management [2]. In load management the main focus is the improvement of energy utilization management. It significantly reduced the chances of major distress and blackout. There are many benefits of load management such as reduction in peak to average ratio, efficient energy consumption by shifting the load from on-peak hours to off-peak hours, minimization in electricity bill and improve the performance such as reliability and stability of the power grid.

DR is defined as the responsive action taken by the consumer against dynamic price models. It provides many benefits to utilities, users and grid operation in financial as well as operational level. There is always a possibility of conflicts between demand and supply as grid condition can change any time. This rapid change is a threat to reliability and efficiency of the grid. Therefore DR is very important because it provide flexibility at comparatively low rate [2].

There are mainly two types of pricing rate schemes: dynamic pricing and flat rate pricing. Dynamic pricing is further characterized as Inclined Block Rate (IBR), Time of Use (TOU), Day Ahead Pricing (DAP), Critical Peak Pricing (CPP), RTP etc. These pricing schemes enable the consumer to shift their high load appliances from on pick hours to off peak hours. This shifting of load minimize the electricity cost and reduce PAR. From above mentioned pricing scheme RTP works more efficiently [4].

The main objectives of SG are minimization in electricity cost, reduction in gross power consumption, minimization in PAR and maximization in UC. To achieve these goals many DSM techniques are proposed in the recent years. In [3,6,7], integer linear programming (ILP), approximate dynamic programming and fractional programming are used in electricity cost minimization. Similarly in [9], author use Nash equilibrium for relatively large number of users up to 1000 and multiple suppliers to efficiently minimize pick load and reduce electricity cost. In this paper, we evaluate two meta-heuristic optimization techniques: Strawberry Algorithm (SBA) and Enhanced Differential Evolution (EDE) to achieve our objective function. Other sections of the paper are comprising as: Sect. 2 consist of related work. Problem statement is presented in Sect. 3. Section 4 comprise of Proposed system model. Section 5 illustrate simulation and discussion. Finally conclusion and future work is presented in Sect. 6.

Fig. 1. Smart grid

2 Related Work

In past few years, many researchers around the world work in DSM and SG. Many techniques have been modeled for residential as well as commercial and industrial users to meet their demand. In these techniques, researchers work to find out optimal solution in order to facilitate both, consumer and supplier. In this aspect, some of the papers are mentioned as follows which are categories on the bases of techniques used to archive optimal solution. Home Energy Management System (HEMS) is also known as smart grid system which allow bi-directional communication between the consumer and utility company. The main objective of HEMS is to minimize cost, reduce PAR, mange load and maximize UC.

1. Programming Approach
In [3], authors propose a consumption scheduling mechanism for residential users; load management in smart grid. Authors use ILP technique to efficiently minimize the hourly peak load. Daily load is schedule by using ILP to achieve PAR minimization, on the other hand UC is compromised. By using a fractional programing approach [7], for a novel concept of cost efficiency, which is based on residential load scheduling framework. The main objective is to improve the economical efficiency of the residential electricity consumption. For this day-ahead bidding process and RTP mechanism is used to minimize electricity bill. In [6], approximate dynamic programming is used to tackle the problems of power

trading in systems and load scheduling. High penetration of renewable energy resources (RERs) is used to achieve the energy expenses and efficiently utilization of RERs. In this model extra energy is sell out to neighbors at a low price. In all above programing techniques, the authors compromise on UC, on the other hand reduction in electricity bill by scheduling the appliances from on pick hours to off pick hours is achieved. As there is always a tradeoff between cost and UC.

2. Hybrid Algorithm Approach
Hybrid technique is known as combination of two or more than two techniques. Hybrid approach is used to achieve desired objective of electricity cost minimization, PAR minimization and UC maximization by shifting load to off pick hours. A Realistic Scheduling Mechanism (RSM) is propose to increase UC level and customer payoff by enhancing appliance utility, a nature-inspired heuristic algorithm. Binary Particle Swarm Optimization (BPSO) is used for scheduling of appliances. These schedules tend to achieve the required objective among appliance utility and cost effectiveness. By classifying appliances effectively within their respective time slot better results are achieved i.e. electricity bill minimization. A heuristic algorithms-based EMC [4], is designed for a residential area in a SG. The authors evaluates five heuristic algorithms, These algorithms are used for load scheduling of appliances in residential area between on-peak hours and off-peak hours. The proposed system is evaluated by using RTP, for maximizing UC and minimizing both electricity cost and PAR. Two type of load scheduling is performed; one for single home and second for multiple homes. Simulation results proved that the proposed GWD algorithm performs far better than the other techniques in the given scenario, usually GA performed better for cost reduction but the proposed algorithm performed much better to reduce the electricity cost by approximately 10%. RES installation is ignored in this paper. A mathematical optimization model [12], of household energy units is proposed. While preserving the user preferences, the major residential energy load is optimally controlled. The objectives are user comfort maximization along with minimum electricity bill, WDO algorithm is defined and implemented. Knapsack along WDO is used to reduced PAR. RESs installation is neglected in this paper.

3. Heuristic Algorithm Approach
In [6], the authors evaluates the performance of HEMC, which is designed on the basis of heuristic algorithms GA, BPSO and ACO. Problem formulation is tackle via multiple knapsack problem (MKP), also RESs is integrated in this system. For energy pricing, combined model is used i.e. TOU and IBR. The authors achieves electricity bill minimization, reduction in PAR and user comfort maximization. The authors present EMC Model for residential energy management system to avoid peak formation within acceptable limits in user comfort level. The simulation results show that GA-EMC > BPSO-EMC > ACO-EMC. The authors Ignored RESs installation cost power consumption is also neglect and the proposed system is not for industrial and commercial area.

4. Other Techniques and Algorithm
The authors in [8], develop a Generic Demand Side Management (G-DSM) model for residential as well as commercial users. The proposed model reduces PAR and total electricity cost. Minimize waiting time of appliances, along with fast execution of the proposed system architecture, on the bases of GA for both single and multiple users. Also the authors [3], use GA and introduce a general architecture of Energy Management System (EMS) in a Home Area Network (HAN) based on the SG. In this paper RTP is used with combination of IBR to obtain cost and PAR reduction. The proposed system on the other hand if, used only RTP it may damage the entire electricity system. Due to the high PAR, this system also neglects user comfort. Using Nash equilibrium [9], the authors considers more than 1 supplier. As there is more possibility in future of having more suppliers then a single one, also there are more than 1000 consumer. In proposed system both sides can participate; the consumer side and the supplier side. The supplier side can send their electricity rates to the DSM center, where the electricity price is computed against the rate of each supplier. After that, the pricing list is sent to the consumer. On the bases of this list, consumer aim to determine optimal load schedule for cost minimization. The results show that the proposed scheme helps in reducing the peak load and increasing the suppliers' profit and the customers' payoff. In [10], authors address the problem of power scheduling for residential consumers in SG. The appliances are categorized in two categories: flexible starting time and flexible power. The proposed strategy of power scheduling is for the residential user. The main objective is to achieve a desired trade-off between the cost and the UC, without tackling peak formation problem.

3 Problem Statement

SG is used to develop an efficient way of distribution, maintenance and planning operation of electricity which is not possible in traditional grid. SG also provides two-way communication between the consumer and supplier. The consumer communicates their demands with utility by using smart meter, which is used in smart homes to manage the energy consumption. On the other hand, utility according to their resources fulfil these demands. In recent years, several optimization techniques are proposed to effectively tackle the energy management problems in SG. Energy optimization problems of SG are:

1. Minimization of electricity cost.
2. Minimization of the accumulated energy consumption.
3. Reduction in PAR.
4. Minimization in waiting time to achieve desirable UC.

In past years, several optimization techniques are proposed to effectively tackle the energy management problems in SG. For this purpose, DSM is used. The main objective of the DSM is to persuade the user to use less electricity during the on peak hours. DSM control the activities of the end user and encourage the user to manage the electricity usage during on peak hours and off peak hours. A residential home has many appliances, such as Cloth Dryer, Vacuum Cleaner, Refrigerator, Air Conditioner, Dish Washer, Water Heater, Electric Vehicle, Television, Electric Stove etc. DSM use scheduling mechanism to schedule these appliances which in return minimizes the electricity cost. From shifting the peak appliances from on peak hours to off peak hours archive the desired objective of cost minimization. On the other hand, waiting time is increased which affect the UC. There is always a trade-off between UC and electricity cost. Appliances are categorized on the basis of their energy consumption sequence into two classes; class A that contain the interruptible appliances whereas class B contains the fixed appliance. Also each appliance has a fixed time slot to operate [6]. Length of Operational Time (LOT) and power rating are shown in a Table 1.

Table 1. Different appliances operation specification

Appliances	Class	Power (kWh)	Starting time (hours)	Finishing time (hours)	Operational time (hours)
Cloth Dryer	A	1.5	06	14	04
Vacuum Cleaner	A	1	06	15	06
Refrigerator	A	0.125	06	15	06
Air Conditioner	A	1	12	24	10
Dish Washer	A	1	08	22	10
Water Heater	A	1.5	06	23	07
Pool Pump	A	2	12	21	08
Electric Vehicle	A	2.5	16	24	05
Television	A	0.25	01	16	05
Ironing Appliance	A	1	06	16	08
Hair Dryer	A	1	06	13	06
Other	A	1.5	06	24	04
Lighting	B	0.5	16	24	06
Electric Stove	B	1.5	06	14	05
Personal Computer	B	0.25	08	24	04
Heater	B	1.5	03	15	03

4 Proposed System Model

The proposed DSM techniques in SG deals with cost minimization with several users and a single utility in residential area for single and multiple homes by enabling more efficient and reliable operations. SG architecture consists of homes equipped with smart meters having Energy Management Controller (EMC) as shown in Fig. 2, to make reliable and stable bi-directional communication between customers and utilities which is connected through a communication network like Local Area Network (LAN) or through Wide Area Network (WAN).

There are many other mode of communication such as wireless which include Wi-Fi, Z-wave. All elements such as sensors electrical appliances smart devices and storage devices give information to EMC which controls the scheduling of appliances and after that send the appliances schedule to the SG. The operation time of the appliances is divided into 24 time slots each representing one hour, so the total period T = 24 correspond to one day. We take appliances classification [3] based on the demand requirement of the user, each appliance can be categorized as either controllable or must-run. This categorization is done by the user and can be different from time to time. So, the EMC has no control over the operation of appliances as must run. On the other hand, EMC has control on the operation of the controllable appliances which is interrupted if needed. Now these can be categorized as interruptible or non-interruptible appliances, and we assume that mode of operation is not predetermine; as controllable or must run appliances, it depends on the need of the user. According to the preference of the user, those appliances may work as controllable and must run appliances. We consider a system with 30 users and 16 appliances in which each user possesses various controllable and must run appliances, we also assume that for each appliance there is a fixed arrival time, that is each appliance have some given time slot in which it must complete its operation as shown in Table 1. All appliances must not start before the earliest starting time and finish respective working hours prior to finishing time horizon. Start and end times of appliances are assumed to be provided by users, whereas, other parametric values are based on [2], and listed in Table 2. The appliances have different power rating. RTP price signal is used to calculate total hourly cost against the energy consumed in one day, by scheduling the appliances our proposed model shift load from on peak hours to off peak hours which in result minimize cost with least effect on UC.

For scheduling we use heuristic algorithm; Strawberry Algorithm (SBA) and meta-heuristic algorithm; Enhanced Differential Evolution (EDE).

EDE [13] is an updated version of DE whose main operations are mutation, crossover and selection. Three vectors are randomly generated for each target vector. Mutant vector is generated from randomly chosen vectors; difference of any two vectors is added into target vector. With successful creation of mutant vector, cross over phase is start in which trial space is generated by combining mutant and target vector.

Following steps are performed during EDE.

Step 1: Initialize the population.
Step 2: Mutation.
Step 3: Cross over.
Step 4: Calculate trails vectors with their fitness function.
Step 5: Selection: Pick trial vector with minimum cost.
Step 6: Find the greatest individual in the population.
Step 7: End.

SBA [5], is inspired by the strawberry plant for solving continuous multi-variable problems. Strawberry plant can be propagated through the so-called runner (or stolon). The runner is a creeping stalk which is produced in the leaf axil and grows out from the mother (parent) plant. At the second node of runner a new plant, called daughter plant, is formed and then again, a new root and runner arises on the daughter plant to generate another new daughter. Below equation is used to generate random population [5].

$$XProp(j) = [Xroot(j)Xrunner(j)] = [X(j)X(j)] + [drootr1drunnerr2] \quad (1)$$

Algorithm 1. SBA for SG scheduling

```
Initialize (Xnew,kmax,n,M, r1, r2)
Generate a random population in the domain of problem by using Eq. 1
while k = 1 → kmax do
  Do test
  for x = 1 → n do
    Number of population
    for x = 1 → N do
      Number of Appliances
      if r2 > uh then
        if a solution (runner or root) lies outside the legal region, we put it in the border
      else
        r2 = uh
      end if
    end for
  end for
  do fitness evaluation
end while
```

5 Simulations and Discussion

In this section, we present simulation results and comparison of performance parameters: cost minimization, PAR reduction and UC maximization for SBA and EDE. Model is designed for residential area in which each home is equipped with appliances. Appliances are categorized into two groups on the basis of their energy consumption sequence. Also each appliance has a fixed time slot to operate [6]. For electricity cost calculation RTP scheme is used [13]. To evaluate the performance of unscheduled, SBA and EDE simulations are conducted in MATLAB. The results are produced ten times and finally average results are evaluated through confidence interval of 95%.

The energy consumption pattern of all appliances using two different algorithms are shown in Fig. 3. By using SBA and EDE more load has been shifted to off peak hours to reduce electricity cost. SBA schedules most of the appliances in first 6 h and last 4 h while EDE schedule most appliances in last 8 h. The remaining appliances are scheduled within the limits of provided time window. It is also shown that maximum energy consumption in unscheduled scenario is 13.550 kWh which is reduced to 8.580 kWh and 7.800 kWh in case of SBA and EDE respectively.

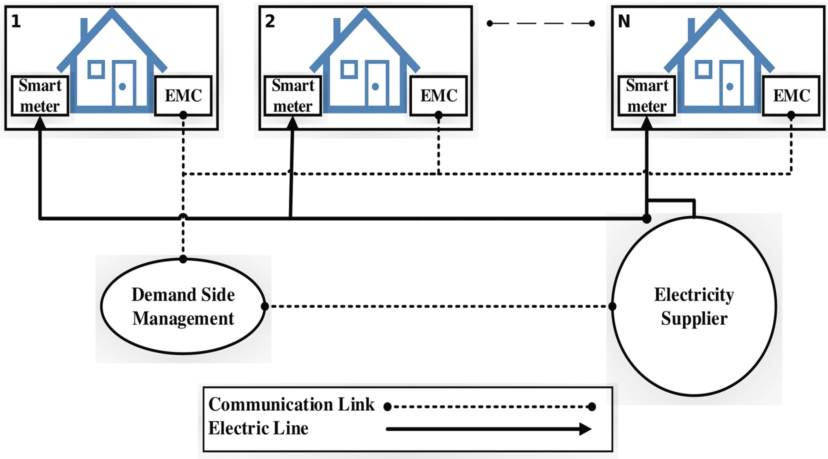


Fig. 2. EMC model for residential users.

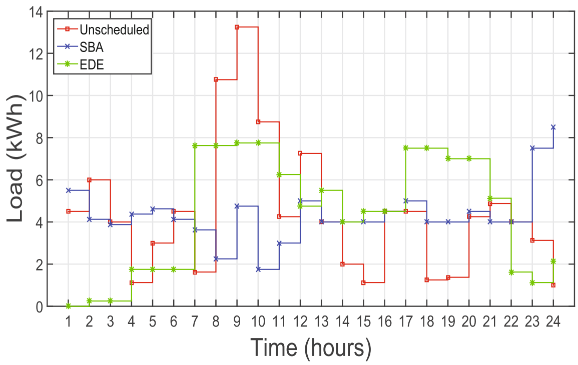


Fig. 3. Load.

Figure 4 shows comparison of hourly power consumption cost for each algorithms. Results shows that SBA has a flat curve during the whole time interval because it utilizes low price. SBA scheduled most of appliances during off peak hours. Maximum cost of electricity bill in SBA is 100 cents. It is increased to 210 cents in case of EDE, while it is increased up to 370 cents in case of unscheduled. EDE does not schedule appliances during time slots 1–4 which in result, creates peak later in time slots with high price. Due to which, total electricity cost of EDE has been increased up to 30% as compared to total cost of SBA, as seen in Fig. 5.

The performance of two different meta-heuristic algorithms: SBA and EDE evaluated in terms of PAR is shown in Fig. 6. From the figure it is clear that the PAR of EDE is 30% less as compared to SBA. However, the total PAR using both techniques: SBA and EDE is less as compared to the PAR of unscheduled load.

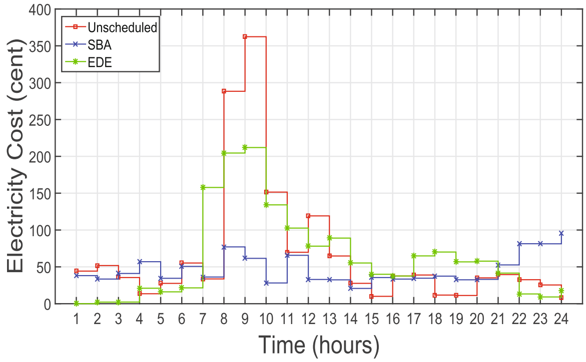


Fig. 4. Electricity cost.

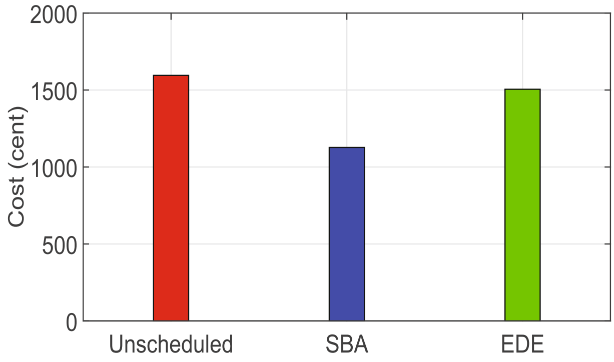


Fig. 5. Total electricity cost.

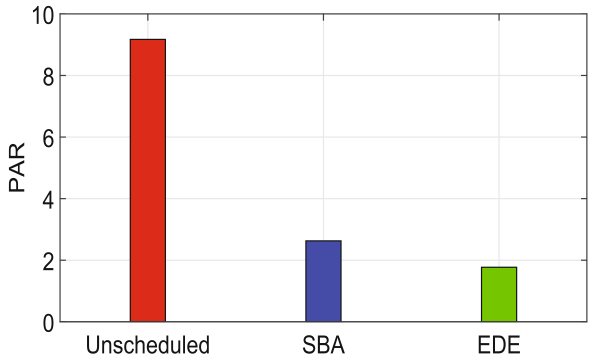


Fig. 6. PAR.

Figure 7 shows the overall waiting time of all appliances using two different meta-heuristic techniques. UC is calculated as the time a consumer waits for an appliance to turn on. Results shows that waiting time of SBA is 25% more as compared to EDE. There always exists trade-off between electricity cost and UC. As shown in Fig. 5, total cost of SBA is almost 30% less as compared to EDE. Both figures show the inverse relationship between UC and cost. UC is not calculated in case of unscheduled, as appliances run whenever they are required, regardless of cost.

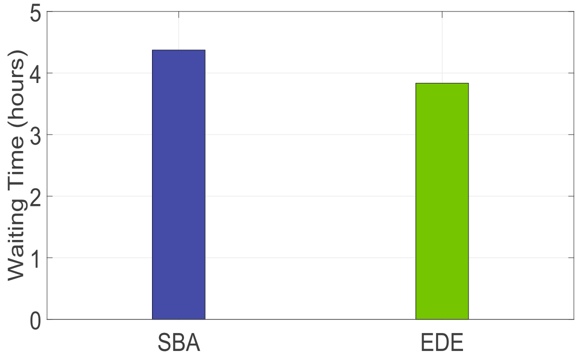


Fig. 7. User comfort.

6 Conclusion and Future Work

In this paper, we propose an EMC to optimally schedule home appliances in order to lower electricity cost and reduce PAR. UC Is also calculated in terms of waiting time. Different types of appliances are schedule on the basis of energy consumption in a given time interval. We evaluate two meta-heuristic optimization techniques; SBA and EDE to achieve the desired objectives, while using RTP scheme. The simulation results show that the proposed techniques; SBA and EDE perform better in term of cost minimization and PAR reduction as compared to unscheduled. Furthermore, EDE performs better than SBA in term of UC. However, electricity cost of EDE is more than SBA, as there exist a trade-off between electricity cost and UC. Energy consumption during on-peak hours also reduced as compared to unscheduled case. In future, we will focus on using of other meta-heuristic algorithms and integration of RES in SG to achieve reduction in electricity cost.

References

1. Rahim, S., et al.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
2. Zhao, Z., et al.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)

3. Zhu, Z., et al.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT). IEEE (2012)
4. Javaid, N., et al.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Merrikh-Bayat, F.: A numerical optimization algorithm inspired by the strawberry plant. arXiv preprint arXiv:1407.7399 (2014)
6. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
7. Ma, J., et al.: Residential load scheduling in smart grid: a cost efficiency perspective. IEEE Trans. Smart Grid **7**(2), 771–784 (2016)
8. Khan, M.A., et al.: A generic demand side management model for smart grid. Int. J. Energy Res. **39**(7), 954–964 (2015)
9. Jalali, M.M., Kazemi, A.: Demand side management in a smart grid with multiple electricity suppliers. Energy **81**, 766–776 (2015)
10. Ma, K., et al.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
11. Mahmood, D., et al.: Realistic scheduling mechanism for smart homes. Energies **9**(3), 202 (2016)
12. Rasheed, M.B., et al.: An efficient power scheduling scheme for residential load management in smart homes. Appl. Sci. **5**(4), 1134–1163 (2015)
13. Zafar, A., et al.: A meta-heuristic home energy management system. In: 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE (2017)

Demand Side Management Using Chicken Swarm Optimization

Zafar Faiz[1], Tamour Bilal[1], Muhammad Awais[1], Pamir[1], Sundas Gull[2], and Nadeem Javaid[1](✉)

[1] COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
[2] Bahauddin Zakariya University, Multan 60000, Pakistan
http://www.njavaid.com

Abstract. In this paper, two meta-heuristic techniques Chicken Swarm Optimization (CSO) and Enhanced Differential Evolution (EDE) are used for demand side management. We have integrated Traditional Grids with Demand Side Management (DSM) We have categorized appliances in two categories; fixed and shiftable/elastic appliances. Real Time Pricing (RTP) is used for calculation of electricity cost. The objective of our work is to minimize electricity bill, increase user comfort, and reduced peak to average ratio. As the simulation results show that CSO gives better results as we compared with EDE in terms of electricity cost and waiting time.

1 Introduction

Today, electricity is the need of every person in technology world. Population is increasing day by day, so the house buildings, industries and commercial areas are also increasing with the passage of time. So traditional grids are unable to satisfy the users need in terms of reliability, scalability and economical. Demands of electricity is increasing with the passage of time, whereas the generation of electricity is not increasing rapidly so it is becoming difficult to manage the electricity load. Every person is not educated that how to save energy and also used it without wastage. In [1], all around the world estimated shortage of energy between country to country is increased up to (30–40)% overall energy consumption. According to (IEO2013), report the energy shortfall is increased up to 56% of the world.

There are several power generation systems used to generate energy and provide energy to grid systems. Some power generation systems are thermal power system, wind driven and solar energy system. Smart Grid (SG) integrates communication, technology and information within traditional grids to make Demand side responsive. In SG system, there involves smart appliances connected with smart meters. Smart meters provide two way communication between the energy user and with the energy providers. The two way communication information is helpful to optimization energy usage and used this information. Demand Side Management (DSM) is used for managing user demands and supply.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_14

Scheduling of appliances according to user demands by smart meter is known as Demand Response (DR). In [2], DSM and DR are the two main functions, which are used in energy optimization. By using DSM electricity load is balanced. It shifts the load form on peak hours to off peak hours for balancing the load.

There are many meta-heuristic techniques used for balancing the load and minimization of energy consumption. In [3], two heuristic techniques Genetic Algorithm (GA) and Binary Particle Swarm Optimization (BPSO) are used for DSM. The main focus of author is on electricity cost, user comfort, and reduction of Peak To Average Ratio (PAR). User comfort is achieved by minimization of electricity cost with minimum waiting time of appliances. In [4], optimal solution for Home Energy Management System (HEMS) to scheduled home appliances by using Binary Backtracking Search Algorithm (BBSA) and BPSO. Two condition are considered when applied BBSA, in condition one home appliances are operationalized in weekdays from 4 to 11 PM and in second condition appliances are operationalized at any time in a day. The BBSA scheduler controller can reduce energy consumption, minimize electricity bill and helps to shift load from on peak hours to off peak hours to reduce cost. BBSA provides better results when it is compared with BPSO.

All the above discussed techniques have the same objective to minimize energy consumption, decreases PAR by reducing waiting time to achieve user comfort and minimize the cost of electricity. In our propose scheme, we have used Enhanced Differential Evolutionary (EDE) and Chicken Swarm Optimization (CSO) for scheduling of home appliances. Residential area is taken into consideration in this paper. Home appliances are categorized into two classes, class A and class B; class A have fixed operational time appliances and class B have shiftable or elastic load appliances. CSO provides better results when we compare it with EDE. In CSO, we achieved user comfort with minimum waiting time of appliances, minimized electricity cost and PAR. Rest of the paper has three sections. Related work is discussed in Sect. 2, proposed system in Sect. 3, whereas results and simulations of CSO and EDE discussed in Sect. 4. In the last section we conclude our proposed technique work.

2 Related Work

In this paper [4], home appliances are considered to schedule in smart grid. Electricity load is scheduled according to appliances power, and working time that called as DSM. Integer linear programing technique is used to schedule the appliances. There are two types of appliances used in home, fixed power consumption appliances, and shiftable power consumption. User comfort is decrease, and cost increased. The proposed method is capable to shift the load from on peak hours to off peak hours. The proposed technique is increasing the user comfort, reliability, performance of the smart grid and decrease the cost. So that the performances of the individual appliance also increased. There is trade of between the PAR and user comfort. Some limitations exist in the smart grid system that is least emphasis on the user comfort. In [5], a cost effective residential load scheduling framework has been introduced by authors to achieve better residential electricity consumption, while keeping the economic efficiency under consideration. The authors have analyzed the effect of load shifting on cost with different households to attain optimized scheduling. Fractional Programming is used for developing a load scheduling algorithm to attain cost efficiency and providing with various consumption patterns to analyze, with the combination of IBR and DAP being integrated to work accordingly. Results depict that user comfort has been increased, however the authors have not considered the effect of PAR and its significance.

In paper [6], the authors is focused on the smart grid system in residential areas for the minimize of power utilization by scheduling the home appliances. The main problem which exists in the grid system is higher electricity cost, and also create PAR. In this paper authors designed an Architecture Energy Management System (EMS) for Home Area Network (HAN). All the home appliances are connected through a Network. Each appliance have a transceiver and data processor. The data processor is work on the received information through transceiver and scheduled the appliance according to time slot. In this procedure Home Gateway (HG) and Demand Response (DR) involved. HG is received the DR information. The electricity cost is calculated by using hybrid technique of RTP Scheme and Inclined Block Rate (IBR). Electricity price is minimized by using Hybrid pricing scheme. In paper [7], authors described the DSM techniques for residential, commercial and industrial areas to manage the load in smart electric grid system. The problems are addressed in this paper: The appliances are in class A non-shiftable appliances time while the shiftable appliances are in class B. Hybrid Genetic Wind Driven (HGWD) technique is used for scheduling the appliances. HGWD consists of GA, BPSO, WDO and Bacterial Foraging Optimization Algorithm (BFOA). The proposed scheme helps to decrease the waiting time of appliances. Demand load is shifted on peak hours to off peak hours without stopping the non-shiftable devices during the running. The objectives of the existing problem is achieved by using the hybrid GWD technique. Cost of the electricity bill is calculated by using the RTP scheme. Some limitations of this paper is ignorance of RES. In paper [8], authors described the HEMS in electric smart grid to manage the load and minimization of energy consumption. There are number of appliances used in smart homes. The problems of the existing system is that some appliances are running in peak hours so that the energy consumption is high during that time. Due to high energy consumption, peak is generated. In peak hours electricity cost is also high which results in increased overall electricity bill. The proposed system is scheduling appliances in two approaches: one is fully distributed approach and second is hybrid approach. In fully distributed approach appliances are intelligently stop running according to time span. However in hybrid approach user forcefully stop the appliance in peak hour. Author used two types of approaches for mathematical modeling, A Game theoretical approach and Nash Equilibrium. Dynamic pricing scheme is used for calculating the cost of electricity. These proposed solution are Decreases waiting time of appliances so that PAR is decrease. The cost of electricity is also decreases. So the limitation of smart grid system is to ignorance of user comfort.

In [9], DR is used to receive the demands of users and smart grid managed the load. Each user demanded from the utility that how much energy is required. Problems existing in the previous system is scheduling of appliances in specific time slots. In this paper a game theory model approached is used to manage the load. According to game theory model, utilities is announced the different pricing schemes for different user demands. Each user is purchased the electricity from the utility as affordable. Objectives of this paper, PAR is reduced and load of the grid minimized. The limitation of the system is to ignore of user comfort.

In paper [10], authors addressed the scheduling problem of appliances in residential areas. DSM is required to schedule two types of appliance, flexible appliances and the appliances running with no interruption during the running. PAR is created between the flexible, and non-interruptible appliances. The utility is announced the high price and low price of electricity according to the power utilization. There are three operational modes are used to schedule the appliances. In first operation mode user comfort not considered and electricity cost minimized. In the second mode of operation electricity cost is not considered, and user comfort maximized. Moreover in third operation mode both problems are addressed. The results of the proposed technique is reduction of PAR, and the user comfort increased. Some limitation and controls are also exists in the smart meters, Electricity cost not significantly reduced.

In [11], the authors has taken three Algorithm GA, BPSO, ant colony optimization (ACO) and Multiple Knapsack Problem (MKP). In this paper, for calculating energy price by combined two schemes time of use and inclined block rates. There are following problem address in this paper like electricity cost, user comfort and PAR. There are also waiting time decreased. The performance of all algorithms (GA-BPSO-ACO and MKP) with respect to PAR when it is compared with unscheduled comparatively decreases. The user will consume the energy from the utility as well as from the Renewable Energy Source (RES), this will increase user comfort. This integration help to minimize the energy production of the utility so that cost of the electricity also decrease while the cost of electricity storage and maintenance is ignored.

In [12], there are two types of problem addressed in this paper, scheduling of power in home according to number of appliances and targeting the generating power. Approximation dynamic programming technique is proposed for scheduling number of appliances in home. This technique is helped in utilization of the Renewable Energy Resource (RER). The RER is very useful for the generation of power. A Game Theoretic Approach is able to sell out the RER. This game theory is helped in intercommunicate between the users and the utility. Furthermore the proposed technique also make it easy for the DSM storage of energy. The stored RER energy is help in approximation dynamic programming technique for scheduling appliances. That energy is used in those appliances having not to stop during the running. Quadratic cost formula is used for calculating the prices of electricity. User comfort is increase as there exist a competition between the different sellers to sell the RER. The cost of electricity is lower as compared with the utilities due to extra energy sell to the local users.

There are some limitation are exist in the system, ignorance of PAR, the extra charges required to developed a RER system and its maintenance charges. In [13], energy consumption is increased in the home management system by generated through RES. Devices are consumed more energy in the peak hours and less consumed in off peak hours. Cost of electricity in peak hours is high as compared with off peak hours. Author is considered domestic areas for scheduling of appliances and generated energy in smart homes by using RES. The problems which are addressed in this paper, PAR and cost of the electricity coming from grid system. Artificial Neural Network algorithm (ANN) is proposed for scheduled appliances and GA is used for optimization of energy. ANN is trained by given a data set. A weekly schedule of appliance is provided to the consumers. RES reduced the energy consumption of grid. In paper [14], appliances are scheduled on the basis of demanded load by the user from utility. DR is used for this purpose. Real time scenario, residential area is considered for testing of Algorithm. The most usable appliances are taken in a set. These appliances are divided into two cases, in first case appliances which are only running 4 to 11 pm in each day and in second case appliances those are running in any time weekend days. BBSA is used to schedule the appliances and BPSO is used to calculating the accuracy of scheduled. The major issues are addressed in this paper shifted the load form peak hours to off peak hours, electricity cost. The proposed BBSA technique is achieved objectives: i.e., reduction of PAR, and minimization of electricity cost. Limitation of the gird system is to ignore the user comfort. In paper [15], the energy demand of the consumers are increased in permanent manners. The DSM load is divided into two layered approaches. These layers are scheduling and real time layers. The PVT batteries are used for the energy storage purpose. The DSM requests are handled by using Non-dominated Sorting Genetic Algorithm (NSGA-II). Fuzzy Logic Decision Maker (FDM) is used to scheduled appliances. FDM is provide flexibility between the appliances scheduling and increase the performances of smart grid system. The problems addressed in this paper is to increase the energy demands and user comfort. By using proposed technique objectives are achieved. Some limitation of the system are ignorance of PAR, the Cost of the PVT panels, and batteries.

3 Problem Statement

In this research paper, Two main issues are target, one is the minimization of electricity cost by optimize of electric energy from the end users and the Second one is the reduction of PAR. The existing system is unable to manage the user requirements. In this paper, we use home energy management controller for optimization of energy in an efficient way.

Many other techniques have been proposed for tackle these kinds of problems as mentioned before. Authors in [11], ACO, BPSO and GA is used in DSM for scheduling of home appliances. RES is integrated with HEMS to achieve the user comfort. In [7], GA, BPSO, BFOA and GWD is used for optimization of electric energy in SG.

DSM controller used for collect usage pattern of energy by the user and scheduled the load according to that pattern. Energy is scheduled such that maximize the user comfort and minimize electricity bill. In our proposed scheme, following problems are addressed as: Optimization of energy consumption pattern, minimize PAR, reduce electricity bill, minimize the aggregated energy consumption, minimize waiting time and balanced the overall electricity load.

4 Proposed System

In proposed system, we managed the electricity load of a single home appliances in residential area. The design of system consist of a single home connected with Smart Meter (SM), controller and with the electricity supply company. Controller provides two way communication between the user and with the utility. In a home there are three types of appliances used (i) fixed, (ii) elastic and (iii) shiftable. Fixed load appliances are kept in class A, Whereas the elastic and shiftable appliances are kept in class B.

RTP pricing scheme is used in proposed system. It generates the pattern of prices on hourly basis. Appliances are scheduled by handler and it generate specific interval for specific hours.

In the proposed system, class A and class B appliances are scheduled by using EDE algorithm and CSO algorithm. In EDE, five different trial vectors are used to find the best solution for scheduled appliances. In CSO, create whole population randomly that is known as roosters. The roosters are divided into three sub categories: (i) population of hens, (ii) population of chickens, and (iii) population of mother of hens. In each iteration these three kinds of population are different and calculate the fitness value of each individuals in each iteration. Population is selected on the basis of fitness values, which has best fitness value. Update all four kinds of population and the fitness values. Provides the best solution of scheduling. In [16], algorithm of CSO is written below.

We used two types of classes which are known as class A and class B. In, class A consist of fixed load appliances and class B consist of shiftable and elastic appliances. In this paper user comfort is maximize, utilities of appliances maximize and cost is minimize. Scheduling is doing on the basis of priority bit. In peak hours the appliances with higher priority do not shift and the appliances which have low priority bit or shiftable are transfer to off peak hours. The details of class A and class B appliances given below in the Table 1.

In proposed CSO and EDE we used RTP pricing signal. That is pattern of price for 24 h.

In the proposed scheme, we used EDE and CSO algorithms for optimization of energy and scheduling of appliances.

5 Simulation and Results

In this research, we have classified home appliances into two classes. These appliances are scheduled by using EDE and CSO algorithm. Electricity price is calculated by RTP scheme. Results of our simulation are described below (Figs. 1 and 2):

Algorithm 1. CSO

Initialize all parameters (D,x1,xu)
total population size *popsize*, the roosters accounts *Nr*, the hens accounts *Nh*, the mother hens accounts *Nm*, updating frequency of the chicken swarm G and the maximum number of generations *itermax*.
Generate a population $X = x_1, x_2, ..x_i, ..x_{popsize}$ of *popsize* chickens with random solutions
Calculate the fitness(xi) and find the best solution X_{best} of the population.
for $iter = 1 \rightarrow itermax$ **do**
  if $Wi = iter\%G$ or $iter$==1 **then**
    Sort all population individuals according to their fitness.
    Divide total population individuals into three subpopulations (called rooster population, hen population, and chickens population) according to their sort criteria, and establish the relationship between the chickens and its mother (hens).
  end if
  Update the rooster population individuals according to Equation $x'_{i,j} = x_{i,j} + \psi_{i,j}.t(n).x_{i,j}$ and calculate their fitness.
  Update the hen population individuals according to Equation $x^{t+1}_{i,j} = x^t_{i,j} + S1.rand.(x^t_{r1,j} - x^t_{i,j}) + S2.rand.(x^t_{r2,j} - x^t_{i,j}) + S3.rand.(op^t_{best,j} - x^t_{i,j})$ and calculate their fitness.
  Update the chicken population individuals according to Equation $x^{t+1}_{i,j} = x^t_{i,j} + FL.(x^t_{m,j} - x^t_{i,j})$, FL ϵ ,[0,2]
  Update the personal best position X^*_i and the global optimal position X_{best}
  Perform local search for the global optimal individual.
end for
Output the best solution X_{best}

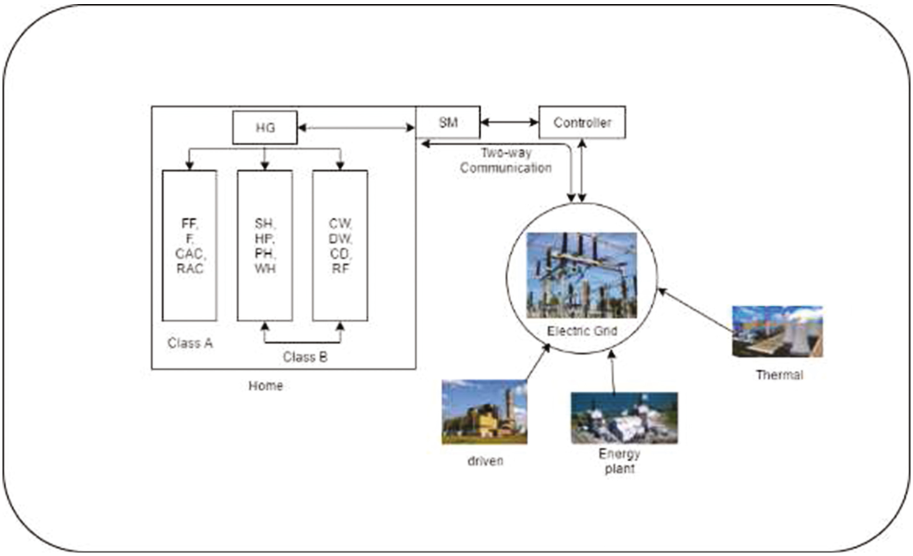


Fig. 1. Home Energy Management System

Table 1. Appliances classification with LOT

Class name	Appliances name	Power rating (KWH)	LOT (Hours)	Deferrable load
Class B	Space heater	1	9	1
Class B	Heat pump	0.11	4	1
Class B	Portable heater	1.00	5	1
Class B	Water heater	4.50	8	1
Class B	Cloth washer	0.51	9	1
Class B	Cloth dryer	5.00	5	1
Class B	Dish washer	1.20	11	1
Class B	First-refrigerator	0.50	24	1
Class A	Fan	0.5	11	0
Class A	Furnace fan	0.38	8	0
Class A	Central AC	2.80	12	0
Class A	Room AC	0.90	5	0

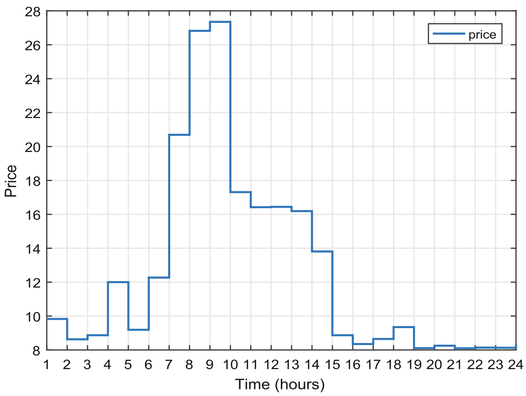


Fig. 2. RTP signal for 24 h

In our proposed solution, RTP pricing signal is used. Electricity bill calculated on the basis of hourly consumption of energy. Utility regulates RTP into two parts. Base bill: It depends on Customer's Base Load (CBL) Hourly prices: Hourly prices are applied according to customer's usage (difference between actual and CBL). In Fig. 3 show the calculated electricity cost.

In Fig. 4, shows the difference of total cost between two proposed techniques EDE and CSO with unscheduled load. Total cost is not more reduce because there trade off between PAR and electricity cost. PAR is more reduce as compared with electricity cost.

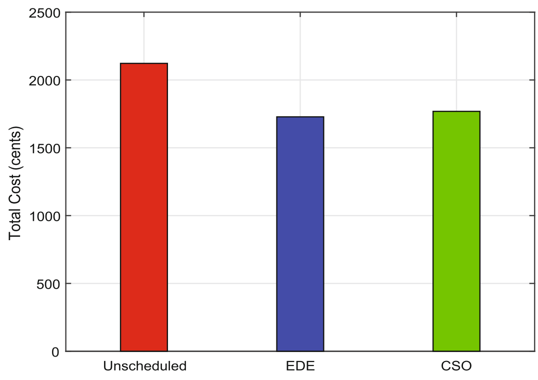


Fig. 3. Electricity cost

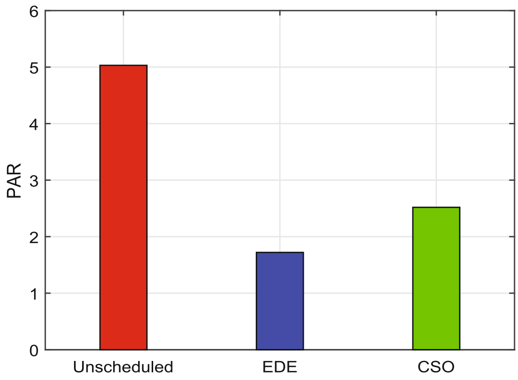


Fig. 4. PAR comparison with EDE and CSO

Here we can see that peak is generated at day time so in case of unscheduled load electricity pricing is very high and when the appliances are scheduled, during day time the load is decreased in on peak hours so overall load is decreased. There is a tradeoff between the cost and PAR. Overall cost of EDE is 40% and in CSO 30% decreases as compare to unscheduled.

Waiting time in CSO is 10% more decreasing as compared with EDE. However in our simulation there are appliances divide into two classes one is fixed load appliances and second is shiftable and elastic. So the waiting time of fixed load appliances is near to zero while the elastic based and interruptible appliances waiting time is scheduled.

User paid more electricity bill while not compromise with satisfaction. PAR value is decrease more but less decrease the cost of electricity. Difference is calculated by comparison of two Algorithm CSO, EDE and Unscheduled shown in Fig. 5.

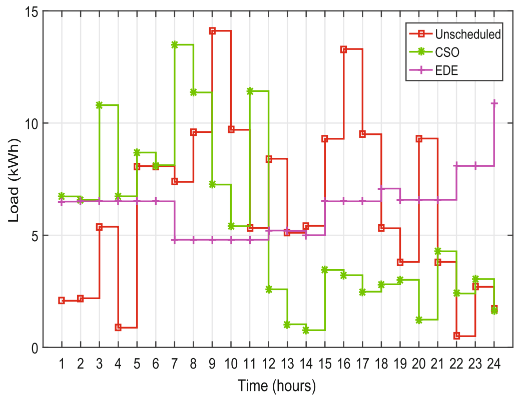


Fig. 5. Load scheduling for 24 h

6 Conclusion and Future Work

In this research work, DSM controller is designed in which two meta heuristic algorithm CSO and EDE proposed for implement HEMS. Our proposed system RTP pricing signal is used for calculating the cost of electricity. CSO reduced the overall cost by approximately 30% and in comparison with EDE electricity cost reduced upto 40%. User comfort is achieved in term of reducing the waiting time of appliances. From our research work simulation results are concluded that there exist a trade-off between PAR and electricity cost. Energy consumption is reduced during the on-peak hours as compared with the unscheduled load. In future we will integrate RES with SG and will propose hybrid of both algorithm EDE and CSO for reduction of electricity cost.

References

1. Khalid, A., Javaid, N., Mateen, A., Khalid, B.: Demand side management using hybrid bacterial foraging and genetic algorithm optimization techniques. In: CISIS (2016)
2. Zafar, A., Shah, S., Khalid, R., Hussain, S.M.: A meta-heuristic home energy management system, January 2017
3. Nojavan, Z., Zare, K., Mohammadi-Ivatloo, B.: Selling price determination by electricity retailer in the Smart Grid under demand side management in the presence of the electrolyser and fuel cell as hydrogen storage system. Int. J. Hydrogen Energy **42**(5), 3294–3308 (2017)
4. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in Smart Grid. IEEE (2011). 978-1-4577-2159-5/12/31.00
5. Ma, J., Chen, H., Song, L., Li, Y.: Residential load scheduling in Smart Grid: a cost efficiency perspective. IEEE Trans. Smart Grid **7**(2), 771–784 (2016)

6. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
7. Javaid, N., Javaid, S., Wadood, A., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid, Version 5 February 2017, submitted to Energies
8. Barbato, A., Capone, A., Chen, L., Martignon, F., Paris, S.: A distributed demand-side management framework for the Smart Grid. Comput. Commun. **57**, 13–24 (2015)
9. Reka, S.S., Ramesh, V.: A demand response modeling for residential consumers in smart grid environment using game theory based energy scheduling algorithm. Ain Shams Eng. J. **7**, 835–845 (2016)
10. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in Smart Grid. Electr. Power Energy Syst. **78**, 320–325 (2016)
11. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khanb, Z.A., Alrajehc, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
12. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), July 2016
13. Yuce, B., Rezgui, Y., Mourshed, M.: ANN-GA smart appliance scheduling for optimised energy management in the domestic sector. Energy Build. **111**, 311–325 (2016)
14. Ahmeda, M.S., Mohameda, A., Khatibc, T., Shareefd, H., Homode, R.Z., Ali, J.A.: Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. Energy Build. **138**, 215–227 (2017)
15. Khourya, J., Mbayedb, R., Salloumb, G., Monmasson, E.: Predictive demand side management of a residential house under intermittent primary energy source conditions. Energy Build. **112**, 110–120 (2016)
16. Qu, C., Zhao, S., Fu, Y., He, W.: Chicken swarm optimization based on elite opposition based learning. Math. Probl. Eng. **2017**, 26 (2017)

Energy Optimization in Smart Grid Using Grey Wolf Optimization Algorithm and Bacterial Foraging Algorithm

C.H. Anwar ul Hassan, Muhammad Sufyan Khan, Asad Ghafar, Syeda Aimal, Sikandar Asif, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. Nowadays, energy is the most valuable resource, new techniques and methods are discovered to fulfill the energy demand. These techniques and methods are very useful for Home Energy Management System (HEMS) in terms of electricity cost reduction, load balancing and power consumption. We evaluated the performance of HEMS using Grey Wolf Optimization (GWO) and Bacterial Foraging Algorithm (BFA) techniques inspired by the nature of grey wolf and bacterium respectively. For this purpose we categorize the home appliances into two classes on the bases of their power consumption pattern. Critical Peak Pricing (CPP) scheme is used to calculate the electricity bill. The load is balanced by scheduling the appliances in Peak Hours (PHs) and Off Peak Hours (OPHs) in order to reduce the cost and Peak to Average Ratio (PAR) and manage the power consumption.

1 Introduction

Technologies made the life of human being easy in different aspects. Electricity is great invention of science for mankind. Increase in the population increases the demand of electricity. Traditional Power Grid (TPG) is inadequate to meet and fulfill the demand of electricity. In this regards Smart Grid (SG) introduced in order to meet the energy challenges. SG increase the customer engagement, power efficiency, storage capacity and, grid sustainability. Wired and wireless technologies, advanced metering and management system are introduced in SG. Although there are number of way through which we generate electricity in SG, some of these are; thermal power plant, nuclear power plant, photovoltaic, wind turbine, hydraulic power generation as shown in Fig. 1.

SG are bi-directional so we manage the electricity demand and use the information flow between the utility and consumer. By this information we optimize the energy at demand Side. Through Demand Side Management (DSM) strategies, we manage the power consumption, as the utility gets the consumer usage pattern of electricity. Managing the power usage reduce the electricity cost by

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_15

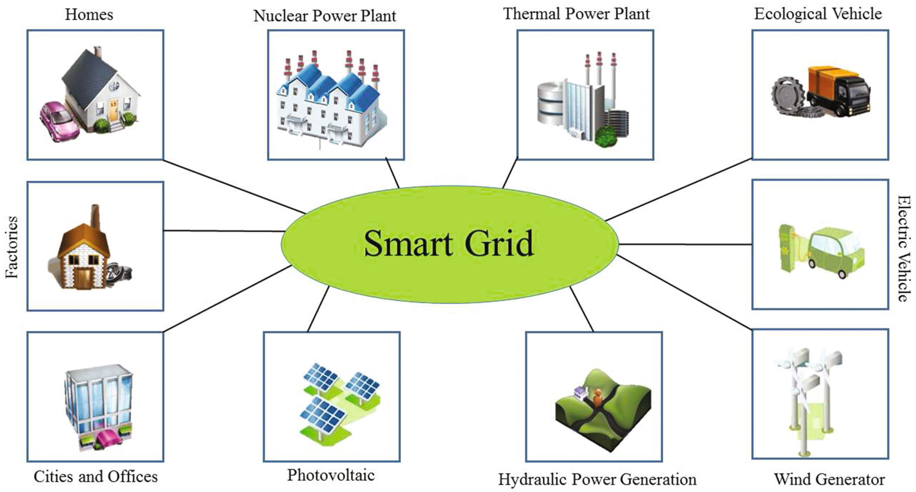


Fig. 1. Smart Grid

shifting the load form PHs to OPHs. Demand Response (DR) and Load Management are two main aspects of DSM. DR reflect according to the dynamic pricing scheme, these are; Inclined Block Rate (IBR), Time of Use (ToU), Real Time Pricing (RTP), Critical Peak Price (CPP). Electricity price is high in PHs by managing the load during these hours reduce the cost.

Managing the electricity usage would minimize the cost. Many optimization techniques are used to overcome the energy management challenges. For example in [3], author use the Genetic Algorithm with RTP and IBR pricing scheme to optimize the energy management. In this paper, we considered the residential area for energy management. We use heuristic optimization techniques these are; BFA and GWO with CPP pricing scheme.

Rest of the paper is organized as: In Sect. 2 we discuss the related work, Sect. 3 problem formulation, Sect. 4 proposed model, Sect. 5 simulation, experiments and results and then finally Sect. 6 conclusion.

2 Related Work

Many researchers around the world work on energy optimization to optimally schedule smart appliances, balancing the load and efficiently manage the energy consumption. In this regard, some of the papers on SG Energy Management Systems (EMS) are discussed in this section. In [1,3], authors discuss the household appliance that operate in PHs. Enhance energy management in smart home through scheduling the appliances. In [1], author proposed Inter Linear Programming (ILP) approach for managing load in smart homes. It is used to manage the power consumption and appliances operational time. The main purpose of the proposed mechanism is to minimize the peak hourly load and balanced daily load usage. Author design the load management system using user interface and smart meter to manage the power consumption by assigning priority to the appliance, maintaining schedule and smart meter are used to get the total consumption of electricity for each appliance. The consumption of energy and the scheduling of appliances will be optimize globally on the basis of collected information via a meter. An efficient optimal scheduling method for demand response in Home Energy Management System (HEMS) is introduced in [2] for scheduling the appliance to reduce the cost and reduce the PAR. Author used GA approach with RTP and IBR pricing scheme to manage the appliance. In [11], authors address energy management by scheduling the appliance, to minimize the peak demand of electricity and reduce the electricity cost. Author proposed two different algorithms; Discrete Non-dominated Multi-objective Particle Swarm Optimization (DNMPSO) algorithm and Manhattan distance based Non-dominated Multi-objective Genetic Algorithm (MNMGA) to schedule the load according to the demand of electricity. Authors efficiently reduce the PAR in [1,3,11]. User comfort is compromised in [1,3] and in [11], author work will be extended to schedule the load by integrating Renewable Energy Resource (RES) as, author discuss it as a improvement.

In [10], author presented the framework to schedule the load, based on cost efficiency. Power trading and load scheduling in smart homes overcome by using RES or Distributed energy resources (DER). Author use Dynamic Programming approach [2], to schedule the appliances. The operational time is divided into 24 h time slots to schedule the appliances and in trading sell the extra generated energy. Author use a game theoretic approach to manage the user power consumption. The proposed algorithm reduces the user expenses by managing electricity usage and facilitate the integration of RES to generate electricity, not only for their use also sell to utility company to obtain large market share and maximized the revenue. In [10], author proposed frame work to schedule the load, residential consumers adopt different load scheduling methods to manage the appliance in order to minimize the cost and energy consumption, sometime these methods increased the power consumption cost. However author not discussed the integration price of RES and DER respectively. Author proposed an efficient EMS in [12], by managing and monitoring the energy usage: energy supplied and energy consumed. The energy consumption in both residential and commercial buildings are discussed. The consumption of energy worldwide including residential and commercial building, data is analyzed. As analyzing the previous energy data, policies are established to manage energy consumption. By managing and monitoring energy consumption reduce the usage of energy, on the other hand this approach neglected the user comfort.

The Author in [6], propose a technique for HEMS maximizing UC and minimizing the electricity bill. Author of [6], design DSM using three heuristic algorithms GA, BPSO and Ant Colony Optimization (ACO) with IBR and ToU Pricing scheme. Objective of DSM is to manage the electricity and demand side control activities in order to avoid the PAR and reduce the electricity bill. In addition, problem formulation is carried via multiple knapsack problem.

Appliance are categories in to three types Fixed, Interruptible and non-interruptible for scheduling the load and Integrate RES to reduce the cost and operate the appliance with RES in PHs. The electricity bill is reduced with the reduction of peak. Simulation performed on all these three heuristic algorithm, by comparatively analysis we found that GA perform better as compared to other techniques. The flaw in [6], author not discussed the implementation price of RES. An efficient energy consumption management approach is presented in [8]. Based on the presented approach author classified the appliance into essential and flexible demand. Flexible demand are further classified into delay sensitive appliance (DSA) and delay torrent appliance (DTA). Centralized algorithm is presented to optimize the energy for both DSA and DST. Due to high priority DSA operate then DTA. When the high load demand of DSA or great number of appliance in queue then waiting time for DTA is raised. In this regard, author upgrade it by probability scheme to overcome the operational delay of DTA. Using probability scheme assigning the high priority to the DTA then they operate before newly arrived DTA. When the information of real time system is not known then neural networks are used to approximately manage the decision. By the proposed scheme, author minimized the delay. The minimization of total cost and flexible demand delays are main focus to achieve optimal energy solutions.

In [5], author schedule the load for residential consumers, to achieve the desirable trade-off between electricity price and UC. Price based demand response program is used to manage the appliance. Day a Head (DaH) pricing scheme is used, utility or service provider announced the electricity price and consumer schedule there appliances according to that. Manage the power usage and time of these appliances respectively, to minimize the cost and energy consumption. Author use Taguchi Loss Function (TLF), to evaluate the cost of discomfort. The Taguchi's loss function describes the relationship between the power deviation and economic loss. Author define three different operation modes with user point of view to achieve the desire objective. Simulation results all of these operation modes perform better by using DaH pricing scheme then flat pricing scheme, it allows the scheduling strategy to achieve a desirable tradeoff between the cost and UC. Authors in [7], proposed a model for DSM using BPSO techniques, which is enhanced version of PSO to reduce the electricity bill and maximize the user comfort. HEMS is used to schedule interruptible appliances load. Appliances are classified in to three type and ToU pricing scheme is used, electricity bill generated according to the usage pattern of appliance. The objective of the proposed approach is the electricity minimization by load shifting and power saving. In [9], author proposed realistic scheduling mechanism for SG to reduce user frustration. For optimal operational time of different appliances and power usage optimization, nature-inspired heuristic algorithm that is BPSO is used with the adaptation of ToU pricing scheme. Shift able appliance are operate in OPHs due to their optimal operational time with the consideration of user's living pattern and working characteristics of the appliance. Shifting the load form PHs to OPHs reduce the cost. By comparing the scheduled and unscheduled load, author analyze the effectiveness of the proposed scheme. The proposed scheme is cost effective. In short, the PAR reduction is actually the reduction of the cost.

3 Problem Statement

In recent years, several optimization techniques are proposed to effectively tackle the energy management problems in SG. In SG, smart meters allow two-way communication between the consumer and the utility. A smart meter is used in smart homes to manage the energy consumption. The Utility gets the energy consumption pattern of user and response according to user demand, this way they effectively plan and distribute the electricity. It is difficult to achieve both user comfort and minimum cost. Mostly UC and PAR is neglected. In SG, energy optimization problems are: minimize the electricity cost, accumulated energy consumption, reduce peak load creation and waiting time to achieve desirable user comfort.

DSM control the activities of the end user, encourage the user to manage the electricity usage. The time in a day when the demand and the usage of electricity is very high that are the PHs of the day, it varies according to the utility companies, in that hours the price of electricity would be high, so manage the usage of electricity in that hours to reduce the cost. This way consumer compromise on comfort but reduce price and manage the power consumption. In order to minimize the cost, load shifted from PHs to OPHs, by this cost is reduced and the usage of energy increase in OPHs [5]. Randomly load shifting increase the demand of electricity in OPHs so, we need to shift the load in certain way, that avoid peak load creation. Author also use high parameter values that increase the execution time [4]. Our proposed algorithm schedules the appliances in such way; to avoid peak load creation and efficiently manage the energy usage.

4 System Model

In order to efficiently tackle the power consumption in residential area; HEMS manage the appliances in smart homes. Every smart home is connected to the smart meter. Utility get the energy usage pattern of consumer as shown in the Fig. 1. It is a bi directional communication, one end of the smart meter is connected to the SG and the other one is connected to the residential area or a smart home.

In residential area DSM, we consider N smart homes and P smart appliances as shown in the Fig. 2. CPP pricing model is used to calculate the per day cost of electricity consumption. Cost varies in PHs and OPHs. In order to optimize the energy management problem in smart homes we design the power consumption model for P number of appliances.

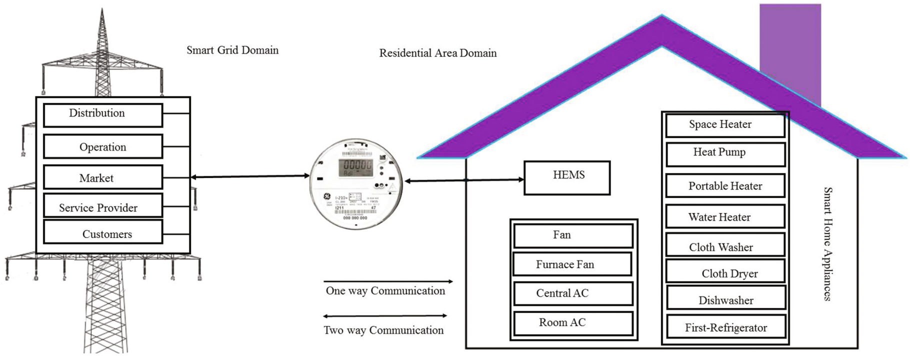


Fig. 2. Smart residential DSM components

4.1 Power Consumption Model

Let $P = \{a_1, a_2, a_3, \ldots, a_p\}$ represent the appliances in N smart homes respectively, 24 h time slot t scheduled such that $T = \{1, 2, 3, 4, 5, \ldots, 48\}$ whereas $t \epsilon T$. By this we calculate the hourly energy usage of each appliance, as follows:

$$E_{a,t} = E_{a,t_1} + E_{a,t_2} + E_{a,t_3}, \ldots, +E_{a,t_{24}} \tag{1}$$

Eq. 1, show the hourly power consumption of each appliance. Total power consumption of all appliance per day is calculated by using Eq. 2 as follow:

$$E^n = \sum_{t=1}^{24} \left(\sum_{p=1}^{D} E^p \right) \tag{2}$$

Our main objective is to reduce the total cost by minimize the energy consumption, as in Eq. 3 total cost is calculated by reducing the peak load creation. By using Eqs. 4 and 5 we calculate the total load and PAR.

$$Cost = \sum_{t=1}^{24} \left(E^{hour}_{Rate} x P^{App}_{Rate} \right) \tag{3}$$

$$Load = P^{App}_{Rate} x App \tag{4}$$

$$PAR = \frac{max(Load_s)}{Avg(Load_s)} \tag{5}$$

4.2 Load Categorization

For energy management is smart homes, we categorized the load into two classes; class A that contain the fixed appliances whereas class B contains the interruptible appliance and elastic appliances, Length Operational Time (LOT) and power rating are shown in a Table 1.

Table 1. Parameters of appliances

Appliance	Class	Power(kWh)	LOT(h)
Space heater	B	1	9
Heat pump	B	0.11	4
Portable heater	B	1.00	5
Water heater	B	4.50	8
Cloth washer	B	0.51	9
Cloth dryer	B	5.00	5
Dish washer	B	1.20	4
Refrigerator	B	0.50	24
Fan	A	0.5	11
Furnace fan	A	0.38	8
Central AC	A	2.80	12
Room AC	A	0.90	5

4.3 Optimization Techniques

Many optimization techniques are used for energy management in SG. In this regards, we are using BFA and GWO to develop an optimal solution. These are the nature inspired algorithms based on the bacterial foraging steps and wolf tracking, encircling and attacking steps respectively. We use these optimal techniques in SG to find the best optimal solution.

BFA's behavior depends upon the foraging strategies, eliminate the bacterium who have poor foraging strategies and favor those who have prosperous foraging strategies. In foraging strategies after many generations poor one's are either redesigned or they are eliminated [4]. The statistic of BFA allow the bacterium cell, swarm or tumble stochastically towards optimal solution. Three main steps are performed in BFA to achieve the desired omen ties: chemotaxis, reproduction and elimination-dispersal. In chemotaxis step we measured the length of life time of bacteria by the number of chemotaxis step. Cost Ji of the bacteria is calculated by comparing it with the new position θi of another bacteria, after a tumble along the manipulated cost surface one at a time by adding step size Çi in the temple direction lie between [−1,1]. Random direction vector $\triangle$i is generated to represent the tumble. In second reproduction step the cells that performed well in their lifetime be the part to contribute for a next generation in third and last step cells are eliminated, discarded and new random samples are inserted. We implemented these step in order to solve the scheduling problem. In BFA initially random population Np is generated for each bacterium, to calculate the fitness function Ji of each bacterium, by performing the steps of BFA we select the best one for scheduling and replace it with the worst one in a population to achieve a desire result.

$$Ji[j,k,l] = Ji[j,k,l] + Jcc(Oi[j,k,l], Np[j,k,l]) \tag{6}$$

GWO's pretend the leadership hierarchy and hunting nature of grey wolf. Social hierarchy of GWO consist of four levels such that Alpha α, Beta β, Delta δ and Omega ω. Alpha is considered as the fittest solution, the second and the third are beta and delta respectively. Rest of these are considered as omega. α guided about hunting, during hunting wolves encircling the prey and the final attacking. In smart appliance we initialize the α, β, δ and ω with some value and generate the population Np randomly. Then calculate the objective function in order to calculate the fitness of each search agent as shown in Eq. 9. In Eq. 10, t is the max iteration we performed, A Eq. 7 and C Eq. 8 are coefficient vectors it can be calculate using Eq. 10. Xp is the position vector of the prey, and X indicates the position vector of a grey wolf Eq. 11. Whereas a components are linearly decreased from 2 to 0 over the course of iterations and r1, r2 are random vectors in [0, 1]. The first three best solutions are saved, then other agent are oblige to update their positions according to the position of the best search agents as shown in Eq. 12. By this we get the most optimal solution. In [13], author use these equations (Table 2).

$$A = 2a.r1.a \tag{7}$$

$$C = 2.r2 \tag{8}$$

$$Obj = (E_{Rate}^{hour} x P_{Rate}^{App}) \tag{9}$$

$$D = |C.Xp(t) - A.X(t)| \tag{10}$$

$$X(t+1) = Xp(t) - A.D \tag{11}$$

$$X(t+1) = (X1 + X2 + X3)/3 \tag{12}$$

Table 2. BFA parametric list

Parameters	Values
Elimination and dispersal	24
Reproduction	5
Chemotaxies	5
Population	30
Swim	2
Step size	0.01

5 Simulation Results and Discussion

In this section, simulation results are discussed to evaluate the performance of proposed architecture. Results shows that how effectively proposed architecture performed. Ten times produced the results than average results are listed.

Algorithm 1. GWO for SG scheduling

```
Randomly initialize the position of search agents
Evaluate the position search agents
while iter < iter_max do
   for i = 1:size(Positions, 1) do
      fitness/obj function using Eq. 9
      if fitness < alpha_score then
         alpha_score = fitness;
         alpha_pos = Positions(i : 1);
      end if
      if fitness < beta_score then
         beta_score = fitness;
         beta_pos = Positions(i : 1);
      end if
      if fitness < delta_score then
         delta_score = fitness;
         delta_pos = Positions(i : 1);
      end if
   end for
   a value linearly from 0 to 2
   for i = 1:size(Positions, 1) do
      for j = 1:size(Positions, 2) do
         r1, r2 randomly initialize the value between 0 to 1
         Eq. 7 and Eq. 8 use to calculate the value of A and C co efficient factors
         Eq. 10 and Eq. 11 to update the position of Alpha (X1), Beta (X2), Delta
         (X3)
         Eq. 12 to get the best positions
      end for
   end for
end while
```

Main objectives are to reduce the cost, minimize the PAR and maximize the user comfort. CPP pricing scheme is used to calculate the electricity bill. The cost, waiting time and load for each class is represented in cents, hours and kWh respectively. Figure 3 shows the energy consumption or a load using both techniques according to CPP. Figure 3 clearly illustrated that when scheduled the load according to the price, the peak load during the high price signal is more in unscheduled case as compared to the scheduled. Using BFA and GWO techniques to effectively reduce the cost by shifting the load from PHs to OPHs without affecting the overall load. Figure 4 demonstrate that how effectively load reduced, by shifting the load from PHs as well as PAR minimized. Energy consumption cost per hour during the whole day is shown in Fig. 4, peak price is high in BFA as compared to GWO. Total cost in scheduled and unscheduled cases are shown in Fig. 5 which shows that proposed approaches BFA reduce the cost 45% and GWO reduce 55% cost compared to the unscheduled. Using GWO approach is more effective in cost minimization as compared to the other approach. In Fig. 6, PAR results are shown. Figure 6 clearly demonstrate that

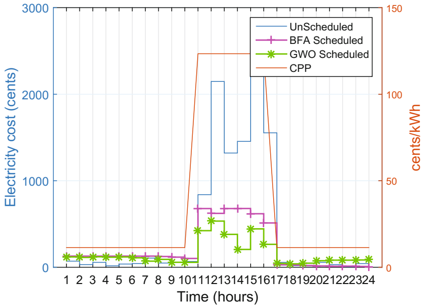


Fig. 3. Electricity cost

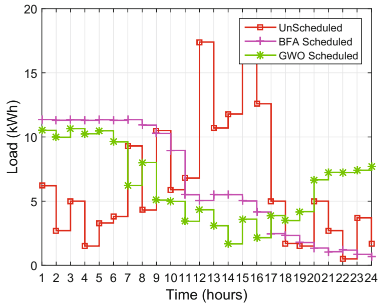


Fig. 4. Energy consumption

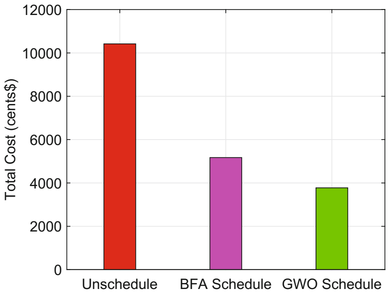


Fig. 5. Total cost

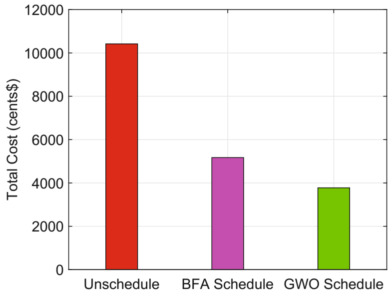


Fig. 6. Peak average ratio

the PAR of BFA is low than GWO and unscheduled. BFA effectively reduce the peak formation, PAR significantly reduce by load shifting. PAR is reduced 40% by BFA as compared to unscheduled load. In Fig. 5 cost of GWO is low compared to BFA so there is trade of between PAR and cost, user need to compromise on one of the objective to effectively achieve other one. UC is measureable in term of waiting time. The average waiting time in of BFA is reduced as compared to average waiting time of GWO. Waiting time is highlighted in Fig. 7. Optimization approaches, GWO and BFA are used to find out the best solution from the given population. GWO work mode depend upon hunting, encircling and attacking nature of the grey wolf whereas BFA depends upon the growth and survival of bacteria. Both optimization approaches are well recognized in order to find the optimum solution. Integrated the functionality of both optimization techniques, for scheduling the appliances, compare the integrated to improve the performance and overcome the deficiencies, limitations of both approaches. Through simulation, we observed that GWO perform best for maximum number of populations required less execution time while in BFA increase the generation steps increased the execution time. BFA shows greatest performance for small population. In short we concluded that GWO is best when user need to minimize the cost but peak value reduced by the BFA.

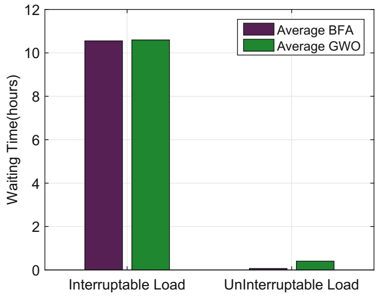


Fig. 7. Average waiting time

6 Conclusion

Proposed techniques are used for HEMS in SG. In this paper, we evaluated the performance of proposed techniques (BFA, GWO) to accomplish the objectives and manage the load of home appliances. The load was balanced as per demand, effectively manage the energy consumption by balancing the load between PHs and OPHs. Our proposed system model use CPP pricing scheme to calculate the electricity bill. Form the simulation results, it is clearly justified that GWO perform better in term of cost reduction. GWO reduce 10% more cost as compared to the BFA approach and manage the energy consumption. However, there is a tradeoff exist between cost and PAR. GWO gives more optimize results as well as the fast convergence. In future, we will focus on integration of RES and using of other algorithms in SG to achieve user comfort and reduction in electricity cost.

References

1. Zhu, Z., et al.: An integer linear programming based optimization for home demand-side management in Smart Grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT). IEEE (2012)
2. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
3. Zhao, Z., et al.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
4. Javaid, N., et al.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Ma, K., et al.: Residential power scheduling for demand response in Smart Grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
6. Rahim, S., et al.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
7. Zhou, Y., et al.: Home energy management with PSO in Smart Grid. In: 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE). IEEE (2014)

8. Liu, Y., et al.: Queuing-based energy consumption management for heterogeneous residential demands in Smart Grid. IEEE Trans. Smart Grid **7**(3), 1650–1659 (2016)
9. Mahmood, D., et al.: Realistic scheduling mechanism for smart homes. Energies **9**(3), 202 (2016)
10. Ma, J., et al.: Residential load scheduling in Smart Grid: a cost efficiency perspective. IEEE Trans. Smart Grid **7**(2), 771–784 (2016)
11. Ranjini, A., Zoraida, B.S.E.: Intelligent residential energy management in Smart Grid. Indian J. Sci. Technol. **9**(45) (2016)
12. Allouhi, A., et al.: Energy consumption and efficiency in buildings: current status and future trends. J. Cleaner Prod. **109**, 118–130 (2015)
13. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)

An Efficient Scheduling of Power and Appliances Using Metaheuristic Optimization Technique

Nasir Ayub, Adnan Ishaq, Mudabbir Ali, Muhammad Azeem Sarwar, Basit Amin, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. Nowadays, Energy become the most valued necessity. Energy crisis becomes a critical issue of this era. Energy demand is increasing day by day, due to which peak load creation occurs. In order to handle the critical situation of the energy crisis, many techniques and methods are implemented. This can be done by replacing the traditional grid with smart grid and scheduling of appliances at Demand Side Management (DSM). Our main focus is on load management and minimization of cost which can be done by load shifting from on peak hours to off peak hours. We have achieved objectives by using two meta-heuristic optimization techniques; Harmony Search Algorithm (HSA) and Earth-Worm optimization Algorithm (EWA). Simulation results show that the approaches we adopted reduce the cost, reduce the Peak Average Ratio (PAR) by load shifting from on peak to off peak hours between the min and max interval with a low difference.

1 Introduction

Science has blessed the human life with many valuable technologies. Electricity is one of the most blessing for human. Electricity is generated by nuclear power plants, hydro power plants and wind power plants etc. as shown in Fig. 1. Electricity is used by industrial and household appliances. As the population increases, the demand of electricity increases. Energy demand is increasing daily with the passage of time [12]. To handle demands, energy provider should stimulate the consumers about the usage of electricity in well-organized manner. The consumption of energy in buildings is much higher than economic sector [13]. This sector needs to be made more economical.

The consumption of electricity can be reduced to some extent without knowing to the user [14]. This can be done by adapting DSM in aspect of electricity demand controlling in a proper way. It is allowed by DSM that the user to be informed about their energy usage [15]. An Energy Management Controller (EMC) is defined at demand side which distributes the energy in efficient way. DSM is a feature of SG which is used for management of energy. DSM is used

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8, DOI 10.1007/978-3-319-65636-6_16

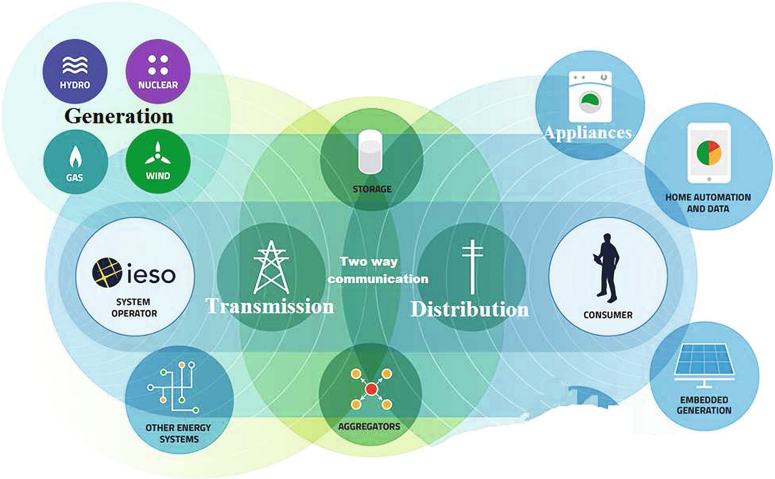


Fig. 1. Smart Grid

in an area like residential, commercial and targets the peak load reduction. Different operational and energy measures are included in SG are smart meters, smart appliances etc.

DSM programs are made for the consumers in order to facilitate them in managing their loads and also provides a balance between supply and demand. The DSM programs encourages the user for shifting their load from high peak hours to the low peak hours. DSM has two main functions; Demand Response (DR) and load management. In load management, the energy is managed in well-manner way. It helps in the reduction of PAR and power consumption. DSM can be explained as deployment, monitoring, and planning strategies of consumer that has an impact on usage and the load [16]. Utility has described pricing schemes for the consumer for the calculation of bill. Dynamic pricing schemes which include Time of Use (TOU), Inclined Block Rate (IBR), and Critical Peak Pricing (CPP), RTP and Day Ahead Price etc. Users are encouraged for shifting their high energy consumption appliances to the low peak hours, which minimizes the electric cost and also reduces PAR. Pricing scheme RTP is considered as more efficient for electricity markets.

The main objectives of our research paper are; Minimization of electric cost and PAR reduction. Many techniques are implemented in recent years to achieve these objectives. In [1,2], integer linear programming, approximate dynamic programming techniques are used to lower the value of electric cost and minimize the PAR. To overcome the lack of the above techniques different types of optimization techniques; meta heuristic techniques can also be used for energy management in SG. In [4,5], author used GA, BPSO and WDO for the reduction of cost and maximizing the UC and peak reduction [8]. A huge amount of energy is consumed in residential area. These are the facts that attracts the attention of researchers towards the scheduling of appliances at home. We use two Meta heuristic algorithms HSA and EWA in order to achieve our objective function.

Section 2 comprises of recent related work. Section 3 depicts the detail description of proposed model. Simulation findings and results are discussed in Sect. 5. Lastly, in Sect. 6, concluding remarks are presented followed by future directions.

2 Related Work

In SG, many methods have been demonstrated and implemented by using different algorithms. The main focus in these techniques is to find an optimum scheduler in aspect of giving benefit to utility and consumer. Different types of scheduling algorithms have been used while considering different parameters in mind such as prices schemes, appliances, user demands etc. While considering these parameters in mind, these algorithms have been implemented.

In [1], author use Integer linear programming (ILP) to schedule the optimal operation time and optimal power for power shiftable and time shiftable appliances respectively, according to the user desire. The proposed technique ILP for DSM in SG is capable of scheduling operation time and load of appliances according to customer perspective. Increasing joint scheduling the load and power trading is another problem with energy renewable resources. A technique proposed in [2] approximate Dynamic Programming [DP] to plan the operation time of appliances and to handle the complexity of tradeoff between UC and electric cost, a benders decomposition approach is used. In this article, plan the operation time of different kinds of appliances and a method was implemented to model the communication of the users with spare power generation. Usage of more energy has been reduced by vend their spare energy to local consumer at a lower price than utility. A user demands for energy from the utility. Utility provides the energy to consumer but consumer didnot utilize all the provided energy. Moreover, some of the energy become spare. User sells that spare energy to another consumer with a low price than utility. User generates their own energy by the integration of RES. User also sells the spare generated energy to other consumers. Author used RTP for calculating the electric bill.

GA is used for the scheduling of home appliances and for decreasing the cost, RTP is used with Inclined Block Rate (IBR) to achieve low cost and PAR [3]. The usefulness of RTP with IBR pricing pattern is very high. The useful features gained by implementing the proposed scheme GA which makes a decrease in the electricity cost as well as delay time of home appliances operations at same time. Reduction in PAR and cost are managed consistently but ignored the user comfort.

Five heuristic algorithms GA, Binary Particle Swarm Optimization (BPSO), BFOA, WDO and a hybrid GWD are used for scheduling appliances in [4]. A tariff signal price RTP is used. Using these heuristic algorithms minimization in electricity cost, PAR reduction and UC was achieved. Hybrid GWD performs best in minimization of cost than the other heuristic algorithms. GA provide an optimal solution for the scheduling of appliances during off peak hours and on peak hours. The GA beat WDO, BFOA and BPSO in aspect of electricity bill cost and energy consumption. Good results are achieved for both single and multi-homes.

To achieve effectiveness in terms of energy consumption, electricity bill, PAR, UC level, execution time, three meta heuristic algorithms are used GA, BPSO, and Ant Colony Optimization (ACO) [5]. GA performs better comparing with other two techniques in aspect of PAR reduction, minimize electric cost, minimize waiting time and optimal integration of RES. Other parameters are achieved but neglects the RES integration cost and power consumption management. [6] Author pointed the unscheduled way of appliances and the maximization of cost in a residential area. Peak load is created in few hours in the day where the demand and consumption of energy is high. To tackle this problem, author used BPSO optimization technique for scheduling of appliances and TOU price signal. It achieves reduction in the PAR by shifting load from high peak to low peak hours. UC are sacrificed and integration of RES is neglected. There is a contradiction between electric cost and UC. In order to reduce the power consumption, it gains reduction in payments with no gain in user comfort.

Optimization power scheduling scheme is used for reducing energy consumption and day ahead price including integer and continuous variable. Using the proposed technique, achieved the reduction in electricity bill [7]. A desired tradeoff is also achieved between payment and user comfort. To maximize UC along with the reduction in electricity cost, a heuristic technique WDO is used in [8]. A model is created by categorizing the appliances into three categories based on power usage model and UC. The classification of appliances is grounded at hourly electric price TOU during low peak hours and high peak hours.

Another model is proposed in [8] is Knapsack based on Wind Driven Optimization (KWDO). KWDO is used to maximize electric cost saving, which is further used as comparison of performance estimation of energy consumption in HAN. In the identical paper min-max regret based knapsack problem is used to reduce the higher cost and user peak load. Optimization practices are used for arranging appliances. During peak hours, RES is incorporated; in order to achieve grid stability, electric cost reduction and maximum UC. This produced a tradeoff between saving the cost and appliance waiting time, finally effect UC. However, achieving the cost reduction along with UC and also integrating RES, installation cost of RES is not defined, which directly effect on the total cost. Energy gain from Grid or RES, when not used is wasted.

An innovative power structure along with bidirectional information movement among service provider and users is employed to achieve balance load and diminish demand supply mismatch. Renewable energy created by consumers which can be vended to further consumers and also to the grid. In this regard, a novel Prosumer based Energy Sharing and Management (PESM) scheme for corporative DSM is presented in [9]. Prosumer; which consume energy as well as vend/share more energy produced by RES with grid or further consumers at community. The PESM scheme incorporate with prosumers that are connected to grid and generation of energy through RES. The scheme minimizes demand supply miss-match and give priority to users; which have shortage of energy at any time. Priority of users is based on different criteria; user who have greater capacity having higher the priority and user who have lower capacity having the low priority. Users having shortage and highest installed renewable capacity will be served first. Usage of energy become efficient by the sharing of energy. As Prosumer can share energy only to grid or within same community [9]. The limitation is that it cannot share energy to another community.

In [10], a system is proposed to develop a generic DSM model for domestic users to achieve low value of PAR, total energy cost, and appliances waiting time along with fast performance algorithm. GA based method is used for scheduling appliances in smart grid situation. Scheduling the appliances by GA technique are taken from utility accordingly to RTP values. The consumption of power in particular time slot can be managed by using the shifting load strategies as a replacement for the reduction in load. Those appliances are scheduled only which can be delayed. The execution stops when the achieving of best solution is confirmed. Appliances that are the most critical for UC and cannot tolerate delay are considered as NDAs. Many of NDAs have small power consumption and little influence on the total power consumption of a user if it is kept in some described range. The GA technique is best to schedule every appliance for Energy Management Controller Unit (EMCU) will select a suitable beginning time of every appliance and regarding the RTP and power capacity limitations. It causes a decrease in PAR and saving electric cost.

3 Problem Statement

In recent years, the problems of energy management in SG are tackled by different optimization techniques. It is difficult to handle the schedule of energy consumption, minimum cost and maximum the User Comfort (UC) in SG. Mostly PAR and UC is neglected. In SG, using a smart meter creates a two-way communication between the utility and consumer. Smart homes uses smart meter to manage its energy consumption. In SG, energy optimization problems are:

- Minimize the electric cost
- Minimize energy consumption
- Minimize the creation peak load
- Minimize waiting time of appliances to achieve user comfort

In some hours of the day, when the consumption of energy is comparatively higher than the demand. It creates a peak in load, at that hours the price of consumption of electricity is comparatively high, which are on peak hours. If the same situation of peak load creation is happened at off peak hours, it changes into on peak hours by utility [7]. To avoid peak load creation, we have proposed an optimization algorithm which schedules the appliances in such a way that energy consumption is carried out efficiently (Fig. 2).

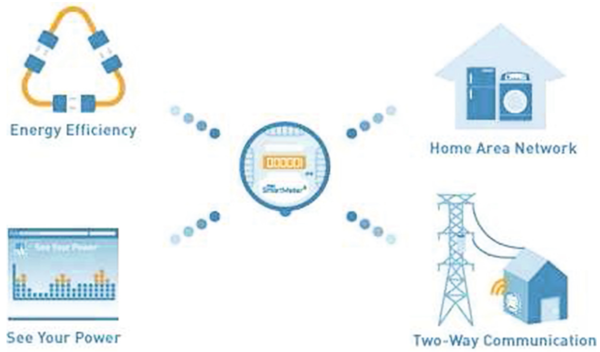


Fig. 2. Smart residential DSM components

4 System Model

In SG architecture, DSM facilitates more proficient and reliable user tasks. The main functions of SG are energy controlling with demand side switch activities for users. In residential area, every home is equipped with smart meters. Demand side activates for the end users are controlled by DSM. It educate the consumer to consume the most of its energy at off peak hours. To solve the optimization problem, several of optimization techniques are used. In our proposed model, we have categorized 6 appliances into 3 classifications; interruptible, non-interruptible and fixed appliances. The categorization of appliances in our paper is according to [7]. Classified appliances are mentioned in Table 1.

Table 1. Parameters of appliances

Appliance	Power (kWh)	LOT (hours)
Washing machine	0.78	5
Air conditioner	1.44	6
Refrigerator	0.50	24
Lighting	0.6	8
Toaster	0.5	2
Kettle	0.8	1

In addition to this we have used TOU as pricing scheme for calculation of bill. The main objective of all this study are: minimize the consumption of energy in order to reduce the electricity cost and reduction in PAR. The main focus is on the minimization of total cost calculated according to the Eq. 1 with PAR reduction.

$$Cost = \sum_{t=1}^{24} \left(E_{Rate}^{hour} * P_{Rate}^{App} \right) \tag{1}$$

$$Load = P_{Rate}^{App} * App \tag{2}$$

$$PAR = \frac{max(Load_s)}{Avg(Load_s)} \tag{3}$$

For optimizing the electricity consumption in HEMS we have used bio-inspired meta-heuristic algorithm; earthworm optimization technique [11]. For that we have initialized the population size as 30 and initialized the iteration as maximum generation index of 50. In EWA, there are two kinds of reproduction; reproduction 1 and reproduction 2. In EWA, reproduction 1 generates only 1 offspring either from male or female while in reproduction 2, generates one or more than 1 offspring at one time. In between them crossover operators are used in order to improve the version of crossover then do mutation is implemented to extract the best value after the iterations. Algorithm of EWA is given below as described in [11]. Next is to compare EWA with BFA. We have merged both technique for same appliances classification. In order to define the solution for better exploration, we have shown the performance of unscheduled appliances as well as scheduled appliances.

4.1 Optimization Techniques

4.1.1 EWA

The reproduction conduct of earthworms state multiple optimization problems, the reproduction steps of earthworms can be perfect by the following guidelines.

- Every earthworm have the ability of producing off springs and each earthworm individual have two types of reproduction.
- Every child of earthworm singular generated holds all the genetic factor whose length is equivalent to parental earthworm.
- The earthworm singular with the finest fitness permit on straight next generation, and cannot be altered by operators. This can be an assurance that population of earthworm cannot fail in the increment in generations.

4.1.2 HSA

It an evolutionary algorithm that have musician's behaviour. Main steps that are involved in HSA are: memory based play, random play and pitch adjustment.

5 Simulations

To assess the performance of our proposed optimization techniques, we have carried out extensive simulations in MATLAB. In these experiments/simulations, we compare our objectives i.e. PAR reduction, pattern of energy consumption, electricity bill reduction and UC (Tables 2 and 3).

Algorithm 1. EWA for SG scheduling

```
Step 1: Start Initialization. At first Set t = 1 which is generation counter
Set the counter of generation t=1
Set population as P of NP earthworm
Select the individual in search space randomly
Set the numbers of earthworms kept nKEW, maximum generation MaxGn, α as
  similarity factor, proportional aspect β, constant γ = 0.9.
Step 2: evaluation of Fitness. Set every earthworm aspect to its position
Step 3:
while till best solution is not achieved or t < MaxGen do
   sorting all earthworms according to the fitness value
   for i = 1 to NP (all earthworm) do
      Generate offspring xi1 through Reproduction 1
      Generate offspring through Reproduction 2
   end for
   Do crossover
   if i > nKEW then
      set the number of particular parents (N) and the produced off springs (M)
      Select the N parents using method i.e. roulette wheel selection;
      Generate the M off springs;
      Calculating xi2 according to offsprings M generated
   else
      Randomly an individual earthworm as xi2
      Update the location of earthworm
   end if
   for j = nKEW + 1 to NP (earthworm individuals non-kept) do
      do Cauchy mutation
   end for
   Calculate the population according to the newly restructured positions;
   t = t + 1.
end while
Step 5: Show Output of the best solution.
End.
```

Table 2. HSA parametric list

Parameters	Values
Appliances	6
Reproduction	5
Max iteration	100
Population	50

Table 3. EWA parametric list

Parameters	Values
Max Itr	50
Population	30

Algorithm 2. HSA for SG scheduling

Step 1: Parameter and problem initialization
Harmony memory size (HMS)
Harmony memory considering rate (HMCR)
Pitch adjusting rate (PAR)
Stopping criteria
Step 2: HM initialization
Initialize the population randomly
Step 3: create a new harmony
Memory consideration
Choosing any value from HM
Value of HMCR specify the probability of selecting value from previous values that are stored in HM
Pitch adjustment
Every component of the New Harmony chosen from HM, is likely to be pitch-adjusted.
Similar to Mutation procedure in genetic algorithm
Random selection
Take a possible range and select random values
Increase the diversity of the solutions
Step 4: Update the Harmony memory
if new values of HM are better than previous worst harmony, then replace the previous worst harmony with new one in the HM **then**
    if xnew $<$ xworst **then**
        Update the HM as xworst = xnew
    end if
end if
Step 5: Termination
If the given criteria is satisfied, selection and calculation terminated.
Otherwise, repeat step 3 to 4

In Fig. 3, the consumption of energy of electrical appliances in the EWA algorithms is low during the (0 to 3) h, while with HSA algorithm appliances consume more energy. During high peak hours (7 to 9) h, In comparison to HSA algorithm, energy consumption of appliances is very low in EWA algorithm. During (10 to 24) h, the normal consumption of energy of both the unscheduled and scheduled cases is the same. Most of the appliances are scheduled by EWA algorithm for a low electricity bill. While the electric prices are very high during (6 to 9) h and (11 to 13) h and maximum number of appliances in these time slots are scheduled by HSA algorithm. On other hand, the EWA algorithm uses the low peak time slots and donot turn on any appliances during the high peak hours and also complete the working cycles. We come to conclude that the EWA algorithm reduce the energy consumption efficiently by scheduling the appliances in mid and low peaks hours.

Figure 4 elaborates the comparison of HSA and EWA algorithms in terms of electric cost. It is clear from Fig. 4 that Energy Management Controller (EMC) schedules the appliances in low time slots to minimize the energy consumption.

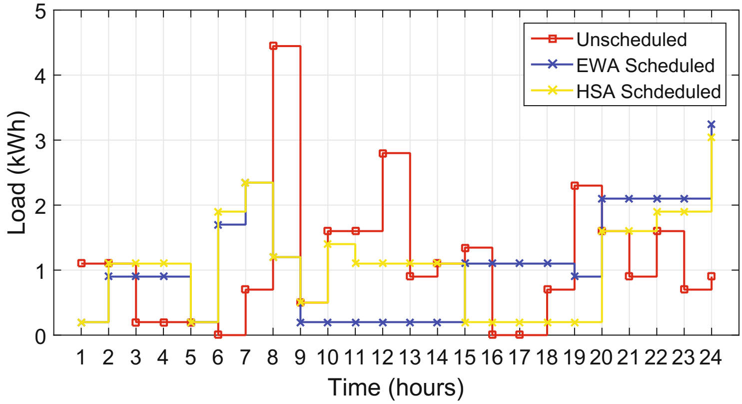


Fig. 3. Load with CPP

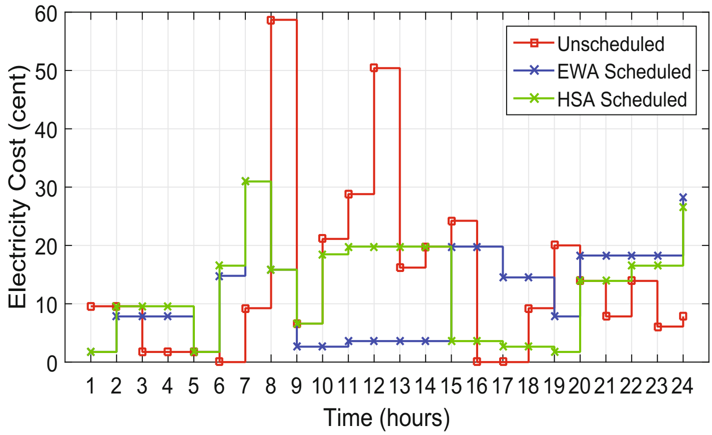


Fig. 4. Cost

During (0 to 6) h, the electricity costs that is scheduled by EWA and HSA algorithms are relatively same, because the scheduling of appliances by EMC is according to the low pricing slots without taking into consideration of the maximum capacity of appliances limit. During high price hours (7 to 9) h, the electric cost of HSA algorithm is higher than the EWA algorithm. It is because greater number of appliances are scheduled by EMC in these time slots. During the shoulder peak hours (10 to 13), HSA turn on a greater number of appliances in comparison with EWA, so the electric bill cost is higher. During the remaining hours (15 to 24) h, the working times of all the appliances are completed. It is clear that most of the electric bill is low during these time slots. In Fig. 5, we come to conclude that EWA algorithm performs better in terms of electric cost than the HSA algorithm.

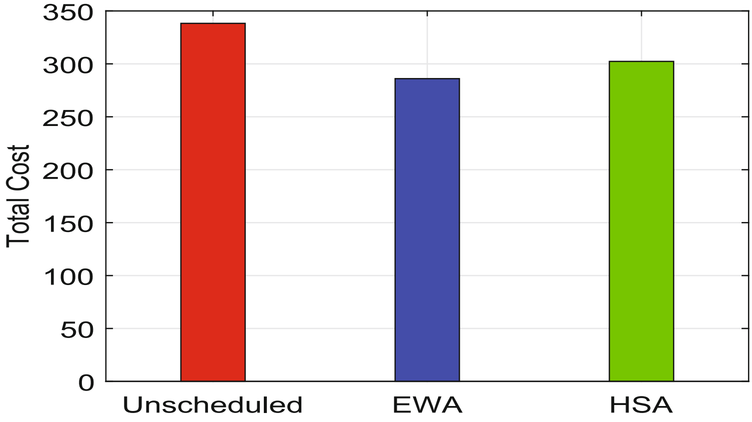


Fig. 5. Total cost

At the start of our discussion, we resulting the reduction of PAR in the residential load when we use our proposed energy optimization algorithm. Consumer wants to minimize their total electric bill, while the utility is attentive to provide balanced energy supply. Figure 6, it is clearly shown by our proposed algorithm that it is very helpful in PAR reduction and balancing the consumption of energy. It is also clear from the Fig. 6 that EWA algorithm shows good performance in the PAR reduction than HSA algorithm. As EWA algorithm reduce the PAR by 6.8 % while HSA algorithm reduce the PAR by 9%. It is because avoiding the creation of peak load congestion by our proposed algorithm, it optimally schedule the home appliances in that hours having low price.

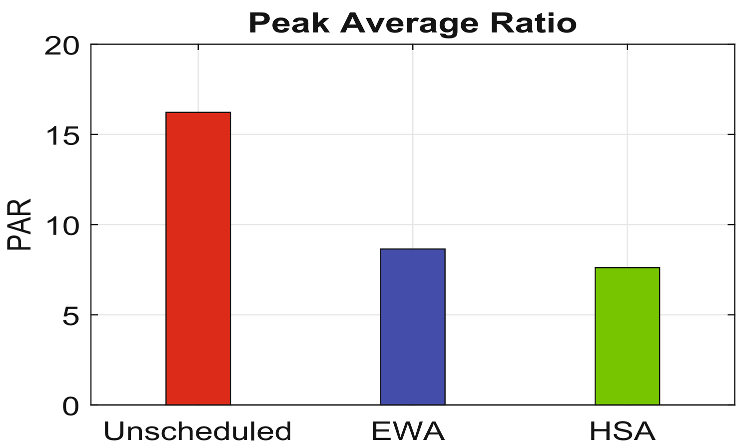


Fig. 6. Peak to average ratio

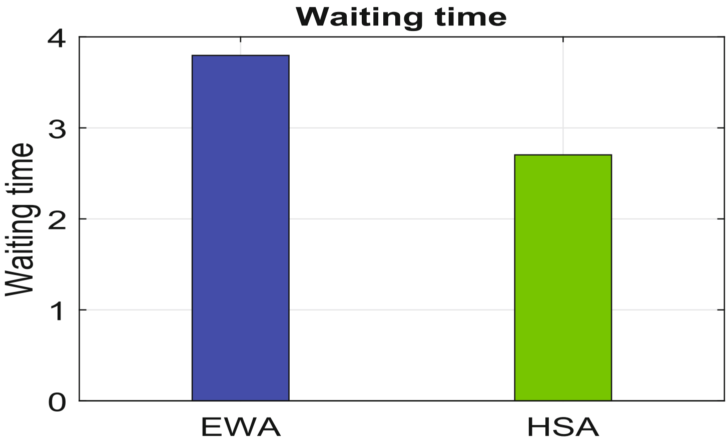


Fig. 7. Waiting time

To achieve reduction in electricity bill, smart user will must follow the scheduling strategy of EMC to operate all appliances. According to scheduling perspective, starting time of any appliance cannot be fixed because variation of price at every hour. Therefore, the scheduling algorithm that we proposed adjusts the initial time of those appliances which are considered as maximum cost saving perceptive. However, this mechanism save the electric cost bill but can ultimately disturbs the life style of the user. Alternatively, scheduling appliances algorithms can also be made to maximize the UC but it will increase the electric cost bill. There is a contradiction between the two objectives which is hard to achieve them instantaneously. The HSA algorithm is designed for those consumers who have no objection on electricity cost and cannot bear compromise on comfort. While the EWA algorithm is considered as for those consumers who are sensitive about electric bill and can on compromise UC. It shows that there is a contradiction between price and waiting time. If price increases the UC decreases and vice versa. Therefore, we designed a scheduling algorithms to minimize the appliances waiting time and electric cost. Figure 7 shows that EWA has more waiting time in appliances than the HSA.

6 Conclusion

In this paper, we have proposed a heuristic optimization techniques for the scheduling of appliances at residential side to avoid peak creations while focusing on the electricity bill reduction by preserving user comfort level to an acceptable limits. We evaluate our designed objective functions using two heuristic algorithms (EWA and HSA) and analysis the comparison for all of them. Our proposed model used TOU as pricing scheme for bill calculation. It is clearly justified that the model we proposed works efficiently with EWA than HSA in terms of PAR reduction, electric cost minimization while considering the UC. In future, we will focus to achieve the user comfort level and to reduce the frustration cost. We will also work on the different optimization techniques in order to achieve more accurate data in less execution of time.

References

1. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: A n integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–5. IEEE (2012)
2. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
3. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
4. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
6. Ullah, I., Javaid, N., Khan, Z.A., Qasim, U., Khan, Z.A., Mehmood, S.A.: An incentive-based optimal energy consumption scheduling algorithm for residential user. Procedia Comput. Sci. **52**, 851–885 (2015)
7. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
8. Ahmad, A., Alrajeh, N., Javaid, N., Khan, Z.A., Qasim, U., Rasheed, M.B.: An efficient power scheduling scheme for residential load management in smart homes (2015)
9. Razzaq, S., Zafar, R., Khan, N.A., Butt, A.R., Mahmood, A.: A novel Prosumer-Based Energy Sharing and Management (PESM) approach for cooperative Demand Side Management (DSM) in Smart Grid. Appl. Sci. **6**, 275 (2016)
10. Khan, M.A., Javaid, N., Mahmood, A., Khan, Z.A., Alrajeh, N.: A generic demand-side management model for smart grid. Int. J. Energy Res. **39**(7), 954–964 (2015)
11. Wang, G.-G., Suash, D., Coelho, L.D.S.: Earthworm optimization algorithm: a bio-inspired metaheuristic algorithm for global optimization problems. Int. J. Bio-Inspired Comput. **1**(1/2), 14–31 (2015)
12. Energy Information Administration. https://www.eia.gov/todayinenergy/detail.cfm?id=12251
13. Logenthiran, T., Srinivasan, D., Shun, T.Z.: Demand side management in SG using heuristic optimization. IEEE Trans. Smart Grid **3**(3), 1244–1252 (2012)
14. Palensky, P., Dietrich, D.: Demand side management: demand response intelligent energy systems and smart loads. IEEE Trans. Industr. Inform. **7**(3), 381–388 (2011)
15. Gellings, C.W., Chamberlin, J.H.: Demand-Side Management, pp. 1–5. EPRI, Palo Alto (1988)

Demand Side Management Using Harmony Search Algorithm and BAT Algorithm

Mashab Farooqi, Muhammad Awais, Zain Ul Abdeen, Saadia Batool, Zunaira Amjad, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. In this paper performance of Home Energy Management System (HEMS) is evaluated using two meta-heuristic techniques: Harmony Search Algorithm (HSA) and BAT Algorithm. Appliances are classified into three categories according to their characteristics. Critical peak pricing is used for electricity price calculation as electricity pricing scheme. The main purpose is electricity cost reduction, electricity consumption, peak to average ratio reduction and maximizing User Comfort (UC) by reducing waiting time. Simulation results show the overall effectiveness of HSA.

1 Introduction

With the advent of the new technologies, human life has become very easy. These technologies help man to achieve their desires. Electricity is one of the greatest invention ever, it made human life very easy many electric appliances have been invent till the day which are facilitating the man kind in different ways. The manufacturing of these appliances is increasing because of the increasing human population, so the demand of electricity is increasing day by day. Electricity is generated through different sources like hydroelectric plants, nuclear plants, thermal power plants, wind mills, solar panels etc. The demand is still increasing and the traditional electricity grid are not enough to meet the current demand of electricity efficiently. This problem is the motivation for introduction of the Smart Grid (SG). SG introduces communication and information technologies in traditional grids. It includes smart monitoring systems, smart meters, energy efficient resource etc.

Smart meters facilitate bi-directional communication between electricity users and utility. The information achieved by this communication is used for energy optimization. One of the most important aspect of SG is Demand Side Management (DSM). DSM encourages users to consume less energy during the peak hours, shifting the load to off-peak hours. It encourage users to store energy during the off-peak hours and utilize them during the peak hours. Consumers can share the burden of utility as they can sell extra electrify to the utility.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_17

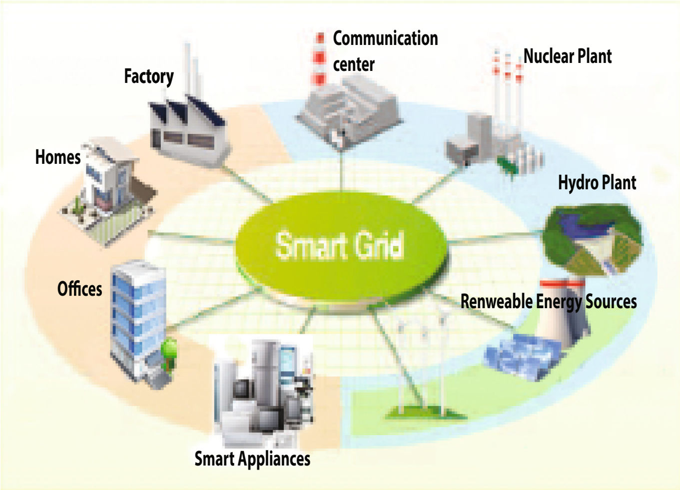


Fig. 1. Smart Grid

Hence, consumers can make profit. The main purpose of DSM is cost reduction and load management, and another purpose it to reduce Peak to Average Ratio (PAR). PAR measures that how much hourly demand is higher than the average hourly demand. High PAR can stress the grid to its limits as the grid has certain threshold above which grid can be damaged. Using DSM electricity providers can prevent system failures if PAR rises too much. As the demand is increasing continuously, this could results a discrepancy in demand and supply (Fig. 1.).

Different pricing schemes have been introduced to calculate the electricity cost, some of them are dynamic pricing schemes and some are flat rate pricing schemes. The dynamic pricing schemes best reflect the supply and demand relationship. Time of Use (TOU), Real Time Pricing (RTP), Inclined Block Rate (IBR), Critical Peak Pricing (CPP), and Day Ahead Pricing (DAP).

Electricity cost minimization, PAR reduction, load management, maximize UC by reducing waiting time are the common objectives of energy management in SG. Many optimization techniques have been used in recent years in DSM to archive above objectives. In [1], Dynamic Programming and Game Theoretic approach is used to achieve reduction in electricity cost and in [2], Genetic algorithm (GA) is used to reduce PAR and electricity cost. Integer Linear Programming (IPL) is used in [3], to archive minimization of peak hourly load. In this paper two techniques: Harmony Search Algorithm and BAT Algorithm are comparatively evaluated. Our main objective is to reduce the electricity cost, PAR, energy consumption and waiting time. Reduced waiting time increases UC but there exists a trade-off between UC and electricity cost.

Rest of the paper is arranged as follows: Sect. 2 presents related work, proposed model is presented in Sect. 3, simulations and results are illustrated in Sect. 4. At the end is the conclusion Sect. 5.

2 Related Work

Many researchers from all over the word are working to optimally schedule the Smart Grid appliances. Some of the work is discussed below.

The paper [4], discusses the scheduling for the residential consumers in the SG. The study conducted to remove the trade-off between the discomfort and the cost by scheduling appliances. The changes which the authors encounter: (1) modeling of the discomfort cost for the appliances. (2) formulation of the optimization problem that achieved the trade-off between UC and the cost by tuning the parameters used by the scheduling algorithm. DAP signal is used. The time horizon is 24 h. Three operation modes are discussed by the author and the comparison between them is done. The power utilization for the entire 24 h under the DAP cost are dispersed more consistently than that under the level cost. PAR is also reduced. The limitation is that the paper does not discussed about waiting time. With the approach of SG clients have chances to plan their energy utilization, in [1], author concentrated on the issue of load management and power exchanging with high infiltration of Renewable Energy Resources (RERs). In systems with high entrance of RERs, the power may spill out of Distributed Grids (DGs) to the substation, which has negative effect on the security of the framework. To address the issue of abundance generation between the clients a Game Theoretic approach is used. With the help of model that is proposed by the author, each user tends to obtain a larger share in the market and to maximize its revenue by approximately selecting its price and generation. The proposed calculation diminishes the vitality installment of clients, contrasted with the situation where exchanging is not applied. It encourages the coordination of RERs and mitigates the turn around power stream issue. Energy cost is reduced in the paper. Limitation of the paper is that the RERs have high installing and maintaining cost. PAR is not considered by the authors. In [5], authors designed a heuristic algorithm-based energy management controller, which is for the scheduling of appliances in a residential area in SG. In this paper five heuristic algorithms, the binary particle swarm optimization (BPSO) algorithm, GA, the bacterial foraging optimization algorithm (BFOA), the wind-driven optimization (WDO) algorithm and the proposed hybrid genetic wind-driven (GWD) algorithm) are compared and evaluated. RTP is used as the electricity price signal. The main focus of the authors is to maximize the UC and to minimize both electricity cost and PAR. The current versions of GA, BPSO, WDA and BFAO are modified in this paper for the optimization. Authors considered cost, energy consumption, PAR and UC in the paper. As for the UC there always exit a trade-off between UC and electricity cost. Consumers and producers can meet the demand of each other due to SG applications and are taking part in the DSM and Demand Response (DR) program. The main problem in SG is energy optimization and electricity cost reduction in today's excess demand of energy. In [6], the techniques used by authors are: BPSO, GA, PSO using the same data set for each technique. The best results show the technique which is better. The results shows that the BPSO outperforms the other two techniques. UC, PAR and energy minimization is not considered by the authors in this paper.

In [2], author presented a general engineering of energy management framework in a Home Area Network (HAN) in view of the brilliant network and afterward propose a productive technique for home power utilization. The whole appliances may operate during low cost hours that can increase the PAR and can damage. Proposed model is the combination of RTP and IBR. GA is used because of the non-linearity of problem. The limitation of the proposed model is that authors did not consider the User Comfort (UC). In [7], performance of Home Energy Management (HEM) controller, designed on the basis of heuristic algorithms GA, BPSO and Ant Colony Optimization (ACO) are comparatively evaluated. Authors proposed a generic architecture of DSM by integrating residential area domain and smart area domain via Wide Area Network (WAN). Problem is formulated via Multiple Knapsack Problem (MKP). Combined model of TOU and IBR is used for electricity pricing. Paper focuses on getting the feasible solution for the designed model and to compare its efficiency with regards to energy consumption, electricity bills, PAR, UC, and execution time. RESs are integrated efficiently in the designed system by modification in the heuristic algorithms. In terms of electricity bill reduction, minimizing PAR while taking into account the user satisfaction, the proposed model efficiently works with GA-EMC than BPSO-EMC and ACO-EMC. Execution time for GA-EMC is less than others. In this paper the author do not considered the integration and installation cost of RESs, also there will be some maintenance cost for the RESs which is also not addressed.

In [3], a scheduling mechanism for power consumption in home area load management in Smart Grid using ILP technique is proposed. The problem addressed by the author is to minimize the peak hourly load in order to attain a balanced load schedule. Proposed model is successful to schedule both the optimal power and the optimal operation time for power-shiftable appliances and time-shiftable appliances. Maximum load is reduced, performance and efficiency of power grid is improved due to the reduction in peak load. PAR and UC is not considered. In [8], an hour wise approach is introduce for minimization of peak demand and the utility cost, starting from start to the end of the day. The author achieve the objective of load shifting through minimization problem. PSO algorithm has been modified for the DSM problem. Author formulated a minimization function, set the constraints for the devices and calculated controllable devices for residential, commercial and industrial areas. Authors are able to achieve the desired results very closely to the expected results. The peak demand and electricity cost for all three areas is reduced averagely about 19.8% and 11.7% respectively. UC, PAR and power consumption is not addressed in the paper. DSM should be efficiently managed using some dynamic power pricing in the SG. However, in the absence of a proper and detailed consumption model it is difficult to set an optimal power price. In [9], authors use user-independent acceptance price to efficiently capture the user consumption behavior. Every user can decide its own acceptance price according to his/her need. A model is presented, which helps the user to choose their acceptance prices without information exchange and cooperation. The authors achieve the desired results with the proposed scheme, the operator maximizes its profit and consumers reduces their electricity price because of the optimal utilization of the renewable power. Author proposed a very efficient method for both the consumer and the operator but PAR and UC is not addressed by the author. In [10], authors present an exact and heuristic algorithm. The heuristic algorithm select the suitable time to schedule the appliances. Greedy strategy is used for scheduling the appliances without any back tracking. The result showed that the heuristic algorithm performs better and it seems a promising candidate for integration into an optimization approach for smart home appliances. The PAR and UC is ignored in the paper. In [11], the author addressed the problem of optimal energy use. The multi-objective mixed integer nonlinear programming (MINLP) model is proposed by the authors for optimal energy use and creating a equilibrium between energy saving and comfortable lifestyle. The simulation is performed using the real data. The results show that the proposed model achieved the desired results as it reduces the domestic energy and ensure the optimal task scheduling and a thermal comfort zone for the inhabitants. The PAR is not addressed in this paper. The authors in [12], addressed the problem of optimal management for a SG and the optimal power dispatch problem for 24-h for distributed systems. A new criteria is formulated by the authors for shiftable load. Number of variables and the adoption of real valued optimization methods are reduced due this new formulation. The original extension of Golwswarm Particles Optimiztion (GSO) is applied which is a novel nature inspired multi-objective optimization algorithm. GSO is enhanced using the Pareto optimality concept. The author achieve the desired results but did not consider the PAR and UC.

3 Proposed Model

With the advent of the SG, consumers are moving towards the HEMS. Smart meters are installed in the houses which are in SG. These meters allow two way communication between consumers and the utility. Utility is informed about their customers electricity usage and pattern. The main purpose of the DSM is to reduce electricity cost and provide user a comfort in cost, also it reduces PAR because if PAR is too high utility may suffer and there are also changes of system damage. Different pricing are used for the electricity cost calculation with respect to different countries. In this paper we are using TOU pricing scheme. In TOU prices are high for the peak hours. We are using two techniques: HSA and BAT algorithm for the reduction of electricity bill, load, PAR and UC. Classification is used from [7] after changing LOT for some appliances. We are taking single home with thirteen appliance which are categorized into three groups according to their characteristics. We take 24-h time slots as the operational time (Fig. 2).

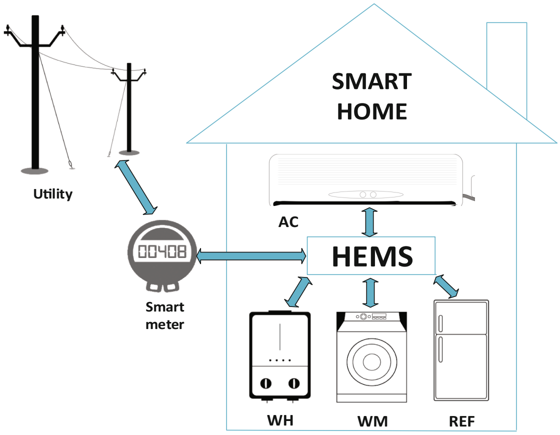


Fig. 2. HEMS

3.1 Load Classification

Load is classified into following three categories:

3.1.1 Fixed Appliances

Fixed appliances contains the appliance whose hours cannot changed. They must have to work in the defined working are. These appliances are: Lights, Fans, Oven, Toaster, and Coffee Maker. Table 1 shows Fixed Appliances.

Table 1. Fixed appliances

Fixed appliances	
Appliances	Power (kWh)
Lighting	0.6
Fans	0.75
Microwave oven	1.18
Toaster	0.5
Coffeemaker	0.8

3.1.2 Shiftable Appliances

These appliances can be ON in any part of the day their operation time is defined. These appliances are: Washing Machine, Dish Washer, Cloth Dryer. These appliances are shown in Table 2.

Table 2. Shiftable appliances

Shiftable appliances				
Appliances	Start time (hours)	End time (hours)	OT	Power (kWh)
Washing machine	9	16	4	0.78
Dish washer	8	13	5	3.60
Clothes dyer	7	18	4	4.40

3.1.3 Elastic Appliances

They are interruptible and fully controllable appliances. These appliance include Air Conditioner, Water Heater, Space Heater. Elastic Appliances are shown in Table 3.

Table 3. Elastic appliances

Elastic appliances			
Appliances	OT	Total hours	Power (kWh)
Air conditioner	6	24	1.44
Water heater	6	24	4.45
Space heater	6	24	1.50

3.2 Optimization Techniques

In this paper two techniques are comparatively evaluated.

3.2.1 HSA

HSA is evolutionary algorithm that was inspired by the behaviour of musicians. It has three steps:
STEP 1: Creating the harmony memory at start randomly, using Eq. 1.

$$x_{ij} = l_j + rand().(U_j - l_j) \tag{1}$$

STEP 2: When the initial harmony is created, generate new vectors, which is depend on the harmony memory consideration rate (HMCR) and pitch adjustment ratio. If the number generated is less than HMCR then the value for harmony is chosen from the stored value.

$$v_{ij} = \begin{cases} x_{randj}, randb() < HMCR \\ l_j + rand().(U_j - l_j), else \end{cases} \tag{2}$$

STEP 3: Elements will be modified according to pitch adjustment rate.

$$v_{ij} = \begin{cases} v_{ij} + -rand().bw_j, rand() < PAR \\ v_{ij}, else \end{cases} \tag{3}$$

3.2.2 BAT

BAT algorithm is a meta heuristic optimization algorithm developed by Xin-She Yang in 2010. This BAT algorithm is in view of the echolocation conduct of micro bats with shifting heartbeat rates of discharge and uproar. A few bats have developed an exceptionally modern feeling of hearing. They emanate sounds that bounce off of objects in their way, sending echoes back to the bats. From these echoes, the bats can decide the measure of items, how far away they are, the means by which quick they are voyaging and even their surface, all in a brief instant. Bats fly haphazardly with speed vi at position xi with a recurrence fmin, changing wavelength and commotion A0 to look for prey. They can consequently modify the wavelength (or recurrence) of their transmitted heartbeats and change the rate of heartbeat discharge r [0, 1], contingent upon the vicinity of their objective. Inspite of the fact that the loudness can shift, we accept that the loudness fluctuates from a large (positive) A0 to a minimum value Amin. Following equations are used for the generation of population and for the optimization. Following equations are used for population generation.

$$f_i = f_{min} + (f_{max} - f_{min})\beta \tag{4}$$

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x_*)f_i \tag{5}$$

$$x_i^{t+1} = x_i^t + v_i^t \tag{6}$$

Algorithm for the proposed technique is written below

```
initialization (POP,D,Fmin,Fmax);
while number of max iterations do
    Generate new population by adjusting pulse rate and location and loudness;
    if rand < r then
        select the best solution and generate local solution;
    else
        generate new solution;
    end
    if rand < r and xi <x then
        Accept new solution ri++ and ai–;
    else
        repeat;
    end
    find the current best solution;
end
```

Algorithm 1. BAT algorithm

4 Simulation and Results

To perform evaluation of HSA and BAT we conduct simulations in MATLAB. Eleven appliance were taken which are further classified into three categories:

Fixed Appliances, Shiftable Appliances, and Elastic appliances. Fixed Appliances includes lightning, fans, microwave, toaster and, coffeemaker. These appliance have fixed hours and must have to work in the defined hours. Shiftable Appliances contain washing machine, dish washer and, cloth dryer. These appliances can be shift between defined hours and have defined Length of Operation Time (LOT) for a day. Elastic Appliances include air conditioner, water heater and, space heater. Their LOT is defined however they can be operated during any hour in a day. CPP pricing model is used for the calculation for the electricity cost. CPP is shown in Fig. 3.

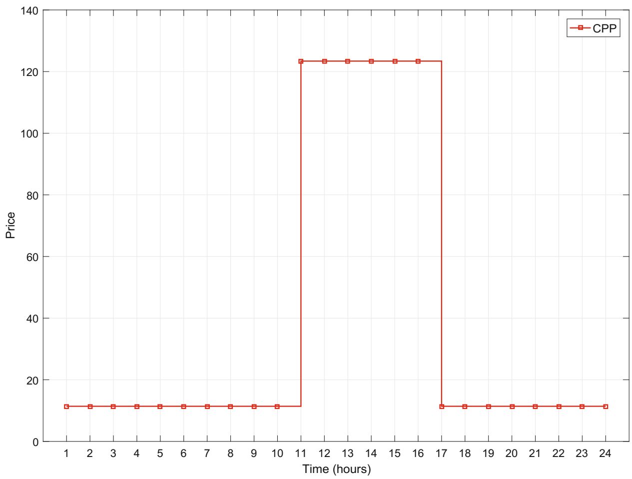


Fig. 3. CPP pricing signal

Figure 4 illustrates electricity cost for the BAT and HSA algorithms. The figure illustrates per hour cost for these algorithms. According to results the electricity cost during the peak hours is less for both BAT and HSA than unscheduled cost. Cost is reduced because load is shifted form on-peak hours to off-peak hours. Here BAT algorithm is outperforming HSA in cost reduction. Figure 5 illustrates the energy consumption for every hour, both before and after scheduling. It is clear that more load is shifted towards off-peak hours for electricity cost reduction. Maximum consumption in unscheduled case is 14.9 kwh which is reduced to 12.6 kwh in case of BAT algorithm and 11 kwh in case of HSA. For HSA load is reduced by 11% as compared to BAT and 26% as compared to unscheduled load. It is clear from the results that in case of electric consumption HSA performs better than BAT algorithm.

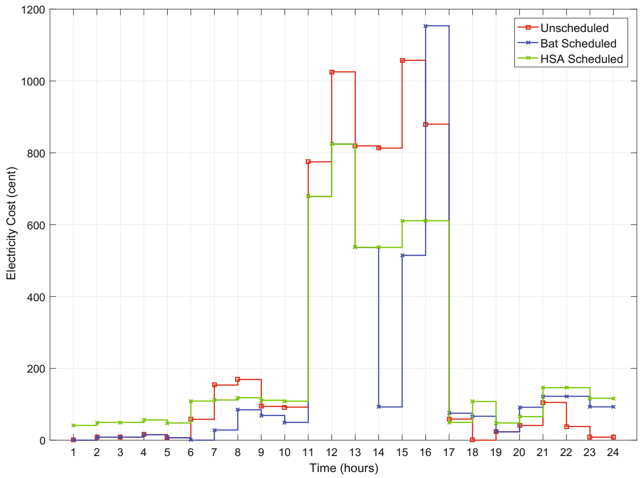


Fig. 4. Electricity cost

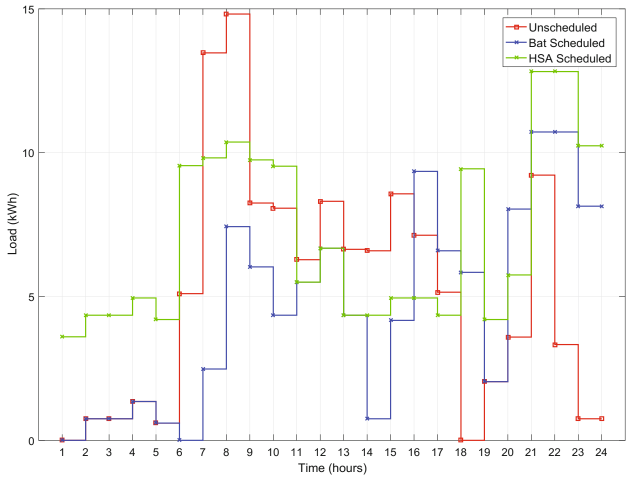


Fig. 5. Energy consumption

Figure 6 depicts the waiting time for both scheduled BAT and scheduled HSA. Waiting time for BAT is reduced by 13.34% and for HSA it is reduced by 60%. Waiting time for HSA is less than BAT, so HSA is outperforming BAT in reducing the waiting time and increasing the UC. There is exists a trade-off between electricity cost and UC. If we increase UC then electricity cost will also increase.

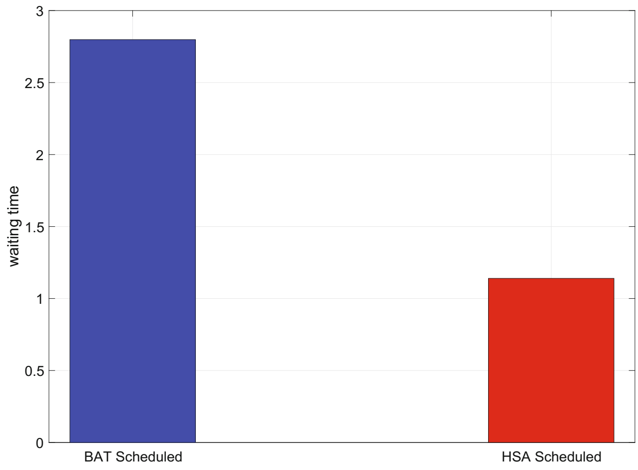


Fig. 6. Waiting time

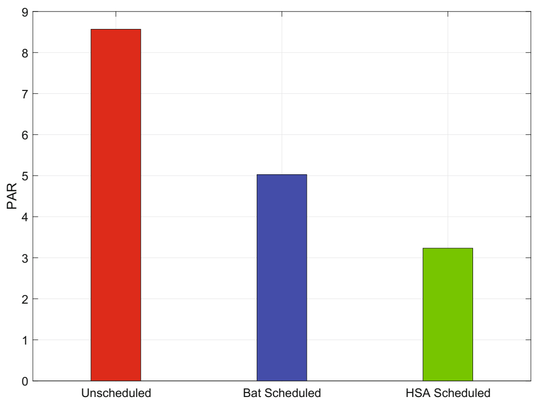


Fig. 7. PAR

Figure 7 represents the PAR for unscheduled, BAT scheduled and HSA scheduled. PAR is reduced for both HSA and BAT as compared to unscheduled PAR. PAR is reduced by 58.9% for HSA and 38.9% for BAT as compared to unscheduled. Clearly HSA is better in PAR reduction as compared to BAT. In DSM consumer is not the only person who gets benefit but utility also. Reducing the PAR helps in retaining the stability of the grid and improves efficiency.

5 Conclusion

In this paper, performance of HEMS is evaluated using two meta-heuristic techniques: BAT and HSA. Both are comparatively evaluated. Appliances are scheduled according to their LOT and energy consumption. Performance of two meta-heuristic techniques are evaluated in this paper on the basis of cost reduction, energy consumption, PAR and waiting time. There exist a trade-off between UC and electricity cost. Results show the overall effectiveness of HSA. Electricity cost is reduced due to shifting of load toward off-peak hours.

References

1. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016). http://ieeexplore.ieee.org/document/7120135/
2. Zhao, Z., Lee, W.C., Shin, Y., Song, K.-B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013). http://ieeexplore.ieee.org/document/6525433/
3. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in Smart Grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–5. IEEE, January 2012. http://ieeexplore.ieee.org/document/6175785/
4. Ma, X.G.K., Yao, T., Yang, J.: Residential Power Scheduling for Demand Response in Smart Grid. Elsevier (2015)
5. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in Smart Grid. Energies **10**(3), 319 (2017). http://www.mdpi.com/1996-1073/10/3/319
6. Iqbal, Z., Javaid, N., Khan, M.R., Ahmed, I., Khan, Z.A., Qasim, U.: Cost and load reduction using heuristic algorithms in Smart Grid. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 24–30. IEEE, March 2016. http://ieeexplore.ieee.org/document/7471167/
7. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016). http://www.sciencedirect.com/science/article/pii/S0378778816306867
8. Gupta, I., Anandini, G., Gupta, M.: An hour wise device scheduling approach for demand side management in Smart Grid using particle swarm optimization. In: 2016 National Power Systems Conference (NPSC), pp. 1–6. IEEE, December 2016. http://ieeexplore.ieee.org/document/7858965/
9. Diamantoulakis, P.D., Pappi, K.N., Kong, P.-Y., Karagiannidis, G.K.: Game theoretic approach to demand side management in Smart Grid with user-dependent acceptance prices. In: 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), pp. 1–5. IEEE, September 2016. http://ieeexplore.ieee.org/document/7881024/
10. Ogwumike, C., Short, M., Denai, M.: Near-optimal scheduling of residential smart home appliances using heuristic approach. In: 2015 IEEE International Conference on Industrial Technology (ICIT), pp. 3128–3133. IEEE, March 2015. http://ieeexplore.ieee.org/document/7125560/
11. Anvari-Moghaddam, A., Monsef, H., Rahimi-Kian, A.: Optimal smart home energy management considering energy saving, a comfortable lifestyle. In: 2016 IEEE Power, Energy Society General Meeting (PESGM), p. 1. IEEE, July 2016. http://ieeexplore.ieee.org/document/7741432/
12. Graditi, G., Di Silvestre, M.L., Gallea, R., Sanseverino, E.R.: Heuristic-based shiftable loads optimal management in smart micro-grids. IEEE Trans. Industr. Inform. **11**(1), 271–280 (2015). http://ieeexplore.ieee.org/document/6834796/ http://ieeexplore.ieee.org/document/6834796/

Energy Optimization in Home Energy Management System Using Artificial Fish Swarm Algorithm and Genetic Algorithm

Muhammad Talha, Muhammad Shahid Saeed, Ghulam Mohiuddin, Musa Ahmad, Muhammad Junaid Nazar, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. In this paper, we have evaluated the performance of heuristic algorithms: Genetic Algorithm (GA) and Artificial Fish Swarm Algorithm (AFSA) for Demand Side Management. Our prime focus in this paper, is to optimally schedule appliances in a smart home in such a way that the Peak to Average Ratio (PAR) and the electricity cost can be reduced. The pricing scheme used in this paper is real time pricing. Our Simulation results validate that the two nature inspired schemes successfully reduce PAR and electricity cost by transferring load of on peak hours to off peak hours. Our results also depict a trade off between electricity cost and comfort of a user.

1 Introduction

If we go just 30 years back in time and have a look at the world, we will be amazed to see how things were different before. There was no technology revolution. Internet was not available to the public. The number of skyscrapers in the big metropolitan cities, were very low. At that time the world did not face that much of electricity crisis as we do now a days. Now major construction has been done all over the world. Big industries, airports, seaports are being built to boost the economic infrastructure of countries. With this major infrastructure development and also with the increase in the number of consumers of electricity, countries are facing serious amount of electricity short fall. [13–15] shows that countries like Pakistan, India, and Bangladesh are still unsuccessful in generating enough power to fulfill their electricity needs. This problem of energy shortfall is being faced by many countries all over the world. Scientists all around the world are trying to come up with plans and new ways to generate electricity so that they can eradicate this dilemma of energy shortfall that keeps getting bigger and bigger every year due to the increase in demand of electricity. Electricity all over the world is being produced through solar, wind, hydro, coal and fossil fuel etc. However, with the exponential growth in the number of consumers, that

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_18

requirement of continuous supply of energy throughout the day is getting harder to meet. The energy generation projects are still falling behind in generating the electricity needed. To cope with issue, scientists are working on Demand Side Management (DSM). DSM is basically a phenomenon to efficiently balance the load of electricity in such a way that the consumer and utility both get benefitted. DSM is a part of Smart Grid (SG). SG is the future, as it holds the power to completely replace the traditional power infrastructure that majority of the countries over the world still use. With a bi-directional communication infrastructure, SG allows the consumer to communicate with the utility and tell the utility about the amount of energy which is required over a certain period of time. A utility can also conveys real time electricity rates to the consumer and tells them how they can reduce the cost of electricity they consume by scheduling their load (Fig. 1).

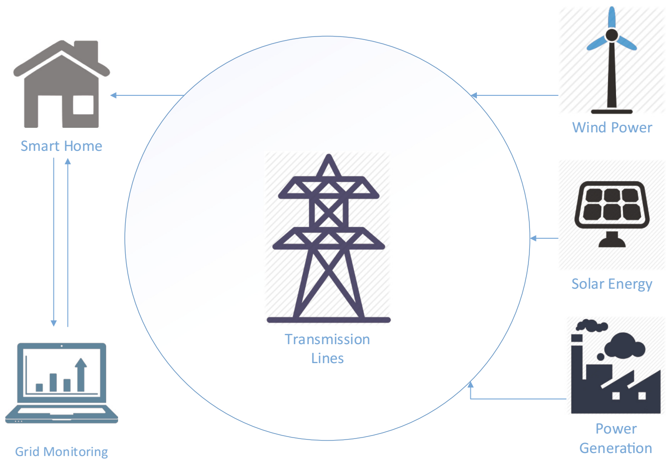


Fig. 1. Smart Grid

With DSM in place now the utility and consumers of electricity can work side by side in efficiently managing the load of electricity. With scheduling in place, the load at the shoulders of utility will be reduced and the consumer would have to bear a lesser cost.

2 Related Work

In [1], the authors used Integer Linear Programming to schedule load for a single and multiple homes. The focus was to schedule, power shift-able and time shift-able appliances. The simulations showed a moderate Peak to Average Ratio (PAR). They also succeeded in balancing the electricity load throughout the day. User Comfort (UC) was also taken into account while scheduling appliances.

The high electricity cost, due to scheduling appliances entirely on the UC was a limitation of [1]. Heuristic techniques like Genetic Algorithm (GA), Binary Particle Swarm Optimization, Wind Driven Optimization, Bacterial Foraging and Optimization Algorithm and the proposed Genetic Wind Driven Optimization were discussed in [2]. The authors addressed problems like cost minimization, PAR reduction, UC maximization and energy load reduction. Appliances were divided into two classes, the authors assigned priority bits based on user's preferences and those bits were used to schedule appliances. Solution proposed in [2], was successful in achieving reduced PAR and Electricity Cost (EC). A tradeoff was present between UC and EC. Numerous simulations were done for single and multiple homes, however, the integration of Renewable Energy Sources (RES) was not considered. No communication infrastructure was suggested either. In [3], the problem of load scheduling was solved using an optimal load scheduling algorithm. Three strategies were proposed: One that maximized UC, another one saved a lot of EC and the third one found a tradeoff between UC and EC. Day Ahead pricing was used in [3], which not only reduced overall cost but also reduced PAR. No automatic scheduler was used that could schedule appliances on its own in [3]. User had to do all the scheduling manually based on the day ahead pricing he got from the utility.

A smart scheduling scheme that ensured reduced PAR and cost was proposed in [5]. GA was used to schedule appliances. Instead of a stand alone pricing scheme, a hybrid pricing scheme of Inclined Block Rate (IBR) and Real Time Pricing (RTP) were used to calculate EC. Although authors in [5], achieved a reduced EC and PAR, however, they ignored UC while scheduling appliances. Five classes of appliances were proposed in [6] and the goal was to minimize cost and improve UC. The authors were presented with a Mixed Integer Linear Problem which they solved using Generalized Benders Decomposition Approach (GBDA). GDBA is basically about dividing a bigger problem into sub problems. The sub problems were then solved individually in order to get the desired result. Cost Reduction and UC maximization was achieved in [6]. However, the inconsideration of multiple homes and multiple pricing schemes limited the proposed solution. No efforts were made to balance load too. The problem of smooth electric power scheduling was addressed in [7]. At first the problem of power profile, which used to fluctuate a lot, was identified. This problem was addressed by using The Majorization Theory. After the power profile was smooth enough, it was passed on to the residential load where an algorithm called Distributed User Benefit Maximization Load Control Algorithm was used to schedule appliances in a home. Smooth power profile and efficient scheduling of load was done along with the reduction of PAR and generation cost. However, the cost of electricity consumption at the consumer end was ignored.

In [8], a scenario was discussed where a Shared Facility Controller (SFC) buys electricity from a grid or a Residential Unit (RU). The RU's generated energy using RES, and they sold extra electricity to the SFC. SFC bought energy at a lower cost as compared to the one it had to pay when electricity was bought from the grid. A scheme was proposed in order to reach a Stackelberg Equilibrium, by using a Non-Cooperative Stackelberg Game. This game helped SFC out in deciding that, from where should it buy electricity? The main objective discussed was the reduction of the cost of electricity bought. Mechanisms were introduced for efficient charging and discharging of storage devices in order to save cost. The UC was not considered. The problems of load management, reverse power flow and reduced EC were solved in [9]. The possibility of energy trade was discussed, as Consumers would sell electricity to his neighbors and earned money, if a local installation of RES happened. The problem of interaction of users and power generation was solved using a Game Theoretic Approach. To solve the trading and pricing problem GBDA was used. For scheduling multiple classes of appliances, Dynamic Programming was used. UC was ignored during scheduling of Appliances.

The problem of high energy cost, PAR and low UC were consider in [4]. Multiple techniques were discussed in this paper including: GA, Ant Colony Optimization and Multiple Knapsack Problem (MKP). A system that had RES and a communication system to ensure reliable communication between user and utility was proposed. The scheduling of appliances was done using the MKP. Simulations showed a reduced PAR and cost. The UC was maximized due to the proposed solution. The initial cost of RES installation was not considered.

In [10], the focus of authors was on energy load reduction, UC maximization and raising the Thermal Comfort Level (TCL). An optimal scheduling scheme was proposed which used the mixed integer non-linear programming. The electricity demand was fulfilled not only from the grid but also from the CHPs. CHP was composed of water tank, a backup boiler and Fuel Cell and was used to generate electricity that fulfilled the consumer's need. Extra energy were stored in batteries and Hybrid Electric Vehicles. Simulations showed the validity of proposed scheme as TCL and UC was maximized quite a bit. RTP was used as a pricing scheme. The limitations of [10], were the fact that they only considered a single home and only one pricing scheme in their experiments. In [11], the problem of Cost Efficiency (CE) was addressed. Different energy consumption patterns were proposed with only a single goal to reduce EC. CE was solved by using Fractional Programming. The load was scheduled in two ways; Day ahead load scheduling: used an algorithm based on Dinkelbach Method. For real time management, consumer had to manage his load all by himself. Higher CE was achieved with low service fee. The CE also increased when DERs were integrated. Only limitation of [11] was the fact that they did not consider energy load management.

In [12], the authors introduced a scheme that not only reduced load on grid but also integrated Distributed Energy Resources (DER). The main objective was to reduce EC and energy load. It was proposed that DERs should be installed to serve a community, not an individual. In this way the cost of installing a DER was divided into a community and the return on investment time also reduced a lot. RUs also produced electricity using CHPs and stored them in Vanadium Redox Batteries. For load scheduling and for addressing Demand Response (DR), Particle Swarm Optimization was used. The charging and discharging patterns were handled by Q-Learning algorithm. The bill calculation was done using the Bill Balancing Algorithm. Successful achievements were EC reduction and UC maximization. Load Balancing of the grid was not considered.

3 Problem Statement

DSM has been a big problem. As mentioned in related work section, a lot of researchers have tried to address the sub problems in a DSM. One of the problems of DSM is to optimally schedule appliances in a Smart Home (SH). Another sub problem of DSM is the reduction of PAR. Most of the work done by researchers in this area is to reduce the electricity cost bear by the user and the maximization of UC. The PAR however is a performance parameter that has been ignored by a lot of researchers in HEMS. In the proposed system model, we have successfully reduced PAR by using AFSA algorithm. We also managed to reduce the cost of electricity in our proposed system model.

4 System Model

To solve the problem of efficiently scheduling the appliances in a home. We propose a system that finds out a better scheduling pattern by using algorithms like Genetic Algorithm (GA) and Artificial Fish swarm algorithm (AFSA). Our goal is to reduce PAR and Maximize UC. Our Classification is inspired from [1]. We have considered a single home which has seven appliances. The appliances are classified into three categories: Non-Shift-able, Power Shift-able and Time Shift-able. The Non Shift-able appliances are our base load. The appliances in this category cannot be shifted to other time slots as it will seriously damage the UC. This category contains three appliances: Hob and Oven, Heater, Fridge and Freezer. The Power Shift-able category contains Water Boiler (WB) and Electric Vehicle (EV) (Fig. 2).

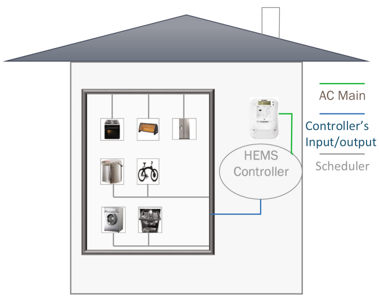


Fig. 2. Different appliances in a Smart Home

Algorithm 1. Artificial Fish Swarm Algorithm

Initialize Parameters: y, tryNum, stepSize, popS, visualScope, swarmLength, ChaseFitness, SwarmFitness, LastFitness.

```
for m = 1 → tryNum do
  for n = 1 → popS do
    if ChaseFitness(i) < SwarmFitness(i) then
      if ChaseFitness(i) < LastFitness(i) then
        LastFitness(i)= ChaseFitness(i) ;
      else
        PrayBehavior;
      end if
    end if
    if SwarmFitness(i) < LastFitness(i) then
      LastFitness(i)= SwarmFitness(i) ;
    else
      PrayBehavior;
    end if
  end for
end for
```

This category can be scheduled based on the power provided to it over a day. EV and WB need a limited amount of energy over the span of a day. So these appliances will be scheduled by giving them different amount of energy over the passage of time until their daily need is fulfilled. The Last category of Appliances is Time Shift-able, this category contains Washing Machine and Dish Washer. Both of these appliances can be scheduled any time in the span of a day. But once they start their operation you cannot pause it until they complete their Length of Operation Time (LOT). For a detailed LOT refer to Table 1.

In our proposed system we are considering a time slot of 24 h. The Pricing scheme that we are using to measure the cost of electricity over the span of 24 h is RTP. The On Peak Hours in our case will from 8am to 1pm. In these six hours, the price of electricity will be higher as compared to the price of

Table 1. Classification of appliances

Name	Type	Operational time (hour)	Power rating (kWh)
1. Hob and Oven	Non-shiftable	7pm–8pm	1 kWh
2. Heater	Non-shiftable	9pm–10pm, 3am–5am	1 kWh
3. Fridge and Freezer	Non-shiftable	24 h	0.12 kWh
4. Water boiler	Power-shiftable	-	3 kWh
5. Electric vehicle	Power-shiftable	-	5 kWh
6. Washing time	Time-shiftable	1 h	1.5 kWh
7. Dish washer	Time-shiftable	1 h	0.8 kWh

Algorithm 2. Genetic Algorithm

```
Input: set of appliances App or P;
Initialization: PH, OPH, t=0, H;
for t = 1 → T do
   for h = 1 → H do
      Generate feasible P randomly;
      for h = 1 → P do
         Fitness Calculation;
         Select best P, pop save in pop1;
         Status check of App using PH and OPH;
         if t == PH then
            wait for OPH;
            if Cosumption == high then
               Check remaining t of all App and check LOT until it is 0 ;
            end if
         end if
      end for
      Generate new pop;
      Perform crossover;
      Save it in pop2;
      Perform mutation;
      Select from pop2;
      Mutate;
      if solution == infeasible then
         Update solution;
         Update with sol in pop2;
      end if
      Update pop best solution;
      Update t=t+1 till 24 h;
      Terminate when t=24 h;
   end for
end for
```

electricity in other hours. Our proposed scheme will try to shift load from On Peak Hours to Off Peak Hours so that cost of energy can be reduced. We are going to schedule our appliances using GA and AFSA. Both of these are nature inspired algorithms. We are mapping the behaviors of both algorithms to Home Energy Management System (HEMS).

5 Simulation and Discussion

As you can see in Fig. 3, the PAR of unscheduled load is very high. The value for PAR in unscheduled load's case is 15.22. In comparison to unscheduled load's PAR, our optimization schemes work really well. The PAR for AFSA is 6.70 which is 56% less than unscheduled load's PAR. The value of PAR for GA is 7.97, this value is 48% less than unscheduled load's PAR. AFSA beats GA with

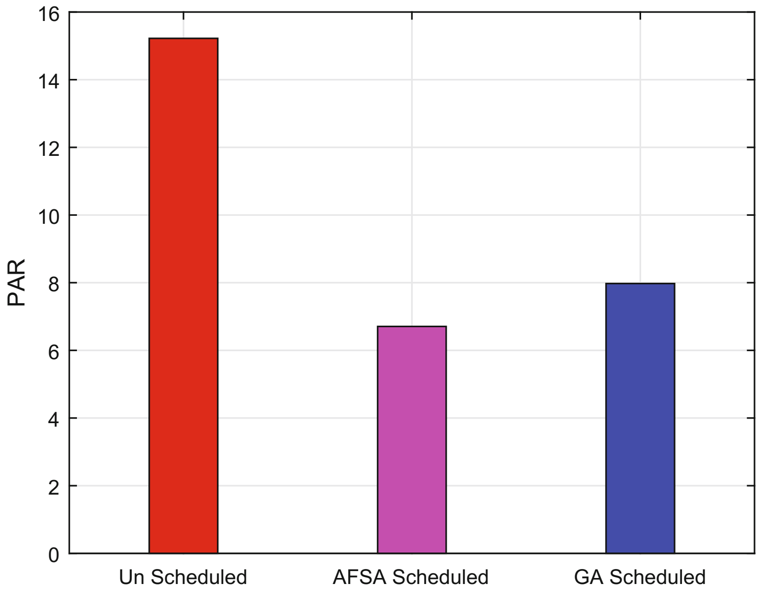


Fig. 3. Peak to Average Ratio

a lesser PAR value in our case. AFSA's PAR value is 10% less than GA's PAR. We achieve our goal of reducing PAR value successfully with the help of GA and AFSA. GA reduces the PAR value by shifting the load from On Peak Hours to Off Peak Hours. The reduction of PAR value means that the utility will have less electricity demand to meet. This will help out the customer, as with a reduced PAR value the cost of electricity charged by the utility will be lower as compared to cost in unscheduled load's case.

In Fig. 4, total cost charged by utility for unscheduled, AFSA scheduled and GA scheduled is shown. The utility charges 1027 cents, when there is no scheduling of appliances. This is quite a lot of money that the consumer has to pay to meet his electricity needs. The cost of electricity for GA is 810 cents. GA successfully managed to bring 21% decrease in the electricity cost that the consumer has to pay. AFSA out performed GA in this case. AFSA managed to achieve a far lesser cost of 720.7 cents. This cost is 30% less than cost of unscheduled electricity. AFSA beat GA's cost by 11%. Our Goal of lesser cost is also achieved here as you can clearly see that the result of GA and AFSA are more economical as compared to unscheduled load of electricity.

Our simulations also consider waiting time for AFSA and GA. Figure 5, shows that the waiting time for fixed load is lowest, as compared to waiting time of power shiftable and time shiftable appliances. The reason behind low value of waiting time for fixed load is the fact that this is base load and it can't be scheduled or interrupted. This electricity load has to work on its own without any interference. The second class of appliances that we are considering is power

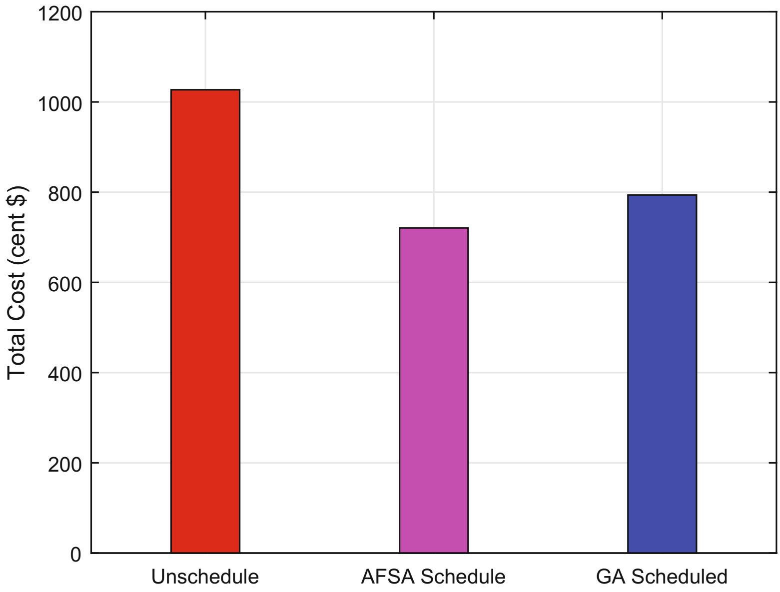


Fig. 4. Total cost of electricity

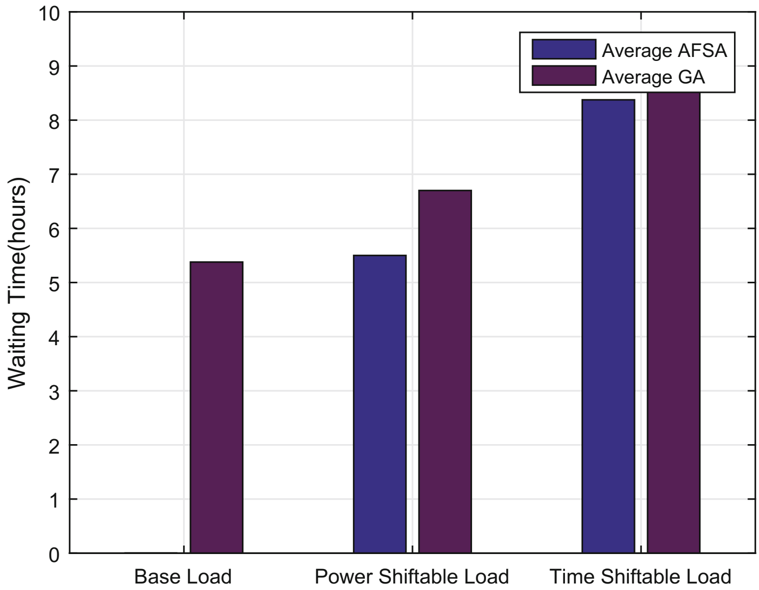


Fig. 5. Waiting time

shiftable appliances. The appliances in this class require continuous supply of electricity. The scheduler can decide the amount of electricity it can provide to the appliances in this class at a certain period of time. The appliances in this class can be scheduled over time but they can't be interrupted once their operation time starts, though the amount of energy supplied to them can vary over time. The waiting time for AFSA is 5.5 h for power shiftable load. AFSA's waiting time is 20% less than GA. Waiting time for GA is 6.9 h. Another class of appliances that we are considering in our simulations is time shiftable load. The appliances in this class can be scheduled overtime and there operation can be interrupted. As Fig. 5 shows that, the waiting time for AFSA is 8.3 h which is 10% less than GA's value of 9.2 h. These values show that the consumer has to wait a lot in order to reduce bills. There is a tradeoff between cost, PAR and waiting time. If cost and PAR are low the waiting time will always be high.

Figure 6 shows the load of electricity bear by the utility throughout the day. Unscheduled load shows high demand of electricity in peak hours. As compared to that, AFSA and GA manages to get a more normalized load graph with not only lesser peaks, but they also manage to get a load with lesser fluctuations. A smooth supply of energy is an achievement of both AFSA and GA. Highest value of load throughout the day for GA is 2.03 kWh. This value is 33% less than unscheduled load of 3.03 kWh. The maximum load value of AFSA is 2.62 kWh which is 13% less than unscheduled load. GA performed 22% better in the reduction of load value as compared to AFSA.

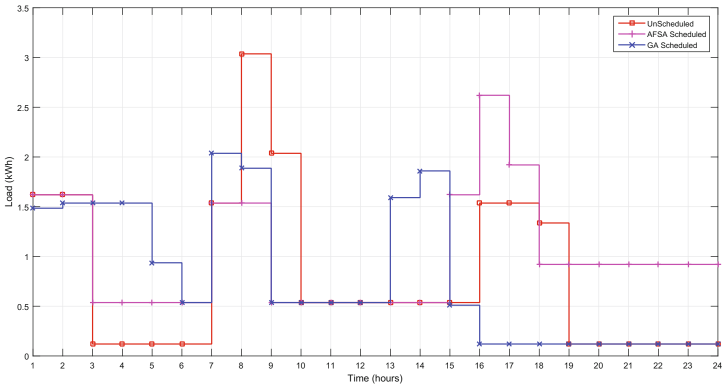


Fig. 6. Load

6 Conclusion

In this paper we evaluate the performance of GA and AFSA for HEMS. Our goal of PAR and cost minimization is successfully achieved. GA managed to reduce the total cost of electricity through out a day. We used RTP as a pricing scheme. The total cost for AFSA was 11% less than GA's cost and 30% less than unscheduled's cost. GA's cost was 21% less than unscheduled load's case. The PAR value is 56% and 48% less than unscheduled load for AFSA and GA respectively. With AFSA and GA, we also managed to find a balance in the load of electricity supplied throughout the day. In future, we will propose a hybrid scheme based on GA and AFSA for the same problem.

References

1. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES, pp. 1–5. IEEE (2012)
2. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
3. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
4. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
5. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
6. Roh, H.-T., Lee, J.-W.: Residential demand response scheduling with multiclass appliances in the smart grid. IEEE Trans. Smart Grid **7**(1), 94–104 (2016)
7. Huang, Y., Mao, S., Nelms, R.M.: Smooth scheduling for electricity distribution in the smart grid. IEEE Trans. Smart Grid
8. Tushar, W., Chai, B., Yuen, C., Smith, D.B., Wood, K.L., Yang, Z., Poor, H.V.: Three-party energy management with distributed energy resources in smart grid. IEEE Trans. Ind. Electron. **62**(4), 2487–2498 (2015)
9. Samadi, P., Wong, V.W., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
10. Anvari-Moghaddam, A., Monsef, H., Rahimi-Kian, A.: Optimal smart home energy management considering energy saving and a comfortable lifestyle. IEEE Trans. Smart Grid **6**(1), 324–332 (2015)
11. Ma, J., Chen, H.H., Song, L., Li, Y.: Residential load scheduling in smart grid: a cost efficiency 483 perspective. IEEE Trans. Smart Grid **7**(2), 771–784 (2016)
12. Jiang, B., Fei, Y.: Smart home in smart microgrid: a cost-effective energy ecosystem with intelligent hierarchical agents. IEEE Trans. Smart Grid **6**(1), 3–13 (2015)
13. Wikipedia: Energy Shortfall in India. https://en.wikipedia.org/wiki/Electricity_sector_in_India. Accessed May 2017
14. Wikipedia: Energy Shortfall in Pakistan. https://en.wikipedia.org/wiki/Electricity_sector_in_Pakistan. Accessed May 2017
15. Wikipedia: Energy Shortfall in Pakistan. https://en.wikipedia.org/wiki/Electricity_sector_in_Bangladesh. Accessed May 2017

An Efficient Scheduling Using Meta Heuristic Algorithms for Home Demand-side Management in Smart Grid

Adnan Ishaq, Nasir Ayub, Arje Saba, Asad Ghafar, Basit Amin, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. Energy consumption demand is comparatively higher than available energy, new approaches are being discovered to fulfill energy demand. This problem can be solved by assimilating Demand Side Management (DSM) with Smart Grid (SG). In this work, we observe the working of Home Energy Management System (HEMS) by using three meta-heuristic techniques; Harmony Search Algorithm (HSA) and Firefly Algorithm (FA) and Bacterial Foraging Algorithm (BFA). Time Of Use (TOU) is used as a pricing signal for calculation of electricity bill. The main concern of this paper is to minimize cost, reduce Peak to Average Ratio (PAR), maximization of user comfort and load management. Load management can be done by shifting load from on-peak hours to off-peak hours. Simulation results show that implemented techniques successfully achieve the defined goals.

Keywords: Smart Grid · Demand Side Management · Demand Response · Heuristic techniques · Harmony Search Algorithm · Firefly Algorithm · Bacterial Foraging Algorithm

1 Introduction

Scientific revolutions have blessed human life in many aspects, electricity is one of those blessings. Increasing population increase electricity demand as mentioned in [1]. Traditional grids are incapable of providing sufficient electricity to fulfill consumption demands. In this regard, SG are used which offers communication in traditional grids. SG contain smart appliances, smart meter (SM), etc. SM interchange information between consumer and utility. DSM is one of the most essential feature of SG for energy optimization. DSM proficiently manage energy request by sharing real time information among consumer and utility. It offers equilibrium between demand and supply. By adopting DSM approaches user are

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_19

fortified to shift their load from on-peak hours to off-peak hours [1]. DSM is responsible for Demand Response (DR) and load management. DR is response of consumer against pricing rates set by the utility. Consumer set his priorities according to announced rates [3]. DR is helpful for both utility and consumer. From utility point of view, reduction of PAR is achieved while it also educate consumer to consume maximum energy during off peak hours that results in reduction of electricity bill (Fig. 1).

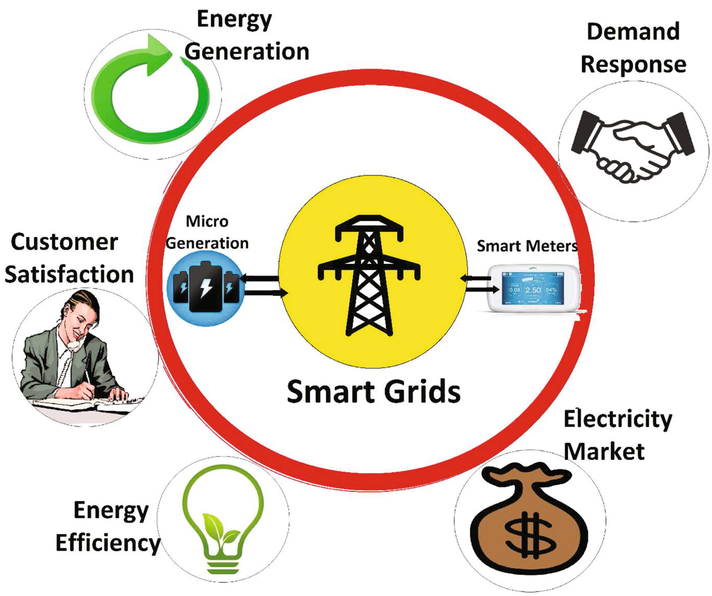


Fig. 1. Smart Grid

HEM is an important application of SG. HEM assure cost minimization and pinnacle load reduction. Moving load from on peak hours to off peak hours may diminish electric cost and reduce peak load, it might decrease the user comfort. Different techniques have been used to encourage the consumer to participate in DR program to proficiently use energy consumption. Some of them are critical peak pricing (CPP), time of use (TOU), real time price (RTP), inclined block rate (IBR) and day ahead pricing (DAP). These pricing schemes motivate the consumer to shift high load appliances to off peak hours this result in reduction of both electricity cost and PAR.

Many techniques have been adopted to achieve defined goals. In [1,2] integer linear programming, approximate dynamic programming are used to minimize cost as well as energy consumption, however they are capable of handling large number of appliances. So that's why different meta-heuristic techniques are used for energy management for example, in [7,8] author used genetic algorithm (GA) to minimize electric cost, binary particle swarm optimization (BPSO) also reduce electric cost and user comfort is achieved in [9,11]. Our central energy consumption source is domestic area that's why researchers show interest to schedule household appliances.

In this work, we consider domestic area. We use two meta-heuristic techniques HSA and FA. Through purpose work, we have an optimized pattern for all available appliances. We categorize appliances into two types interruptible appliances and non-interruptible appliances on the account of their usage. We have used TOU as pricing signals in our work. The above mentioned optimization techniques schedule the operation of interruptible and non-interruptible appliances. The purpose schedule make it promising to plan the operation of various appliances independently. Independent scheduling considerably reduce the complication of scheduling algorithm.

Simulation result shows that our implemented techniques shrink the energy expenses of consumer as compared to situation where scheduling is not applied. Section 2 comprises of recent related work. Section 3 explain problem statement. Purposed model is defined in Sect. 4. Simulation and results are in Sect. 5. Concluding remarks followed by future work are in Sect. 6.

2 Related Work

In paper [1], author propose a scheduling mechanism for load management. The authors use Integer Linear Programming (ILP) for load management. The main objective of the presented model is to decrease price and PAR. Presented system is capable of balancing load for both interruptible appliances and non-Interruptible appliances. By scheduling appliances, appliances have to wait for their time of operation that effect the user comfort. In paper [10], authors elaborate the energy management mechanism. Adaptive Dynamic Programming (ADP) is used to solve the scheduling problem. Residential SG network is used in this paper. Author divides the appliances into two parts. 1st is delay sensitive and 2nd is delay tolerant. Objectives of this paper are cost minimization, PAR reduction and user comfort maximization. The achievement of this technique is to improve the delay time of appliances. The author gave high priority to the delay sensitive appliance. Results show that author achieve the goal. Beside this, parameter consideration is considered as flaw.

In paper [3], author propose a scheduling mechanism for house hold appliances. Main goals of this work are to reduce electricity cost and PAR efficiently. GA is used for scheduling appliances in this paper. The author use combination of pricing schemes (RTP+IBR) in this work. By adopting this combined pricing model the proposed power scheduling method would effectively reduce both the electricity cost and PAR. In this way, the system becomes more stable. Although this is a good solution for home energy management. The main flaw in this scheme is that the user comfort is compromised as there is inverse relation in cost and comfort. To make system more efficient, those demands are considered which have high impact rather than normal impact (Table 1).

Table 1. Related work table

References	Techniques/Targeted areas	Objectives	Achievements	Limitations
[1]	Integer Linear Programming (ILP)	Minimize peak hour load balanced daily load	Efficiently reduces the peak hour load	User comfort neglected/waiting time increased
[2]	Approximate dynamic programming	Manage load and power exchange	Reduces the energy expenses utilization of RERs	Ignored energy management, RES integration cost
[3]	EMC architecture according to DR RTP, IBR	Dropping electricity cost and PAR	Efficiently manage the peak hour load and reduce the electricity bill	User comfort neglected RES integration are not used
[4]	GA, BPSO, BFOA, WDO, GWD, hybrid GWD RTP single home, multiple homes	Maximizing UC and minimizing electricity cost and PAR	Energy consumption Total cost, PAR, UC	Compromising on RES, Inconsideration of Parameters Tuning, 4 h Delay
[5]	DR, day-ahead price including integer and continuous variables	Reduce power consumption reduce payments	Wanted exchange off between the installments and the inconvenience	RES integration, Ignored PAR management
[6]	BPSO, GA, ACO multiple knapsack problem (MKP) IBR TOU	Lessening electricity bill, PAR reduction and UC level maximization	Minimize electricity bill, minimize waiting time, optimal integration of RESs	Neglected power consumption management, RES integration cost.
[7]	GA	Electricity cost and peak formation minimization	Effectively reduce electricity bills while preserving user comfort	Trade Trade off user Comfort and energy consumption.
[8]	BPSO	User frustration and electricity bill reduction	Encourage consumer participation and enhance EMS	Integration of RESs are not considered
[9]	Residential energy load	Increase UC, decrease charges and PAR	Effectively reduce the power consumption to minimize cost and PAR	Explicit pressure values degrade performance
[10]	Residential SG networks	Decrease cost and interruption minimization	Reduced the operation delay	Inconsideration of Parameters Tuning
[11]	Nikaido-Isoda function based relaxation algorithm. (ii) Newton method toward centralized coordination. Appliances	Reduce the discomfort of customer and reduction of PAR. Also addresses the scheduling problem of the residential smart	DSM game is able to achieve a superliner convergence rate that is much faster than the traditional. Peak to average ration is minimized	Author reduces the PAR but regarding user comfort electricity cost is not given importance.
[12]	Kp+GA+RTP+IBR	Utility electricity cost minimization and peak formation reduction	Effective model for both single and multiple users	User comfort level and integration of RES are neglected while achieving desired objectives

In paper [2], the author suggests an energy mechanism for DSM. To manage appliances author used a technique known as Approximate Dynamic Programming (ADP) and a game theoretic approach was embraced to show the connection of the consumers with overabundance control era. Using this technique user can store and sell electricity. By using this, user is enable to utilize RES and also capable of selling excess energy. It is also helpful to minimize the cost of electricity. Beside this, author do not focus on the energy management. Author suggest the integration of Renewable Energy Source (RES) however cost of implementation is ignored. To manage the load and maintain user comfort, author proposed a scheme in paper [9]. The algorithm proposed in this paper is WDO with the integration of RES to gain the steadiness. The main concern of author is to maximize the user comfort, minimize the electricity cost. TOU is used as a pricing scheme in this paper. Instead of providing benefit to only consumers, the proposed scheme also facilitates service provider by matching waiting time of appliances and load. The simulation results show that the proposed work is better than existing work. Beside this, explicit values may degrade the performance.

In article [5], the main focus of author is on the scheduling problem of load in demand side. Author categorized appliances into two type flexible appliances and fixed appliances. DR based technique is used in this paper for scheduling purpose. The pricing scheme for this is Day Ahead Pricing. The main objectives are minimizing power consumption of the appliances and then to reduce the cost. Author achieves all these goals efficiently. The tradeoff between the cost and user comfort is minimized. Beside this, author have compromise on some facts, PAR is ignored in this scheme, as more load is shifted on peak hours. Another deficiency is the cost of RES is overlooked. In paper [8], author suggests the integration of RES. Consumer also have to focus on the bills in addition to the utility. There is a difference between the optimal load scheduling and the scheduling according to user ease. The technique used in this paper is Bacterial Particle Swarm Optimization (BPSO) with the integration of RES. The main focus of this work is to minimize the bill of electricity and user frustration. It also helps to reduce the PAR. Author used the TOU pricing scheme in this paper. Main achievement of this work is that it can effectively reduce the energy consumption. Author achieves all this in good way. Beside this, author neglect the integration cost of RES.

In article [4], author proposed an Energy Management Controller (EMC) algorithm for residential areas in smart grid system. In this author used five heuristic algorithms GA, BPSO, Bacterial Foraging Optimization Algorithm (BFOA), Wind Driven Optimization (WDO) and Genetic Wind Driven (GWD). The main purpose of this scheme is to schedule the load between on peak hours and off peak hours. The pricing scheme used in this paper is RTP. Author successfully maximize the user comfort along with the reduction of electricity bill and PAR. This scheme is proposed for both the single home and for multiple homes. Author implements this technique efficiently. Beside this, author also compromise on some things. The waiting time is increased in this scheme. In paper [6], author have used multiple knapsack. The main focus is to minimize the cost and peak to average ratio and also increase the user comfort level. Author used the IBR with addition of TOU for calculation of electricity price. Simulation result shows that the cost is reduced as well as the waiting time is also maximized and appliances are used in scheduled pattern. The applied algorithm has ability to work in any condition which make it more effective.

Results also show that the proposed scheme has better results than existing techniques. It reduced cost 10 percent more than the other techniques. By giving priorities to the appliances, the waiting time is reduce to minimal level. Beside all this the author neglect to manage the power. Additionally author did not consider implementation cost of RES.

In paper [7], author used GA for scheduling house hold appliances. He implements the scheme on the bases of consumer preference. Dynamic pricing scheme is used in this paper for calculation of electricity bill. The main focus is to reduce the cost and PAR. As users want the electricity bill to be as lower as possible. User can schedule appliances according to his usage and can change scheduling pattern. Author achieves all these goals efficiently. Cost is reduced and PAR is also minimized. User comfort is also increased. Beside this, the only one fact the author do not focus is that as user comfort increase, there is increase in the energy consumption. In paper [12], author use GA for scheduling house hold appliances with combination of RTP and IBR as a pricing scheme for electricity bill calculation. The main objectives of this work is a quick implementation of scheduling algorithm. Along with this, author also focus on reduction of waiting time for appliances, reduce electricity cost, and minimize PAR. The proposed model CHEMA consists of six layers. The proposed model can be useful for both unit home and multiple homes. The simulation results show that author achieve what he aimed. This work could be better if author consider integration of RES. In paper [11], author addresses well-organized depletion of thin load flowing in DSM and management problem of utilizations. Objective of work is to minimize PAR and minimize user discomfort. Firstly, procedures are applied to re-model the DSM for improving the scarcity to minimize the disruptions then develops a bi-directional framework for attaining the Nash equilibrium in DSM. Non-smooth Newton approach is used to accelerate the convergence. At last, a fast gradient method is designed to quickly solve consumer's response problem to improve the performance of SM. For the reduction of PAR, authors use Nash equilibrium. DSM game model consists of these dispersed optimization procedures: (i) Function Based Easing Procedure (ii) Newton Method Toward Centralized Coordination. The authors also tend to achieve user comfort by using the sparse patterns, meanwhile cost of electricity is neglected in achieving user comfort as there is tradeoff in user comfort and cost.

3 Problem Statement

In SG, load management and reduction of electricity bill are main concerns as there is irregularity in energy consumption pattern. Generally user comfort is neglected while reducing electricity bill. In DSM we have to schedule appliances in such a way that it reduces total electricity bill and maximize user comfort. There are following optimization problems that need to be addressed in SG.

- Reduce total power consumption
- Minimize electricity bill

- Reduce PAR
- Maximize user comfort level
- Provide an optimize schedule of appliances

Many techniques have been adopted in [5–8] to tackle the said issues, however there is tradeoff between user comfort level and electricity bill. To solve these problems heuristic optimization techniques are used. These techniques have been able to deal with homes having large number of appliances with less computational time and less computational complexity. In our purposed work we have used three techniques HSA, FA and BFA to achieve our goals and compare their results.

4 System Model

Electricity price is increasing day by day so we need a system to minimize average cost of electricity. HEMS enables the consumer to regulate and observe residential appliances. HEMS facilitate the user in term of minimization of electricity wastage as well as reduction of electricity bill. The appliances we schedule are classified into interruptible appliances and non-Interruptible appliances. Interruptible appliances are those which can be disturbed during their time of operation while Non-Interruptible appliances are those which cannot be interrupted during their execution time. Appliances with their categorization is shown in Table 2. User can communicate with utility, this communication inform consumer about his recent consumption. In this work we use TOU as a pricing scheme, through which load is shifted from peak hours to Off peak hours. It enable us to know how much energy you are using and how much budget you

Table 2. Classification of appliances

Interruptible	Non-interruptible
Cloth dryer	Lighting
Refrigerator	Pool pump
AC	Electirc vehicle
Dishwasher	Heater
Vacuum cleaner	
TV	
Electric stove	
PC	
Water heater	
Ironing appliance	
Hairdryer	
Other	

are paying per hour. When we know our consumption one can reduce the usage that results in the reduction of bill. In our work we have used three optimization techniques HSA and FA and BFA (Fig. 2).

Power rating of interruptible appliances and non-interruptible appliances is shown in Tables 3 and 4. We have used TOU as pricing scheme for calculation of bill. The main objectives of all this study is to minimize the consumption of energy in order to reduce the electricity cost and reduction of PAR. The main

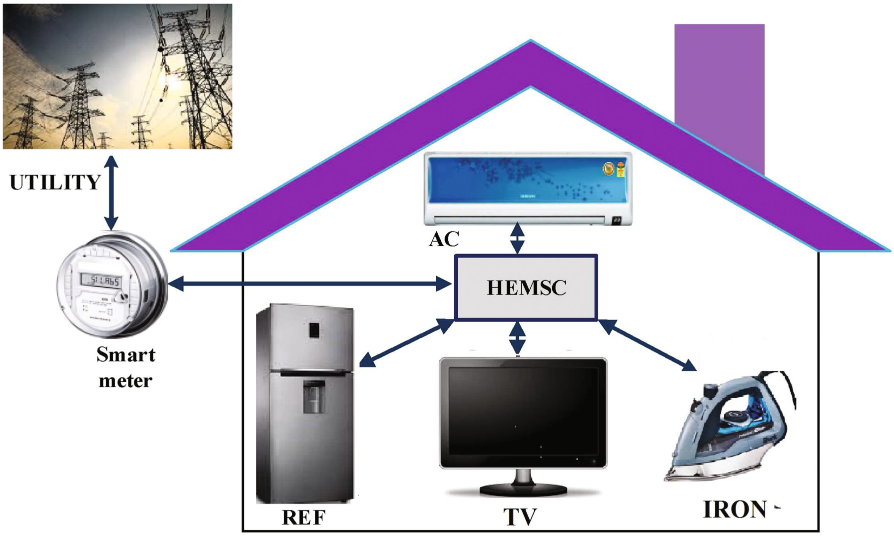


Fig. 2. DSM

Table 3. Interruptible appliances with power rating

Interruptible appliances	Power rating (kW)
Cloth dryer	0.5
Refrigerator	0.125
AC	1.00
Dishwasher	1.00
Vacuum cleaner	1.00
TV	0.25
Electric stove	1.00
PC	0.25
Water heater	1.5
Ironing appliance	1.00
Hairdryer	1.00
Other	1.5

Table 4. Non-Interruptible appliances with power rating

Non-Interruptible appliances	Power rating (kWh)
Lighting	0.6
Pool pump	1.00
Electric vehicle	2.50

focus is on the minimization of total cost calculated according to the Eq. 1 with PAR reduction.

$$Cost = \sum_{hour=1}^{24} (E_{Rate}^{Hour} \times E_{Rate}^{App}) \tag{1}$$

Equation 2 is Calculating the load and calculation of PAR using the Eq. 3 as in [5]

$$Load = P_{Rate}^{App} \times App \tag{2}$$

$$PAR = max(load^s)/Avg(load^s) \tag{3}$$

4.1 Optimization Techniques

4.2 HSA

HSA was presented by Geem et al. in 2000. HSA is used to schedule house hold appliances. Primary strides includes: initialize imperatives (harmony memory size (HMS), harmony memory consideration rate (HMCR), pitch Adjustment rate (PAr), ending criteria). Then Initialize the HM (instate the HM population contain HMS vector produced haphazardly) and then Improvise the New Harmony (memory consideration, pitch adjustment, random selection) then Update HM. HSA can deal with alteration that is the reason HSA is exceptionally successful. HSA works productively in given scenario. HSA is very simple to execute. In HSA different consonant gatherings can be utilized parallel.

4.3 FA

FA was presented by Xin-She Yang in 2007. Fundamental steps include in FA are (1) introduce objective function (2) Generate population (3) Determine the light intensity (4) Calculate attractiveness (5) Movement of less brighter firefly towards brighter firefly (6) Update the light power rank of the firefly and discover current best. FA can manage exceedingly non-direct multimodal enhancement issues normally and proficiently. There is no utilization of velocities so there is no issues of velocity. The speed of merging of FA is high in probability of finding optimize solutions. It has the adaptably of mix with other optimization techniques.

4.4 BFA

BFA was presented by Kevin Passino in 2002 The algorithm was aimed for application of optimization problem fields. Loops in the algorithm can constitute plenty of ways to draw different search behavior. Generally we have large number of chemotaxis iterations, and small numbers of the other iterations. The step size is usually a small fraction of the search space. In reproduction, normally half the population with a low health metric rejected, and two copies of each member from the first (high-health) half of the population are taken. The probability of elimination and dispersal is generally relatively huge.

Algorithm 1. FA for SG scheduling

```
Begin
Objective function
Create first population of fireflies
Express light intensity I
Define absorption coefficient Y
while t ¡ MaxGen do
   for p = 1 : n do
      for q = 1 : n do
         if  I(q)¿I(p) and I(q)¿I(p) then
            attractiveness change with distance
            move firefly i towards j
            Assess new solutions and update light intensity
         end if
      end for
   end for
   Rank fireflies and find the current best
end while
results and visualization
End
```

5 Simulations and Results

In this section, execution of the proposed architecture is assessed and results are demonstrated that speak to how much the proposed approach is viable. Load consumption in FA, HSA, BFA and unscheduled is shown in Fig. 3. Main attention in this simulation was cost reduction, PAR reduction and load balancing on DSM for TOU signal. The maximum load consumption value in FA is 9 kWh, in HSA its value is 8.2 kWh and in BFA its value is 8.3 kWh while in unscheduled its maximum value is 11 kWh. By using meta heuristic techniques load is balanced between off-peak hours and on-peak hours as shown figure. Result shows that implemented techniques handle the load better as compared to unscheduled scenario. In Unscheduled load, load is increased during peak hours that increased the electric cost and energy consumption while in implemented schemes load

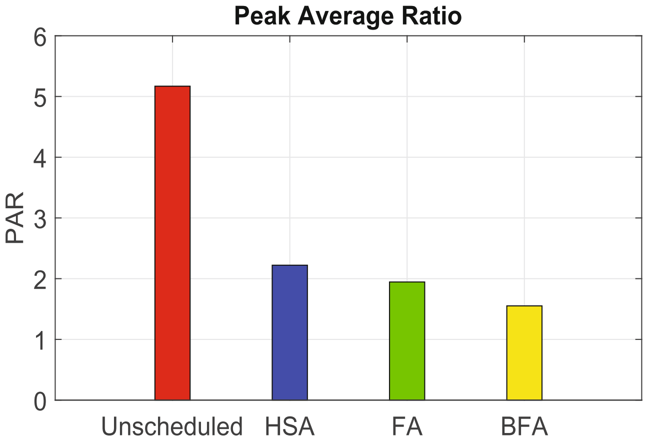


Fig. 3. Peak to average ratio

is managed by shifting the load in to off-peak hours that reduce the energy consumption of specific hour. In Unscheduled, power consumption is increase due to which electricity cost also increased that results in peak curve formation but in implemented schemes peak curve is not formed so power consumption is reduced during peak hour by using TOU cost minimization is achieved. In Schedule load, load is schedule in such a manner that loads shifted from on-peak hours to off-peak hours. From simulations it is verified that our designed model achieve significant results.

PAR is increased in unscheduled load because the max load is shifted to a specific hour of a day and creates the peak. So PAR is increase due to unscheduled load. In FA, when load is schedule it reduces the PAR by shifting the load to off-peak hours. PAR in our scheme FA is minimum. As shown in Fig. 4, the PAR is significantly reduced as compared to HSA and greatly reduced as compared to unscheduled. PAR and cost will affect each other. In FA PARs value is 2 while in unscheduled its value is 5.2 and in HSA its value is 2.3. It shows that PAR is significantly reduced in FA as compared to unscheduled case and HSA to avoid peak formation in any hour of a day. Figure clearly confirmed that peak load for the duration of the high price signal when scheduled load is low as compared to the unscheduled. This clearly indicate the effectiveness of implemented scheme in which load is shifted without effecting overall load. This will reduce cost, decrease PAR as well as perform load shifting.

Electricity cost per hour for implemented schemes is shown in Fig. 5. Results illustrate that cost paid in scheduled load is low as compared to unscheduled load because load is shifted from on-peak hours to off-peak hours. Figure 6 displays the variation of total cost between unscheduled and scheduled schemes. It is conformed from the figure that FA attains minimum cost as compared to HSA and BFA. HSA attains maximum cost among all scheduling schemes. However, cost is reduced in all implemented scheduling approaches as compared to

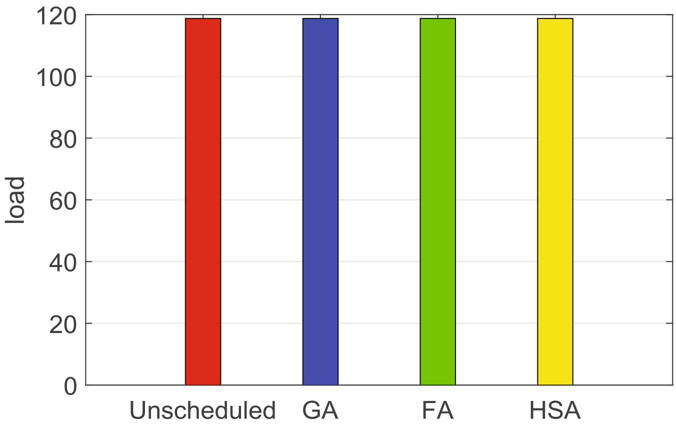


Fig. 4. Total load

unscheduled situation. Electric cost is increased in unscheduled because price is increases in emergency situation according to price signal TOU. Figure 5 shows the cost for each hour through the day, here the price signal by HSA in peak hours is highest when contrasted with other two techniques. This influences the general cost every day as presented in Fig. 6 which demonstrates that price using FA is low as compared to other implemented techniques. The implemented approaches reduces 15 % cost when contrasted with unscheduled cost.

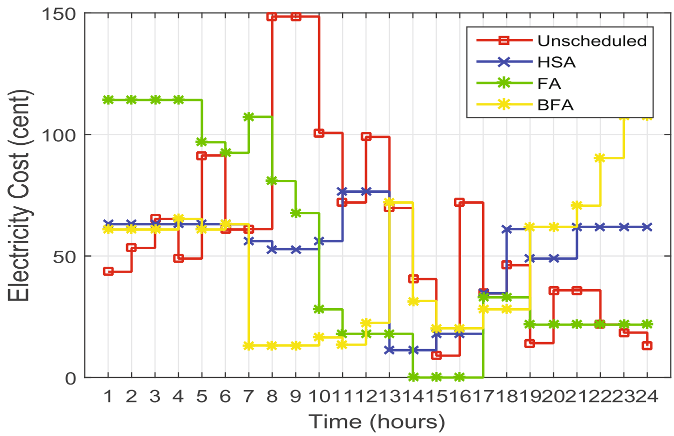


Fig. 5. Cost

While reducing the cost and PAR, waiting time cannot be neglected which is shown in Fig. 7. User comfort is figured as waiting time. Waiting time is the time

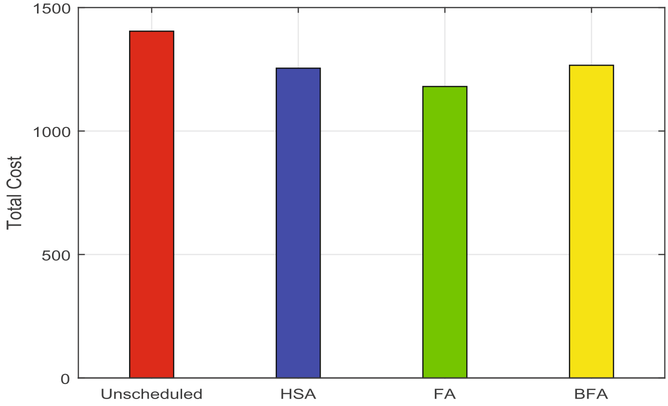


Fig. 6. Total cost

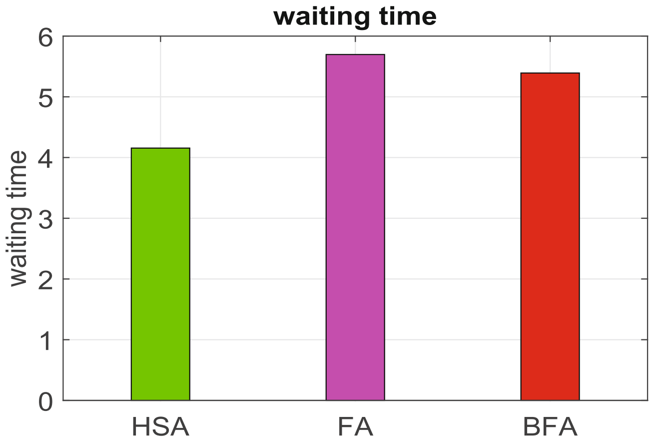


Fig. 7. User comfort

consumer have to hold appliance to gets its turn to work. User comfort and electric cost has inverse relation. To decrease electricity bill, low price hour must be favored by consumer. While, if comfort is favored at that point they should trade off on cost (i.e. they should pay high cost). Figure 6 denotes total cost of all implemented approaches. Where Fig. 7 represent waiting time of consumer for every single scheme. The implemented approaches reduce the waiting time of the appliances to maximize the user comfort. Waiting time is increased because we reduce the PAR by shift the load from on-peak hours to off-peak hours. User comfort also sacrifice when we reduce the electricity cost because of their inverse relation. Figure 7 demonstrates that waiting time for all three schemes from which it is concluded that HSA have lowest value as compared to FA and BFA.

6 Conclusion and Future Work

DSM is exceptionally viable with regards to load management. Generally consumers shift load on RES however it is insufficient to completely facilitate the user. To handle such circumstance AI based optimization approaches are adopted. In this paper, we have implemented FA, HSA and BFA and examined its working by implementing and comparing it with unscheduled load. The firefly algorithm is very effectual. Simulation results offer finest schedule plan of appliances. Our purpose technique works better than existing algorithms in term of cost, competence and attainment rate. Our implemented techniques are stronger in resolving DSM. It is possible to further improve the quality by decreasing randomness. FA can be implemented to solve multi objective problems. In addition, FA chained with other techniques could be electrifying work in future.

References

1. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home Demand-Side Management in Smart Grid. In: Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES, pp. 1–5. IEEE (2012)
2. Samadi, P., Wong, V.W., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
3. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for Demand Response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
4. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for Demand Side Management in Smart Grid. Energies **10**(3), 319 (2017)
5. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for Demand Response in Smart Grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
6. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
7. Rasheed, M.B., Javaid, N., Awais, M., Khan, Z.A., Qasim, U., Alrajeh, N., Iqbal, Z., Javaid, Q.: Real time information based energy management using customer preferences and dynamic pricing in smart homes. Energies **9**(7), 542 (2016)
8. Mahmood, D., Javaid, N., Alrajeh, N., Khan, Z.A., Qasim, U., Ahmed, I., Ilahi, M.: Realistic scheduling mechanism for smart homes. Energies **9**(3), 202 (2016)
9. Rasheed, M.B., Javaid, N., Ahmad, A., Khan, Z.A., Qasim, U., Alrajeh, N.: An efficient power scheduling scheme for residential load management in smart homes. Appl. Sci. **5**, 1134–1163 (2015)
10. Liu, Y., Yuen, C., Yu, R., Zhang, Y., Xie, S.: Queuing-based energy consumption management for heterogeneous residential demands in Smart Grid. IEEE Trans. Smart Grid **7**, 1650–1659 (2016)
11. Li, C., Yu, X., Yu, W., Chen, G., Wang, J.: Efficient computation for sparse load shifting in Demand Side Management. IEEE Trans. Smart Grid **8**(1), 250–261 (2017). doi:10.1109/TSG.2016.2521377
12. Khan, M.A., Javaid, N., Mahmood, A., Khan, Z.A., Alrajeh, N.: A generic demand side management model for Smart Grid. Int. J. Energy Res. **39**, 954–964 (2015)

A New Two-Stage Algorithm for Constructing Group Multicast Tree with Bandwidth Constraint

Meng Sun[1(✉)], Xinchang Zhang[1], and Jianwei Zhang[2]

[1] Shandong Provincial Key Laboratory of Computer Networks, Jinan, China
{sunm, zhangxc}@sdas.org
[2] Shandong Computer Science Center
(National Supercomputer Center in Jinan), Jinan, China
zhangjw@sdas.org

Abstract. The traditional idea of constructing an optimal group multicast tree (GMT) mainly adopted two-stage algorithm. The first stage constructed multiple single-source multicast trees (SMT), named the initial GMT, the second stage modified the initial GMT to obtain an optimal GMT. In this paper, we propose a new two-stage algorithm. Our algorithm adapt the second stage algorithm in two items. Firstly, for the purpose of quickly reducing the number of saturate links, we propose a new method of selecting the need-modified SMT. Next, aiming at minimizing the need-modified SMT's cost and the number of the saturate links, we propose a dynamic modification algorithm using the existing topology of the SMT. Following that we combine the two steps into an entire iteration process. Our algorithm can significantly decrease the iteration number and can greatly reduce the computational time. Experimental results demonstrate that our algorithm has better performance than the previous two-stage algorithm.

1 Introduction

The group multicast is a communication mechanism from multiple source nodes to multiple destination nodes, its routing problem usually models as constructing group multicast tree (GMT). The GMT is composed of multiple single-source multicast tree (SMT). These SMTs need to share network resources (i.e., nodes and links), that may cause congestion on some links and reduce network utilization. Therefore, the constructing an optimal GMT includes two important sub-problem. The one is how to construct a SMT, the other is that how to handle the congestion of all the SMTs on some links.

There have been some research about the group multicast routing problem in IP network until now. The method of group multicast routing can be split into two categories. The one method is using the Core-based Tree (CBT) protocol [1] to construct a core based tree. Nevertheless, using CBT will result in side effects and too large delay variation.

Another method is to construct a source specific routing tree for each member. It has been proved that finding optimal solution for the group multicast routing with

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_20

bandwidth reservation is NP complete [9]. In this paper, we consider this idea. The existing research are as follows.

Paper [2] present a coordinated strategy called overhead comparison to generate a set of multicast trees. The basic idea was that when occurring saturate links, the author give up the tree whose alternative overhead is least. By this way, the set of trees coordinated with each other can make the overall cost minimum. Paper [3] proposed a new routing algorithm called Feasible Solutions using adapted Takahashi and Matsuyama algorithm [10] for the group multicast routing. The author mainly focused on how to achieve the maximum success rate of building a set of multicast trees and proved that their algorithm always finds out at least one feasible solution, but it may produce arbitrarily high cost for the resulting trees. Papers [4–6] all used adapted genetic algorithm to resolve the group multicast routing problem under the corresponding constraints, respectively.

Papers [7–9] all adopt an iteration modification algorithm, their idea can be concluded that they all adopted two-stage algorithm. The first stage constructed multiple SMTs, the second stage iteratively modified some SMTs until the GMT doesn't contain saturate link. In the second stage, they all select the busiest link and the SMT that occupied the busiest link to modify. Each paper's method had difference when modification SMT, some focused on making use of the existing SMT to modify the SMT's topology, some adopt to rebuild a new tree. However, in these algorithm, the SMT that needed to be modified are all random, the new tree construction algorithm also may need higher time cost, the GMT's total cost are also not be considered.

Therefore, on the basis of analyzing the algorithms in papers [7–9], we propose a new two-stage algorithm to construct an optimal GMT. Our algorithm focus on handling the following problem: how to implement the iteration modification can reduce the iteration number, how to design a new method of selecting SMT can quickly reduce the number of saturate links, how to modify a SMT can minimizing the its cost and its number of saturate links.

The rest of this paper is organized as follows. In Sect. 2, we will introduce the group multicast routing problem formulation and analysis the previous work. Section 3 give the first stage algorithm, i.e., the initial GMT construction algorithm based on TM algorithm. In Sect. 4, in the second stage, we will propose a new iteration modification algorithm for the initial GMT, including a new method for selecting the need-modified SMT, the dynamic modification for a need-modified SMT. Section 5 will perform the experiment and compare our algorithm with the existing algorithm.

2 Problem Formulation and Analysis

2.1 Problem Formulation

Assume a network be modeled as a weighted Graph $G = (V, E)$, where V is the set of nodes and E is the set of links. Let $T = \{T_1, T_2, \cdots, T_m\}$ represent the group multicast, T_i is a multicast from one source node to many destination nodes, the source node of each T_i is different. Let $d(T_i)$ represent the total cost of T_i. The problem can be described as follows.

$$\min \sum_{i=1}^{m} d(T_i) \quad (1)$$

Subject to, for any link $l \in T$,

$$\sum_{s=1}^{k} B_s x_l^s \leq M_l \quad (2)$$

The optimization goal (1) ensures that the total cost of the GMT is minimum. Equation (2) ensures that the total bandwidth required by all the multicast request on link (i,j) should be less than the available bandwidth of the link (i,j). Where $x_l^s = 1$ represents $l \in T_s'$, otherwise $x_l^s = 0$. B_i is the bandwidth requirement of T_i. M_l is the total bandwidth of link l.

2.2 Previous Work Analysis

Papers [7–9] all adopt the two-stage algorithm. The first stage constructed multiple SMTs, the second stage iteratively modified some SMTs until the GMT doesn't contain saturate link. Their implementation can be introduced as follows.

In paper [7], Low and Wang proposed an algorithm that is based on adaptation of an algorithm by Takahashi and Matsuyama (TM) to construct the SMTs. Next in the second stage, while existing saturate link, select a SMT randomly that occupied the busiest link, rebuild a new SMT using TM algorithm, update the saturate link set. Repeat the process until the GMT doesn't contain saturate link.

In paper [8], Low and Song constructed the minimum delay SMTs using Dijkstra algorithm. And then in the second stage, while existing saturate link, select a SMT randomly that occupied the busiest link, modify the SMT by deleting a link, finding a new path that satisfied bandwidth constraint not cost constraint and connecting the two sub-tree, update the saturate link set. Repeat the process until the GMT doesn't contain saturate link. Call this algorithm as L.S. algorithm. The author demonstrated that their algorithm can always find out at least one solution that is feasible.

In paper [9], Yan-lin proposed two routing algorithms for group multicast with delay and bandwidth constraints. Firstly, construct the SMTs using a new metric that described the links' characteristics. Secondly, while existing saturate link, select a SMT that occupied the busiest link, the SMT had the smallest bandwidth requirement, rebuild a new SMT using the first stage algorithm, update the saturate link set. Repeat the process until the GMT doesn't contain saturate link.

The above three algorithms mostly adopted the idea. In the first stage construct some SMT for each source node. In the second stage, selecting a busiest link and a multicast tree that occupies the busiest link, iteratively modify the old multicast tree or construct a new multicast tree, constantly update the saturate link set and the need-modified multicast tree until the GMT doesn't contain the saturate link.

Analyzing the above three algorithm, we can see that construction an optimal GMT with bandwidth-constraint includes the following key problem. Firstly, for an initial GMT, if existing some saturated links, which saturate link we should choose to migrate multicast flow. The second is which multicast tree we should choose to modify its routing path. The third is that how to obtain a proper multicast tree, construct a new multicast tree or modify the old multicast tree. Therefore, we consider the following adaptation to increase the efficiency and success rate of building a GMT.

- The above algorithm selected the SMT all only consider the busiest link, this way may generate the result that the saturate link decreased very slowly. So, we can adapt the selection method of the SMT that needed to be modified.
- For the SMT that needed to be modified, paper [7, 9] reconstruct a new SMT. The two algorithms need to use the static algorithm to calculate minimum cost SMT. Recalculating a SMT has low computational efficiency compared with modify the old SMT using the existing topology. Paper [8] modified the old SMT by deleting the busiest link. Only deleting one link can result that the SMT still have saturate link. So, we can adapt the modification method by deleting the SMT's all saturate links to achieve the minimum cost and minimum saturate link.
- For the iteration modification process, with updating the saturate link set, if the number of saturate links decreases very fast, the iteration can be completed very fast. So, we can adapt the iteration process by the adaption 1 and 2.

On the basis of the above adaption, we propose a new two-stage algorithm to construct an optimal GMT. Different from the previous work, our solution mainly adapts the second stage in two-fold. Firstly, with the goal of quickly reduce the number of saturate links, we design a new method to select the need-modified SMTs. Secondly, aiming at minimizing the need-modified SMT's cost and minimizing the number of saturate links of the need-modified SMT, we propose a dynamic modification algorithm based on TM algorithm to modify a SMT. Thus, the computation space and time of each iteration can be improved, and the total iteration number can greatly be reduced compared with the existing algorithm.

3 The Initial GMT Construction Algorithm

In the first stage, we construct a source specific routing tree for each member, called all the SMTs as the initial GMT. SMT construction, i.e. multicast routing, is closely related to the Steiner tree problem [10]. A shortest path heuristic algorithm proposed by Takahashi and Matsuyama [11] is a simple heuristic method proven good performance bound. It is based on a greedy strategy. The detail implementation can be described as follows.

Given a network model $G = (V, E)$, where V is the set of nodes and E is the set of links. Suppose exist a single-source multicast, s represents its root node, $D = \{d_1, d_2, \ldots, d_m\}$ represents its destination nodes.

Step1: Define two sets of node, $V = \{s\}, D = \{d_1, d_2, \cdots, d_m\}$.
Step2: Select a node from D, denoted as d_i. Finding the shortest path using Dijkstra algorithm with bandwidth constraint from the node to all the nodes in V, called the set of path as P. Choose the minimum cost path in P, named as p_i. Add all the node of p_i into V. $D = D - \{d_i\}$.
Step3: If $D \neq \emptyset$, return step2, else stop.

We construct each SMT under the available bandwidth using Dijkstra algorithm. Each SMT constructed by the above three steps all satisfies the optimal goal (1). All the SMTs comprise the initial GMT, denoted as T. Following that we will calculate whether existing saturate links in the initial GMT. To simplify the calculation, the following data structure is employed.

A link and tree belonging list, for each link, to record which SMTs the link is belong to, denoted as $CorLT(i,j)$. Assume a link (i,j), the link is belong to k SMTs. The link and tree belonging list can be noted as $CorLT(i,j) = \{(i,j), T_1, T_2, \cdots, T_k\}$.

For all the links of the initial GMT, we can build its link and tree belonging list, and then can calculate its residual bandwidth. Given a link (i,j) and its list $CorLT(i,j) = \{(i,j), T_1, T_2, \cdots, T_k\}$, let $total\_band_{(i,j)}$ represents its total available bandwidth, req_band_r represents the requirement bandwidth of T_r. So its residual bandwidth can be calculates by Eq. (3).

$$res\_band_{(i,j)} = total\_band_{(i,j)} - \sum_{r=1}^{k} req\_band_r \tag{3}$$

After calculating the residual bandwidth of each link, we can distinguish the initial GMT may occur the following result.

- The residual bandwidth of all the links in the initial GMT are more than 0. T is the final solution.
- If the residual bandwidth of some links in the initial SMT are less than 0, i.e., some links are saturate links, we need to modify the SMTs that share these saturate links until the GMT doesn't contain saturate link.
- One or more SMT can't be generated. The reason is that the network can't supply sufficient bandwidth.

4 The Initial GMT Modification Algorithm

In the second stage, the previous work mostly adopt the idea, selecting a busiest link and a multicast tree that occupies the busiest link, iteratively modify the old multicast tree or construct a new multicast tree, constantly update the saturate link set and the need-modified multicast tree until the GMT doesn't contain the saturate link.

However, we need to consider the following problem. Firstly, for an initial GMT, if existing saturate links, which saturate links should be chosen to migrate out multicast flow. The second is which multicast tree should be chosen to modify its routing path. The third is that how to obtain a proper multicast tree, construct a new multicast tree or modify the old multicast tree. In this paper, we adapt the second stage algorithm in two-fold. The specific implementation is as follows.

4.1 The New Need-Modified SMT Selection Algorithm

Traditionally, this work all select a busiest link and a multicast tree that occupies the busiest link to modify. The goal of modifying the initial GMT is that reduce the number of saturate links until it is equal to 0. If only consider the busiest link, this way can result in the number of saturate links decreases very slowly.

So, we try to design a new method to improve the efficiency, i.e., quickly reduce the number of saturate links. If we select the SMT, whose saturate links' number is maximum, named need-modified SMT, next to modify the SMT, the number of saturate link can decrease faster than the previous method. The algorithm of selecting need-modified SMT can be described as follows.

Algorithm1. the algorithm of selecting a need-modified SMT

Input: the network graph G including the cost and bandwidth of each link, the bandwidth request of each multicast , the initial GMT

Output: the need-modified SMT and its saturate links

Step1:For all the links of the initial GMT, calculate its corresponding list between the link and its belonging trees, all the lists can compose a matrix, denoted the element of the matrix as $CorLT(i,j) = \{(i,j), T_1, T_2, \ldots, T_k\}$.

Step2:According to the matrix in step1, calculate the residual bandwidth of each link by equation (3), if the residual bandwidth is less than 0, Add the link into the saturate link set.

Step3:For each link by step2, according its list in step1, statistic the number of each SMT, choose the SMT that the number of saturate links is maximum.

4.2 Dynamic Modification Algorithm for the Need-Modified SMT

Given a need-modified SMT T_i, whether constructing a new tree or modifying the old tree must make the new tree satisfies $\min d(T_i)$. The previous work mostly used the static algorithm to construct a new tree, which need completely to recalculate the shortest routing and the minimum cost multicast tree. This lead to low computational efficiency and require long time to obtain the GMT.

In this paper, we adopt the dynamic modification algorithm for the need-modified SMT. Assume the new SMT T_i', the optimal goal (1) is equivalent to the following optimal goal

$$\min\left|d(T_i') - d(T_i)\right| \tag{4}$$

Subject to

$$n = 0 \tag{5}$$

where Eq. (4) represents the new multicast tree still have the minimal cost, and the new multicast tree still satisfy the optimization goal in Eqs. (1) and (2). n represents the number of saturate links of the new SMT T_i', it is equal to 0.

For the need-modified SMT, let s represent its root node, $D = \{d_1, d_2, \cdots, d_m\}$ represent its destination nodes. $L = \{l_1, l_2, \cdots, l_k\} = \{(s_1, e_1), (s_2, e_2), \cdots, (s_k, e_k)\}$ represents its saturate link set. For example in Fig. 1.

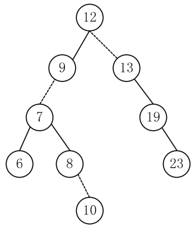


Fig. 1. An instance of a need-modified SMT that contains three saturate links.

As shown in Fig. 1, the need-modified SMT contains three saturate links $12 \rightarrow 13$, $9 \rightarrow 7$ and $8 \rightarrow 10$. If we only remove one saturate link, obviously the efficiency is not high. Thus, we removes all the saturate links and can obtain the following conclusions.

Conclusion 1: Removing k saturate links in a need-modified SMT can generate $k + 1$ sub-trees. These sub-trees may contain many nodes or contain only one node. We split the sub-trees into two-categories. The one is the sub-tree that contains the root node, called root sub-tree. The other is the sub-trees that doesn't contain the root node, called non-root sub-tree.

As shown in Fig. 1, removing three saturate links $12 \rightarrow 13$, $9 \rightarrow 7$ and $8 \rightarrow 10$ can generate four sub-trees, as Fig. 2a, b, c, d.

Fig. 2. Four sub-trees after removing three saturate links.

After removing all the saturate links, we need to consider that how to attach all the non-root sub-trees to the root sub-tree and can satisfy the Eqs. (4) and (5). As shown in Fig. 2, we need attach the sub-trees in Fig. 2(b), (c), (d) to the sub-tree in Fig. 2(a). The idea contains two key problem. Firstly, which sequence should we adopt to attach all the non-root sub-trees to the root sub-tree. Secondly, how should we connect the selecting non-root sub-tree and the root sub-tree.

As shown in Fig. 2, we compare the following two case to demonstrate our idea. The first case is we find the shortest unsaturated links for node pairs $(12, 13)$, $(9, 7)$ and $(8, 10)$. In this way, the new SMT isn't influence by the attached sequence of the sub-trees in Fig. 2(b), (c), (d). The second case is we let $V = \{12, 9\}$, $D = \{7, 10, 13\}$ and choose the attached sub-tree according to the path cost of node pairs

$(12, 7)$, $(12, 10)$, $(12, 13)$, $(9, 7)$, $(9, 10)$, $(9, 13)$. We select the sub-tree whose root node is the destination node of the minimum cost path. And then, we update the root sub-tree and the number of the non-root sub-trees decreases 1. Because the paths obtained in first case all consist of the paths obtained in second case, it is more possible to obtain a shorter path in the second case than the first case, i.e., the second case can bring a shorter cost SMT. Before describing our algorithm, we give the following definition and conclusion.

Definition 1: (the connected-path between two sub-trees) Given a root sub-tree $T_i^1 = \{v_i^1, v_i^2, \ldots, v_i^r\}$ and a non-root sub-tree $T_j^1 = \{v_j^1, v_j^2, \ldots, v_j^q\}$, there exists $r \times q$ source and destination node pair and $r \times q$ paths. The minimum cost path of two sub-trees is equal to the minimum path in the $r \times q$ paths. All path can be found by Dijkstra algorithm with bandwidth constraint.

Conclusion 2: Given a need-modified SMT, it contains k saturate links. We let $V = \{s, v_1, v_2, \cdots, v_r\}$ is the source node set, where s is the root node of the SMT, $v_i (1 \le i \le r)$ is all the nodes contained in the root sub-tree. let $D = \left\{d_1, d_2, \cdots, d_q\right\}$ is the destination set, where $d_i (1 \le i \le q)$ is the root node of all the non-root sub-trees. In this way, the element number of the destination set D is equal to k.

Conclusion 3: If we remove k saturate links in a need-modified SMT, the number of the connected-path is equal to k, denoted as $P = \{p_1, p_2, \ldots, p_k\}$. The new SMT can be constructed by $T_i' = P \cup T_i^1 \cup T_i^2 \cup \cdots \cup T_i^{k+1}$.

Based on the above analysis, we can know that the idea similar to the second case is dynamic process, our idea updates the initial root sub-tree and the non-root sub-trees. Thus, we propose a new dynamic algorithm to attach all the non-root sub-trees to the root sub-tree. The detail can be described as follows.

Algorithm2. Dynamic modification algorithm for the need-modified SMT
Input: a need-updated SMT and its saturate links
Output: a new SMT
Step 1:Remove all the saturate links L from the tree,
Step2:Construct the initial source node set and destination node set V and D according to Conclusion 2,
Step 3:Choose the attached sub-tree and find the connected-path, denoted as $\{p_1, p_2, \ldots, p_k\}$

3-1 Find the connected path from V to D according to definition 1, denoted the path as p_i.

3-2Add all nodes contained in the path p_i to V.

3-3 Remove the destination node of p_i from D.

3-4 If $D \neq \varnothing$, return to step 3-1. Else, stop.

Step4: Connect all the sub-trees in step1 with $\{p_1, p_2, \ldots, p_k\}$ to obtain a new tree.

4.3 The Iterative Modification Algorithm for the Initial GMT

The initial GMT constructed by the algorithm in Sect. 3 may contain many saturate links, so we propose an iterative modification algorithm to handle all the saturate links until the GMT doesn't contain saturate link. And the iterative process must satisfy the Eqs. (1), (2), (4) and (5), the algorithm can be described as follows.

Algorithm3. the iterative modification algorithm for the initial GMT
Input: the bandwidth of each link, the cost of each link, the initial GMT
Output: a new GMT that satisfies equations(1)(2) and doesn't contain saturate link
Apply step1 and 2 of algorithm1 to obtain the saturate link set L of the initial GMT.
while $L \neq \emptyset$ do
  Apply algorithm1 to obtain a need-modified SMT and its delete links.
  Compute the residual bandwidth after deleting the saturate links.
  Use algorithm2 to modify the need-modified SMT.
  Update the GMT and obtain a new GMT.
  Compute the residual bandwidth and the saturate link set L of the new GMT.
End

In the above iterative process, the step 2-1 guarantees the selected need-modified SMT contains the maximum number of saturate links. The step 2-2 guarantees the new SMT after modification have the minimal cost and it doesn't contain saturate link. So, the iterative process satisfies the Eqs. (1), (2), (4), (5) and can obtain a new optimal GMT.

4.4 Time Complexity Analysis

Assume a network $G = (V, E)$, the number of the edges is equal to e, the number of the nodes is n. The group multicast contains m single-source multicast. Using TM algorithm [11], each SMT can be built in $O(mn^2)$. The initial GMT can be build in $O(m^2n^2)$. In Algorithm3, the saturate link set in step1 can be found in $O(e)$. The need-modified SMT and its delete links in step 2-1 can be found in $O(e)$. Computing the residual bandwidth can be finished in $O(e)$. The modification of SMT in step2-2 can be accomplished in $O(mn^2)$. Updating the SMT can be achieved in $O(e)$. Hence, each iteration of the while loop in step2 takes $O(mn^2)$. In worst case, step2 iterates m-1 times. So, the time complexity of our algorithm is $O(m^2n^2)$. But, generally, in our algorithm, the iteration of step2 is less than the previous work, the specific experiment result is demonstrated in Sect. 5.

5 Experimental Results

We use the following items to evaluate the performance of our algorithm.

- Because all the algorithms of constructing GMT must satisfy the optimal goal (1) and (2), for our algorithm, whether it can provide a smaller cost GMT than the previous algorithm.

- In the iteration process of our algorithm, whether the descending speed of the saturate links' number is faster than the previous algorithm.
- Because our algorithm modifies the old SMT in the second stage not to rebuild a new tree, whether the computation time is lower than the previous algorithm.

The algorithm in paper [9] had been proved that had advantages over the algorithm in paper [8]. So, we compare our algorithm with the algorithm in paper [9] in the above three items. We use the method proposed by Waxman [12] to generate our simulation network. We generate a random bi-directional network containing 26 nodes and 65 edges. The total bandwidth of all links has a uniform distribution from 100 Mb to 1000 Mb. The distance of each node pair is defined their Euclidean distance. In our experiment, group members are selected randomly from the set of nodes.

Assume the size of group multicast is respectively equal to 10 and 15. Under the two cases, we analysis the total cost and the number of the saturate links with the iteration index increase. When the size of group multicast is equal to 10, the total cost of the GMT vs. the iteration index is shown in Fig. 3, the number of saturate links vs. the iteration index is shown in Fig. 4. When the size of group multicast is equal to 15, the total cost of the GMT vs. the iteration index is shown in Fig. 5, the number of saturate links vs. the iteration index is shown in Fig. 6.

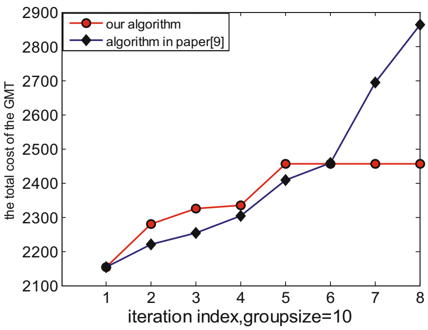


Fig. 3. The total cost vs. the iteration index, when the group size is equal to 10.

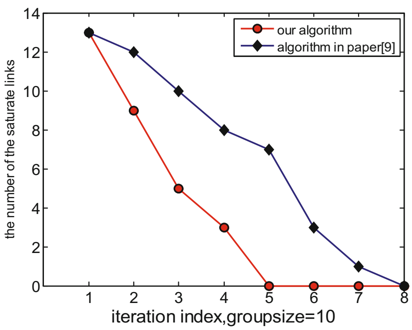


Fig. 4. The number of saturate links vs. the iteration index, when the group size is equal to 10.

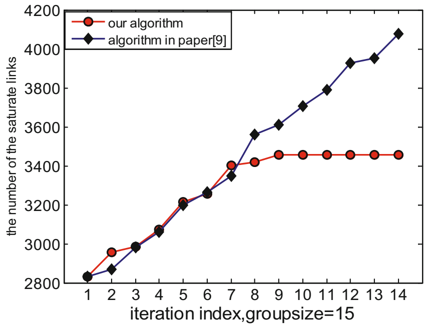


Fig. 5. The total cost vs. the iteration index, when the group size is equal to 15.

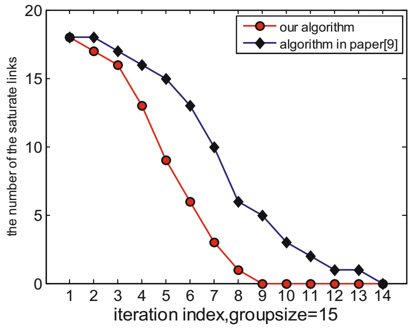


Fig. 6. The number of saturate links vs. the iteration index, when the group size is equal to 15.

From Figs. 3 and 5, we can see that when the group size is fixed, in our algorithm, the final total cost of the GMT is smaller than the algorithm in paper [9], and our algorithm find the optimal solution faster than the algorithm in paper [9]. This advantage also can be proven in Figs. 4 and 6, with the raise of the iteration index, the number of the saturate links decreases faster than the algorithm in paper [9], i.e., our algorithm need less iteration than the algorithm in paper [9].

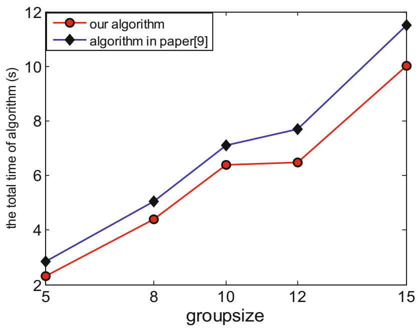


Fig. 7. The computation time of the two algorithms vs. the group size.

Moreover, we implement an experiment to compare the computation time between our algorithm and the algorithm in paper [9]. We let the two algorithms have the same network topology and the same group multicast. The result has shown in Fig. 7. We can see that for each group size, our algorithms' computation time is always less than the algorithm in paper [9].

6 Conclusion

This paper have studied a new two-stage algorithm for constructing group multicast tree with bandwidth constraint. Through analyzing the previous two-stage algorithm, we adapt the second stage algorithm. Our adaption mainly included two items, firstly, we proposed a new method to choose the need-modified multicast tree, which can quickly reduce the number of saturate links. Secondly, we proposed a dynamic modification algorithm, which can provide a smaller cost and a lower computation time. The two items greatly improve the performance of our algorithm. The simulation results have shown the superior performance compared with the previous algorithm. Furthermore, we will consider another constraint, such as, delay, to extend our algorithm.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant No. 61472230, and the Shandong provincial Natural Science Foundation of China under Grant Nos. ZR2014YL042, ZR2015FL025 and ZR2016YL008.

References

1. Billhartz, T., Cain, J.B.: Performance and resource cost comparisons for the CBT and PIM multicast routing protocols. IEEE J. Sel. Areas Commun. **15**(3), 304–315 (1997)
2. Jia, X.H., Wang, L.S.H.: A group multicast routing algorithm by using multiple minimum Steiner trees. Comput. Commun. **20**, 750–758 (1997)
3. Wang, N., Low, C.P.: On finding feasible solutions to group multicast routing problem. IEICE Trans. Commun. **E85-B**(1), 268–277 (2002)
4. Din, D.-R.: Genetic algorithms for multiple multicast on WDM ring network. Comput. Commun. **27**, 840–856 (2004)
5. Sanna Randaccio, L., Atzori, L.: Group multicast routing problem: a genetic algorithms based approach. Comput. Netw. **51**(14), 3989–4004 (2007)
6. Xuan, M., Qing, L.: Particle swarm optimization for multiple multicast routing problem. J. Comput. Res. Dev. **50**(2), 260–268 (2007)
7. Low, C.P., Wang, N.: An efficient algorithm for group multicast routing with bandwidth reservation. Comput. Commun. **23**, 1740–1746 (2000)
8. Low, C.P., Song, X.Y.: On finding feasible solutions for the delay constrained group multicast routing problem. IEEE Trans. Comput. **51**(5), 581–588 (2002)
9. Yan-lin, W.: Based QoS constrained group multicast routing for multimedia communication. In: International Conference on Computer, Mechatronics, Control and Electronics Engineering, pp. 296–299 (2010)

10. Kou, L., Markowsky, G., Berman, L.: A fast algorithm for Steiner trees. Acta Informatica **15**(2), 141–145 (1981)
11. Takahashi, H., Matsuyama, A.: An approximate solution for the Steiner problems in graphs. Math. Japonica **6**, 573–577 (1980)
12. Waxman, B.M.: Routing of multipoint connections. IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (1988)

Research on the Connotation and Measurement of Uncertainty for Reassembly Dimensions Based on Entropy

Conghu Liu[1,2](✉) and Kang He[1]

[1] Suzhou University, Suzhou 234000, China
lch339@126.com
[2] Sino-US Global Logistics Institute,
Shanghai Jiao Tong University, Shanghai 200030, China

Abstract. The uncertainty of reassembly dimensions is directly related to the accuracy and stability of remanufactured products. Therefore, the purpose of this paper is to improve the mathematical understanding of the uncertainty of remanufacturing. First, the uncertainty connotation of reassembly dimensions is analyzed by contrasting the ideal parts, the manufactured parts and the remanufactured parts combined with the uncertainty principle. Second, the uncertainty measurement model for reassembly dimensions is constructed, which can realize quantitative measurement by entropy. Then, we study the uncertainty coupling mechanism of reassembly dimensions, and its corollary has to be in conformity with the reality. Finally, the uncertainty of remanufactured heavy-duty engine crankshaft is measured to verify the validity of the model. These conclusions are useful in the development of the uncertain optimization decision for remanufacturing, and it provides theoretical support for lean remanufacturing.

Keywords: Remanufacturing · Assembly dimensions · Uncertainty · Entropy

1 Introduction

In 21 Century, with the rapid development of the world economy, the industrial waste is a substantial increase, the energy consumption is huge, and the ecological environment is deteriorating. Remanufacturing has been an effective way to address the resource crisis and environmental pollution problems. It is a key technology to build circular economy, and of great significance for the construction of a resource-saving and environment-friendly society [1, 2].

Due to the short time of birth, there are still some basic technical problems that have not been solved in remanufacturing engineering, which is the multi discipline. Remanufacturing uncertainty theory is one of the key issues. The research of remanufacturing uncertainty theory is not only the demand for remanufacturing production management, but also the key to ensure the quality of remanufactured products is not lower than that of manufactured products. Based on the related literatures of remanufacturing, we discover that the "uncertainty" has the highest frequency. There are multiple uncertainties, which are mainly treated as background or constraints in the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_21

remanufacturing process [3, 4]. Some experts and scholars researched on remanufacturing uncertainty, and their details are listed as follows:

Product demand uncertainty directions: Qiang et al. established convergence of the proposed algorithm that can allow for the discussion of the effects of competition, distribution channel investment, yield and conversion rates, combined with uncertainties in demand, on equilibrium quantity transactions and prices [5]. Amin and Zhang proposed a three-stage model including evaluation, network configuration, and selection and order allocation, which is supposed that demand is an uncertain parameter [6]. Inderfurth's paper shows that uncertainty in returns and demands can be a considerable obstacle to follow a consequently environmental-benign recovery strategy within a reverse logistics system [7]; and so on.

Remanufacturing planning uncertainty directions: Kenné et al. dealed with the production planning and control of a single product involving combined manufacturing and remanufacturing operations within a closed-loop reverse logistics network with machines subject to random failures and repairs [8]. Su and Sha presented a remanufacturing production planning method based on mixed-uncertainty and evidence theory, considering stochastic and fuzzy parameters in the process [9]; and so on.

Remanufacturing rate uncertainty directions: Ferguson et al. considered a tactical production-planning problem for remanufacturing when returns have different quality levels [10]. Xuhong studied the effect of quality uncertainty of parts on performance of reprocessing system in remanufacturing environment [11].

Remanufacturing machining path uncertainty directions: Li et al. proposed a Graphical Evaluation and Review Technique (GERT) network based on remanufacturing process model with respect to the uncertain remanufacturing process in remanufacturing systems [12]. Liu [13] studied the tolerance redistributing of the reassembly dimensional chain on measure of uncertainty; and so on.

The uncertainty exists in the whole life cycle of remanufactured products. Remanufacturing is a lot more uncertain than manufacturing [2, 3]. But the study on the uncertainty of reassembly process is relatively few. The higher uncertainty leads to incomplete control of the reassembly process which affects the quality and service security of remanufactured products. Research on reassembly uncertainty obtain useful information as much as possible to monitor the reassembly process and help to make proper decisions for improving the quality of remanufactured products. Therefore, research on remanufacturing uncertainty becomes the theoretical basis for the development of remanufacturing industry.

The uncertainty of remanufactured parts attributes is the one of the main factors causing the remanufacturing uncertainty. The uncertainty of reassembly dimensions is representative, as it is directly related to the accuracy and stability of remanufactured products. We analyze the uncertainty connotation of reassembly dimensions and build a measurement model to quantitatively measure. The assembly uncertainty coupling mechanism for reassembly dimensions is explored and the law that the reassembly quality is lower than the manufactured is revealed. These research provide the basic theories for reassembly process production management.

2 The Uncertainty Connotation of Reassembly Dimensions

Whether remanufactured products or manufactured products are the hybrid structure made up of several parts types in accordance with certain combinations of assembly sequence. Compared with remanufactured parts, manufactured parts are new ones processed by founding, forging, stamping, rolling and other methods to get a relatively better quality. Remanufactured parts are reuse parts or repair parts, which have greater uncertainty and instability since they have larger discrete degree and variance.

To quantitatively describe the difference between reassembly dimensions and assembly dimensions, we do a comparison as show in Fig. 1. Here we assume that the ideal variance of the assembly size is zero.

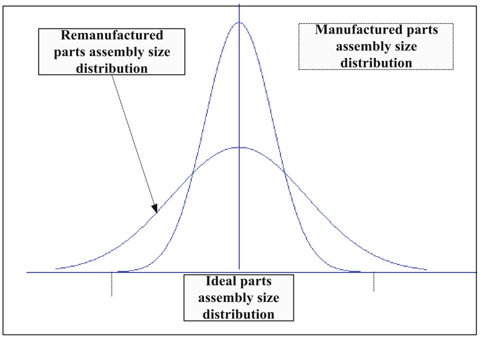


Fig. 1. Distribution of parts assembly dimensions

Figure 1 illustrates that the uncertainty of the remanufactured parts is higher than the manufactured from the view of assembly size.

As we know, uncertainty is absolute, certainty is relative. Owing to the particularity and complexity of remanufacturing, the uncertainty of remanufacturing is much higher than manufacturing, which is one of main differences between remanufacturing and manufacturing. Study on uncertainty theory for reassembly process is the premise thing and key to optimal management of remanufacturing process.

The information entropy theory provides a quantitative measure for the uncertainty measure [14], which is used in production system. such as: Frizelle and Suhov [15] proposed two kinds of manufacturing system complexity measurement model based on information entropy. Efstathiou et al. [16] proved that the system entropy is equal to the complexity of the manufacturing system. Kuzgunkaya and ElMaraghy [17], Zhang [18] and Liu et al. [19] applied entropy to the optimal decision of the manufacturing system, they described its complexity, disorder and uncertainty, and provides an effective solution to the uncertainty description of reassembly dimensions by entropy.

Shannon uses entropy as a measure of the uncertainty of a random event or the amount of information. As the information uncertainty on the opposite, as to eliminate the uncertainty of things, the amount of information that can be used to eliminate the uncertainty is how to measure the information entropy, so set up a bridge between the event probability information uncertainty [20]. The quantitative measure of uncertainty for reassembly dimensions, in essence, is to describe and quantify the degree of uncertainty of the reassembly accuracy. In this study, we use entropy measure the uncertainty of reassembly dimension to quantitatively describe the uncertainty of reassembly accuracy.

3 Uncertainty Measurements for Reassembly Dimensions

We assume that parts X has T assembly attribute points, and the sample data is n. $x(t)$ is the t-th assembly attribute point of parts X. $x(t)_i$ is the i-th sample value of $x(t)$.

$P(t)_i$ is the occurrence probability of $x(t)_i$, and $P(t)_i \geq 0$, $\sum_{i=1}^{n} P(t)_i = 1$

$x(t)_0$ is the expectation of $x(t)$, which describes the central tendency of assembly dimensions.

So, the variance of $x(t)$ is listed as follows:

$$\sigma[x(t)]^2 = \frac{\sum_{i=1}^{n} [x(t)_i - x(t)_0]^2}{n} \tag{1}$$

where, $\sigma[x(t)]$ means the dispersion degree of assembly dimensions, and its ideal value is 0.

Entropy is a tool to measure uncertainty. So we use the entropy $h[x(t)]$ to measure the uncertainty of $x(t)$.

According to the central limit theorem, small, independent and uniform random variables obey the normal distribution:

$$f[x(t)] = \frac{1}{\sigma[x(t)]\sqrt{2\pi}} \exp[-\frac{(x(t) - x(t)_0)^2}{2\sigma[x(t)]^2}] \tag{2}$$

According to the definition of entropy, we can get the following assembly dimensional uncertain entropy formula:

$$h[x(t)] = -\int f[x(t)] \ln f[x(t)] dx \tag{3}$$

The nature of the normal distribution shows that: when the variance is constant, normal distribution entropy is maximum and the entropy value is:

$$h[x(t)] = \ln\{\sigma[x(t)]\sqrt{2\pi e}\} \tag{4}$$

From the above Eq. 4, the following properties are available: (1) when $P(t)_i = 0$, assembly size uncertainty is minimum and its value is zero.

(2) When $P(t)_i = 1/n$, assembly size uncertainty is maximum.

Therefore, the uncertainty of the parts assembly dimension can be defined as the parts attribute entropy which is expressed as:

$$H(x) = \sum_{t=1}^{T} h[x(t)] = \sum_{t=1}^{T} \ln\{\sigma[x(t)]\sqrt{2\pi e}\} \tag{5}$$

Its simplification formula is:

$$H(x) = T \ln\sqrt{2\pi e} + \sum_{t=1}^{T} \ln\sigma[x(t)] \tag{6}$$

4 Coupling Mechanism of Reassembly/Assembly Uncertainty

As the two assembly dimensions are assembled together, the uncertainty is as follows:

One parts' assembly dimensions conform to normal distribution $N(x_1(t)_0, \sigma[x_1(t)]^2)$ and the other parts is $N(x_2(t)_0, \sigma[x_2(t)]^2)$.

Two parts assembled ideal size is: $x_1(t)_0 + x_2(t)0$.

After assembly, the variance is: $\sigma[x_1(t)]^2 + \sigma[x_2(t)]^2$.

Therefore, the assembly uncertainty entropy value is

$$h[x_1(t)_0, x_2(t)_0] = \ln\sqrt{2\pi e\{\sigma[x_1(t)]^2 + \sigma[x_1(t)]^2\}} \tag{7}$$

Similarly, if there are m assembly dimensions in an assembly node, as shown below:

The assembly uncertainty entropy value is

$$h[x_1(t), x_2(t), \ldots\ldots x_i(t)] = \ln\sqrt{2\pi e\sum_{i=1}^{m}\sigma[x_i(t)]^2} \tag{8}$$

On the basis of the research on uncertainty measurement model of parts assembly dimension, we measure the assembly uncertainty of ideal manufacturing, manufacturing and remanufacturing, the following two conclusions are drawn:

Corollary 1:

The uncertainty of reassembly dimensions is higher than the manufactured ones.

This paper considers that the entropy of the ideal parts assembly dimension is 0.

The assembly dimension entropy value of the manufactured parts is

$$H(x) = T \ln\sqrt{2\pi e} + \sum_{t=1}^{T} \ln\sigma[x(t)] \tag{9}$$

The assembly dimension entropy value of the remanufactured parts is:

$$H'(x) = T \ln \sqrt{2\pi e} + \sum_{t=1}^{T} \ln \sigma'[x(t)] \tag{10}$$

where, $\sigma'[x(t)]$ is remanufactured parts assembly dimensional tolerance. $\sigma[x(t)]$ is manufactured parts assembly dimensional tolerance.

Based on the uncertainty connotation of reassembly dimensions, we can get: $\sigma'[x(t)] > \sigma[x(t)]$.

Obviously, $H(x) < H'(x)$.

Corollary 1 is proved.

Corollary 2:

Under the same condition, the uncertainty of remanufactured product assembly is higher than manufactured.

A product has n assembly dimension chains, and each assembly dimension chain has m assembly dimensions.

So the assembly dimension entropy value of the manufactured product:

$$H = \sum_{j=i}^{n} h_j \left[\sum_{i=1}^{m} \pm x_i(t) \right] = \sum_{j=i}^{n} \ln \sqrt{2\pi e \sum_{i=1}^{m} \sigma[x_i(t)]^2} \tag{11}$$

The assembly dimension entropy value of remanufactured product:

$$H' = \sum_{j=i}^{n} h_j \left[\sum_{i=1}^{m} \pm x_i(t) \right] = \sum_{j=i}^{n} \ln \sqrt{2\pi e \sum_{i=1}^{m} \sigma'[x_i(t)]^2} \tag{12}$$

It is known that $\sigma'[x(t)] > \sigma[x(t)]$.

Therefore, $H' \geq H$.

Corollary 2 is proved.

These show that: under the same condition, the uncertainty of remanufactured product assembly dimension is higher than the manufactured. In other words, the accuracy of remanufactured products is lower than manufactured.

5 Example Simulation

This paper takes the remanufactured heavy-duty engine crankshaft as an illustrative example. Journal wear is one of the remanufactured engine crankshaft's main failure modes. Crankshaft journal wear mainly contains the spindle journal wear and the rod journal wear. The wear of rod journal is more terrible than the spindle journal. After wearing, the size of spindle journal is smaller and its shape is oval and the position which faces the rod journal is terribly worn. Additionally, because of lube impurities and the lubrication failure, the spindle journal will be scratched. Repair technologies are also different for the structure, material and damage degree differing between the bearing journal and the rod journal, such as chroming, iron-plating or thermal spraying, and so on.

We take the manufactured crankshaft and the remanufactured crankshaft as measuring objects. The data (Table 1) is calculated below (dimension unit: μm):

Table 1. Contrast between the bearing journal uncertainty and the rod journal uncertainty

Part	Classification	Assembly size	Tolerance	Variance	Entropy
Crankshaft bearing journal	Ideal manufactured	54000.00	0.00	0.00	0.00
	Manufactured	54000.00	30.00	213.16	4.01
	Remanufactured	53750.00	30.00	392.04	4.40
Rod journal	Ideal manufactured	48000.000	0.000	0.000	0.000
	Manufactured	48000.000	30.000	201.640	4.072
	Remanufactured	47800.000	30.000	345.960	4.340

To further study the impact of remanufactured product quality uncertainty, we collected quality data of remanufactured heavy-duty engine (RWD615) and manufactured heavy-duty engine (WD615) from January to December in 2016 (Fig. 2).

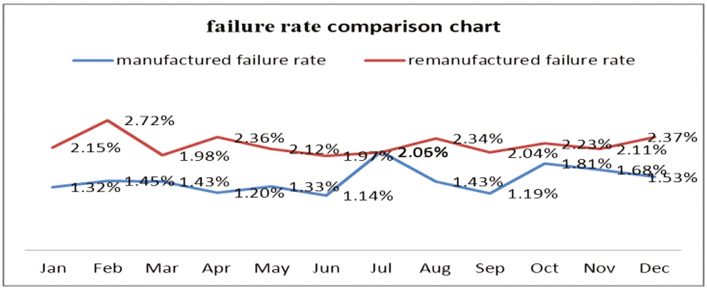


Fig. 2. Comparison chart of failure rate

It is found that the heavy-duty engine failure rate is shown below:

The compensation amount and frequency in 2016 contrast between RWD615 and WD615 is listed in Tables 2 and 3.

Table 2. Compensation detail of RWD615 and WD615(ten thousand CNY)

No.	Engine model	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total compensation amount
1	RWD615	24.7	14.49	16.01	28.54	24.48	18.69	20.3	15.03	18.5	23.89	35.3	19.6	259.53
2	WD615	15.51	12.62	13.18	18.87	11.46	11.7	10.51	10.53	12.12	10.75	17.33	11.73	156.31

Table 3. Compensation frequency detail of RWD615 and WD615 (times)

No.	Engine model	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total compensation frequency
1	RWD615	144	99	104	164	157	104	149	92	126	112	197	138	1586
2	WD615	76	50	51	87	52	54	52	45	68	69	86	52	742

It proved that there is a positive correlation between the uncertainty of assembly dimensions and the product quality. This also shows that remanufacturing uncertainty research is necessary. It is the basic theory to ensure that the quality of remanufactured products is not lower than the manufactured products.

6 Conclusions

Research on reassembly uncertainty is critical to guarantee the remanufactured product quality and its service safety, and it also provides one of the basic theories for the development of remanufacturing industrialized. How to understand and apply the reassembly uncertainty has become one of the key to optimize the management of remanufacturing process.

The uncertainty connotation of reassembly dimensions is analyzed and the uncertainty measurement model for reassembly dimensions is built. We explore the assembly uncertainty coupling mechanism and reveal that the remanufactured assembly quality is always lower than the manufactured. The Comparative results of the example and the inference of the coupling mechanism are worrying. But from the optimistic side, it opens the research direction of the uncertain optimization decision for remanufacturing. In the future, we will further study the coupling mathematical relationship between uncertainty and remanufacturing assembly accuracy.

Acknowledgments. This research is in part supported by Supported by China Postdoctoral Science Foundation (2017M611574), Natural science research project of universities in Anhui (KJ2017A438 and KJ2017A439) and Professor of Suzhou University (PhD) project for Research Foundation (2016jb08).

References

1. Xu, B.-S.: Remanufacturing Engineering and its Application. Harbin Institute of Technology Press, Harbin (2005)
2. Subramoniam, R., Huisingh, D., Chinnam, R.B.: Remanufacturing for the automotive aftermarket-strategic factors: literature review and future research needs. J. Clean. Prod. **17** (13), 1163–1174 (2009)
3. Ilgin, M.A., Gupta, S.M.: Environmentally conscious manufacturing and product recovery (ECMPRO): a review of the state of the art. J. Environ. Manag. **91**, 563–591 (2010)
4. Liu, M., Liu, C., Xing, L., et al.: Study on a tolerance grading allocation method under uncertainty and quality oriented for remanufactured parts. Int. J. Adv. Manuf. Technol. **87**, 1–8 (2013)

5. Qiang, Q., Ke, K., Anderson, T., et al.: The closed-loop supply chain network with competition, distribution channel investment, and uncertainties. Omega **41**, 186–194 (2013)
6. Amin, S.H., Zhang, G.: A three-stage model for closed-loop supply chain configuration under uncertainty. Int. J. Prod. Res. **51**(5), 1405–1425 (2013)
7. Inderfurth, K.: Impact of uncertainties on recovery behavior in a remanufacturing environment: a numerical analysis. Int. J. Phys. Distrib. Logist. Manag. **35**(5), 318–336 (2005)
8. Kenné, J.P., Dejax, P., Gharbi, A.: Production planning of a hybrid manufacturing–remanufacturing system under uncertainty within a closed-loop supply chain. Int. J. Prod. Econ. **135**(1), 81–93 (2012)
9. Su, C., Sha, Y.: Remanufacturing production planning based on mixed-uncertainty and evidence theory. J. Southeast Univ. (Nat. Sci. Ed.) **40**, 712–716 (2010)
10. Ferguson, M., Guide Jr., V.D., Koca, E., Souza, G.C.: The value of quality grading in remanufacturing. Prod. Oper. Manag. **18**(3), 300–314 (2009)
11. Tang, X., Mao, H., Li, X.: Effect of quality uncertainty of parts on performance of reprocessing system in remanufacturing environment. J. Southeast Univ. (English Ed.) **1**, 020 (2011)
12. Li, C.C., Li, C.B., Cao, H.J., et al.: Uncertain remanufacturing process routings model for used components based on GERT network. Comput. Integr. Manuf. Syst. **18**(2), 298–305 (2012)
13. Liu, C.: Tolerance redistributing of the reassembly dimensional chain on measure of uncertainty. Entropy **18**(10), 348 (2016)
14. Weaver, W., Shannon, C.E., et al.: Mathematical Theory of Communication, pp. 623–656. University of Illinois Press, Champaign (1949)
15. Frizelle, G., Suhov, Y.M.: An entropic measurement of queueing behaviour in a class of manufacturing operations. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, vol. 457, no. 2011, pp. 1579–1601 (2001)
16. Efstathiou, J., Calinescu, A., Blackburn, G.: A web-based expert system to assess the complexity of manufacturing organizations. Robot. Comput.-Integr. Manuf. **18**(3), 305–311 (2002)
17. Kuzgunkaya, O., ElMaraghy, H.A.: Assessing the structural complexity of manufacturing systems configurations. Int. J. Flex. Manuf. Syst. **18**(2), 145–171 (2006)
18. Zhang, Z.: Complexity of cellular manufacturing systems based on entropy models. J. Mech. Sci. Technol. **24**(11), 2275–2280 (2010)
19. Liu, C., Liu, M., Xing, L.: A research on the uncertainty measure and its application to the journal surface roughness of remanufactured engine crankshaft. Qiche Gongcheng/Automot. Eng. **37**(3), 341–345 (2015)
20. Liu, C., Liu, M., Wen, H.: Based on the uncertainty measure of remanufactured crankshaft's balance control. J. Mech. Eng. **51**(5), 171–177 (2015)

Multi-agent Collaborative Planning in Smart Environments

Flora Amato[1(✉)], Nicola Mazzocca[1], Francesco Moscato[2], and Fatos Xhafa[3]

[1] DIETI, University of Naples Federico II, Naples, Italy
{flora.amato,nicola.mazzocca}@unina.it
[2] DiSciPol, University della Campania Luigi Vanvitelli, Caserta, Italy
francesco.moscato@unicampania.it
[3] Department of Computer Science, Technical University of Catalonia, Barcelona, Spain
fatos@cs.upc.edu

Abstract. Nowadays *Smart* systems have become commonplace in our lives: domotics, social networks, automotive, smart application, virtual reality are having each time more and more users. One recent example of smart spaces can be found in domains like cultural heritages sites, museums or libraries where the use of new technologies grows up fast, namely, distributed sensors networks, virtual reality and smart systems are now being widely used to aid in preserving archaeological findings and sites as well as to enhancing presentation of cultural heritage assets. This work focuses on a problem that face the visitors at all large museums and ruins: the problem of scheduling tours depending on users preferences and on the time they can reserve for their visit. In particular, in museums and sites with a large number of visitors, the problem of queues to access to particular areas is also well-known. Sometimes waiting times are so high that visitors are not able to end their tours in time. In this work we present a modeling methodology and a planning technique able to redirect visitors tours in order to optimize their experiences within the desired available time. In addition, the system is able to face security and safety problems, providing a mean to redirect users to safe areas in case of emergency problems.

1 Introduction

An important problem for museums and archaeological sites is the organization of tours able to optimize visitors experiences and to increase tours appeals. In addition, tours must meet several other constraints like visitors safety (this may decrease significantly during tours organized for large numbers of groups or for museums with many visitors) as well as the preservation of cultural heritage items and assets.

Although organized tours usually have fixed visiting paths, due to the presence of many groups in a museum or in a site may introduce several delays in visits, as well as it may produce safety concerns. In addition, fixed tours do

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_22

not provide dynamical storytelling features able to adapt tours to visitors needs and their preferred criteria: for example, looking for paintings belonging to the same geographical region or historical period or school. These criteria are usually addressed by semantics based procedures.

In addition, usually visits take limited time: visitors have the necessity of ending their tour before a fixed deadline.

Sometimes available time is too short (or a museum is too large) that it is impossible to complete the visit to all the spaces in the area. In addition, the number and the characteristics of groups of visitors that are already in the site, can overload areas that become accessible only by using queuing policies for safety and security reasons.

Since organized tours are usually defined statically, it is not possible to focus directly on each user preferences, and to adapt the tour according to actual load. On the other hand, it is difficult for users to build personal routes as they may realize too late that their scheduled tours might not completed.

In this paper we discuss a model and a technique able to apply both classical and novel planning algorithms for recommending re-planned routes to visitors.

Our methodology is based on a Multi-Agent model and logics, and on Model Checking over Timed Automata. The problem of finding a schedule for a tour is managed as a reachability problem, namely, visitors agents express the *Goal* of visiting some items, rooms and areas within given deadlines. The formal analysis of the related Multi-Agent Model leads to a sequence of actions that visitors can enact to fulfill their needs. Such actions include *moving* from a site to another, *visiting* an area, *looking at* an item, *reading or hearing* a description and so on. This is often addressed as a *planning* problem.

In classical planning, one of the most used approaches to find a sequence of actions achieving a goal state, is the State Space Informed Search. It is based on the definition of heuristic functions used by planners in order to search paths to goals in the most efficient way. In the last years many works have tried to develop new strategies to optimize the planning problem by proposing other approaches. For example, in [1] the authors propose a domain independent approach for heuristics definition based on Pattern Databases approach. Works in [2] and BFBnB [3] introduce efficient search state planners and algorithms. Authors of [4,5] extend the classical planning problem by introducing time and resources constraints. Nevertheless, planning literature usually neglects a common problem, namely, the feasibility of the existence of a plan. Approaches based on State Space Search are not efficient in detecting if a Goal is not reachable in any way. In this work we are interested on the problem version, which includes the control of the non existence of planned route for a visitor (or a group of visitors) as well as the definition of a new route that satisfy users requirements and needs.

This problem is very important in Multi-Agent Systems, where concurrent execution of multiple agents complicate the state search problem because of resource sharing and agents interactions. Sometimes, especially when dealing with real-time systems [6], resource sharing [7,8] and interaction may lead to complexities in reaching the goals. In these cases, a way to determine in a short time goals non-reachability is really challenging since agents may choose to change goals without wasting time.

In order to change goals agents have to realize that it does not exist any plan reaching current goal. The work in [9] for example, proposes an efficient way to determine if, for a given goal, a plan of length n exists by examining the satisfiability of properly defined formulas. In our work we propose a framework composed by a Modeling Tool that provides the means to model Multi Agents Systems [10], and a planner based on counter example. This framework is fully integrated in the Unity 3D Engine, so it offers an easy way to realize 3D simulations. In the counter example approach, the planning problem is defined aiming to answer the following question: "is it true that from the initial state, only states where the goal conditions are not satisfied are reachable?". If the answer of this question is "yes" than no plan for the current goal exist. Otherwise, it is possible to find a "counter example" showing that at least one path exists from the initial state to a state where the goal conditions are satisfied; this counter-example, as we will show, represent a plan for the current goal.

The agent modeling approach is based on the Beliefs, Desires and Intentions (BDI) model and the action modeling is based on First Order Logic (FOL) STRIPS [11,12] formalism: for each action we define (a) a list of preconditions; (b) a list of postconditions (in spite of add and delete lists of STRIPS) (c) the domain definition for each of the variables in the previous lists. This last definition is very important because differently from the FOL semantics, where each agent belief can assume only one of two values (true or false), in our modeling approach we are able to associate any domain to beliefs. The counter example search is performed by the UPPAAL [13] Model Checker. This needs the application of proper model translation techniques in order to produce a timed automata representation of the multi-agent system. Counter examples are returned in form of UPPAAL traces and then translated into sequences of actions.

In the *Modeling* phase we provide a Multi-Agent System model of visitors and of the environment they move in. The *Verification* phase has the target of study reachability of users' goals and to provide a plan to reach them within specified deadlines. In addition, we provide a *simulation* environment in order to test visiting loads off-line. Finally, we provide a *run-time* system that collects real data from sites [14,15] that dynamically suggests new schedules to visitors whenever old one becomes unfeasible for any reason.

2 Agent Based Planning Problem

In this section we provide a simplified version of the model we use to define our planning problem. The model is based on a Multi-Agent System (MAS) representation of visitors and of the site where tours take place. We consider here a variant of Beliefs, Desires, Intentions (BDI) logics [16] for our Multi-Agent System model, namely a quadruple:

$$(Agents, World, \mathscr{TS}, \mathscr{F})$$

where *Agents* is the set of all agents in the system; $World$ represent the environment where agents execute; $\mathscr{TS}$ is a transition system that resumes possible state transitions of agents in the environment and, finally, $\mathscr{F}$ is a a set of formulas expressed in first order logics that characterize each state in $World$.

We use a triple $\langle n, d, v\rangle$ to define variables evaluations that describe states in the World, where n is the name of a variable, d represents its domain and v a value assigned to the variable. A state $s \in World$ is a set of variable evaluation.

In addition, we call *state conditions* of a state, the set of all formulas in $\mathscr{F}$ that holds in a state s:

$$StateCondition(s) = \{\phi \in \mathscr{F}, s \in World : s \models \phi\}$$

In addition, ϕ cannot be a sub-formula of other formulas holding in s.

In this work we consider $\mathscr{TS}$s and states with only one condition per state. If $s \models \psi$; $s \models \phi$ and $s \models \psi \wedge \phi$; then we consider only the last formula as state condition in s.

An Agent is in turn a triple:

$$(Actions, Beliefs, Goals)$$

where *Actions* is a set of possible actions an agent is able to perform. Actions modify the environment changing $World$ representation. They can also require the intervention of other agents in order to achieve common goals and, in general, they include *communication* and *execution* actions. In addition, an action can be *reactive* if its execution depends on external events or messages; or *proactive* if its execution is decided directly by the agent. We call *Proactive Agent* an agent with at least one proactive action; an agent with no proactive actions and with at least one reactive action is a *Resource Agent*; an agent with no reactive or proactive actions is classified simply as a *Resource*.

Further, $Beliefs$ include the knowledge the agents have about the $World$; the agent itself and other agents.

Notice that an agent may have a belief about the $World$ which in turn is *not* true in the environment: in general, beliefs of each agents may not be exact.

Goals is a set of states in $\mathscr{TS}$ that represent goals an agent wants to reach. Since $\mathscr{TS}$ is not available when agents are defined, by simplifying the notation we identify Goals with formulas in $\mathscr{F}$ that are satisfied in goal states. We call these formulas: *Goal Conditions*. Notice that a goal condition is a State Condition for a goal state.

Beliefs are managed as Worlds variables and states (they are practically local World representation in each agent). Agents define the $\mathscr{TS}$ Transition System on $World$ states by means of *Actions*. An action $\alpha \in Actions$ is a triple:

$$(name, Precondition, Effects)$$

where *name* is trivially the name of the action; *Effects* is a set of formulas that hold in the new state; *Precondition* is a formula that *evaluates* true in a state s in order to *apply*(execute) the action and to produce a state transition $(s \overrightarrow{\alpha} s')$, produces a transition from the state s to s'. If Precondition of α evaluates true in s, s' will be the same of s, except for variables involved in Effects evaluation. The values of these variables have to change in order to satisfy *all* effects in s':

$$\forall \phi \in Effectss' \models \phi$$

In addition we must consider that an agent executing an action can access only to its local representation of $World$, i.e. to its beliefs. Hence, if agent's beliefs and World State are not synchronized (i.e., if the agent has a wrong belief about the world), it is possible that Precondition is evaluated true on beliefs, but *not* on $World$ state.

In order to apply an action, we must execute the following two steps:

1. an agent tries to apply Effects in a state s producing a transition from s to s' if precondition is evaluated true on its *beliefs*;
2. if Precondition evaluates true in *World too*, then $s \overrightarrow{\alpha} s'$ both in agent's Beliefs and in $World$ too.

$\mathscr{TS}$ is then the Transition System defined by the application of all actions in any state of $World$, performed by *all* Agents in the model. The execution of an action to build $\mathscr{TS}$ must follow the two above-mentioned steps. State transitions apply both to agents $Belief$ and to $World$. Anyway Precondition control is enacted on Beliefs first. If evaluation fails on Beliefs, the action is not applied even if Precondition would evaluate true on $World$. In this model, a **Plan** to reach a Goal $\mathscr{G}$ with given Goal Condition is a *path* from a starting state to a state where the Goal Condition holds. Notice that in a Multi-Agent System, actions in a transition system can be executed by different agents, even concurrently. We consider here a path as a linear scheduling of concurrent applications of actions.

A Planning problem hence, is the problem of finding such a path or to state that the requested goal is unreachable in the current environment.

3 Methodology and Framework

The methodology used in this work is based on the model presented in Sect. 2 (see [17] for details).

In the *modeling* phase, we provide a representation of systems that is compliant with the model in Sect. 2.

The environment ($World$) is modeled in terms of states defined by means of $\langle n, d, v \rangle$ triples.

In addition, Agents are modeled in terms of Beliefs and Actions. Actions requires the definition of $(name, Precondition, Effects)$ triples, but we introduce in this phase an extension of the model since we need some additional information during planning. Hence we extend Actions definition to the quadruple:

$$(name, Precondition, Effects, VarInfo)$$

where $VarInfo$ set is necessary to specify variable domains, quantification (universal or existential), and eventually other properties.

During *Model Translation* and *Planning Phase*, a *Planner Engine* processes the system model. This produces a scheduling of actions representing the plan to achieve the requested goal. In the last phase, Agents perform plan's action by action at run time. The *Execution Environment* monitors correct behaviors of agents: if for any reason the current goal is no longer reachable, or if scheduled actions cannot be executed, (because preconditions no longer meet or because Beliefs were different from *World* real configuration), the Execution Environment enacts a *Replanning* action that tries to retrieve new plans for the same goals (if they exist).

In this work we consider communication among agents at different layers. The interaction model from Proactive Agents and Resource Agents to Resources is very simple: they can use a resource if it is *free* (i.e. if it is not used by other high level agents). Interactions between Proactive and a Resource Agents is more complex: a Proactive Agent can *ask for a service* to a Resource Agent (and this latter *serves* this kind of request). When a proactive agents asks for a service [18,19], it assigns a schedule of actions to a resource agent, providing a scheduling for a plan it wants to execute. Hence, in this work, a planning actions execute with the following steps:

1. A Proactive Agents finds a plan in an environment with one ore more Resource Agents and resources. Actions in the plan are related to Resource Agents actions which in turn may use simple Resources.
2. The proactive Agent with the new plan, assigns sub-goals to Resource Agents that, in turn, executes planning actions at a different layer of abstraction to reach sub-goals.
3. Resource Agents executes their sub-plans to reach sub-goals assigned by proactive agents. This usually involves the use of simple Resources.

Results obtained by the application of our methodology can be analyzed by a simulation framework [20]. The Simulation framework is developed by using Unity3D Engine[1] so it easily allows 3D simulations of planning problems. The framework uses the UPPAAL Reasoning Engine to perform counter-example searching. Two kinds of planners are implemented: a *Classical Planner* that is a simple implementation of the Breadth First and A* state space search algorithms, and *Counter Example Planner*. The presence of different planning strategies enables the implementation of a *Multi Expert* system. Multiple planners can start in parallel processes improving performances or providing solutions at different layer of abstraction. The Plan search algorithm in the Counter Example Planner works in four steps: (1) The TA (Timed Automata) translator implements an algorithm for Agent models translation from the model in Sect. 2 to Timed Automata formalism; (2) The TA is passed to the UPPAAL Reasoning Engine in order to produce a counter example (if any) to the following formula: "**A [] !'GoalCondition'**" that is *"is it true that from initial state, only states where the goal condition is not satisfied are reachable?"*; (3) If the previous property is satisfied, then no plan exists for the current goal; otherwise, UPPAAL

[1] https://unity3d.com/.

returns a *counter example* that is a path from the initial state to a state where the goal condition is satisfied, in form of an *Uppaal trace*; (4) The Uppaal trace translator perform the translation from trace to a sequence of actions that is the agent plan.

Automata generation follows a *Self Learning* approach: the planner expands current state of the agent considering only possible performable actions. The planning automata is updated only with new state during expansion. This considerably reduces the number of states to analyze during planning and monitoring phases.

4 Experimental Study

We evaluate how agent systems can be modeled with our methodology, through the definition of two applications scenarios.

By these applications, we also demonstrate the good performances of the Counter Example planner when solving the problem of determination weather a visitor's route cannot be completed with a given deadline.

We compare Planner results with the ones from a classical planner based on state space search.

The application context is the case of a museum where different visitors and groups of visitors are present. They may have different goals depending on their needs: someone may visit all areas in the museum, other may be interested in particular areas, authors etc. The Environment ($World$) consists in the description of the areas in the museum as well as of the entry and exits points of each area. Wireless sensors near items and doors in the area provide information about position of visitors, their number in areas etc. Let us assume that the museum exposes M items, and that N groups and K single visitors are in the museum.

We model entities in the museum as agents: Areas and Items are considered as *Resource Agents* since they have no reactive or active behaviors; visitors (and group of visitors) are modeled as Proactive agents.

Resource Agents Beliefs (i.e. Areas and Items) include the position and the name of the Area, the name of the author of the Item, the period and the date when the item was produced, the school the item and the authors belong etc. We address these beliefs with a dotted notation like $Item_i.author$ or $Item_i.date$. In addition, proper beliefs define how many visitors can stay near an Item and how many visitors can be queued to visit and area or an Item in the Museum:

- $item_i.visitors$: the number of visitor seeing an item;
- $item_i.weight$: the *importance* of the item in the museum;
- $item_i.maxVisitors$: The maximum number of visitors allowed for an item at the same time;
- $item_i.meanVisitors$: The mean number of visitors on an item at the same time;
- $item_i.meanTimetoVisit$: the mean time visitors stay near the item;

- $item_i.queueLimit$: the maximum number of persons in queue for an item;
- $item_i.visitorsInQueue$: the number of persons in queue for an item;
- $item_i.queueLimit$: the maximum number of persons for a queue;
- $item_i.meanTimeinQueue$: the mean time in queue for an item;

Similar beliefs are defined for areas (Fig. 1).

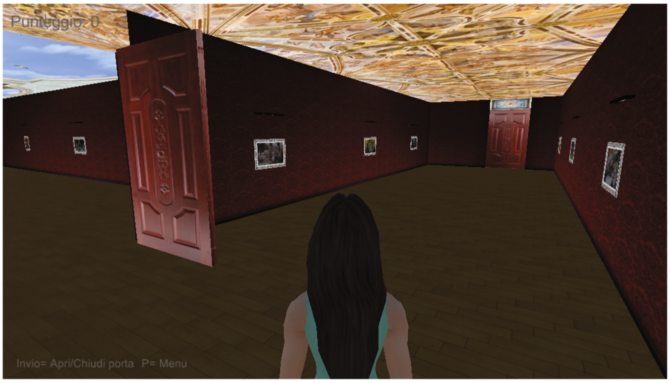


Fig. 1. Museum simulation route

The environment is modeled by the following formulas:

- isAdjacent($item_i$,$World.item_j$): a facts defining the adjacency of items and areas in the Museum.
- inArea($Area_i$, $World.item_k$): a fact defining if an item is in $Area_i$ in the Museum
- isFull($Area_i$): when an Area is full and cannot accept more visitors;
- isQueueFull($Area_i$): returns true if the queue to access to the area is Full
- isInArea($Visitor_i$, $Area_i$): returns true if a visitor is inside an area;
- canMove($Visitor_i$, $Area_i$, $Area_j$): states if a visitor can move directly from an area to another;
- isOnItem($Visitor_i$, $item_j$): defines if a Visitor in near an Item;
- isOnQueueforArea($Visitor_i$, $Area_j$): defines if a Visitor is queued to enter into and are;
- isOnQueueforItem($Visitor_i$, $item_j$): defines if a Visitor is queued for an Item;

The main Beliefs characterizing Visitors are:

- $Visitor_i.time$: the time spent by visitor at the museum;
- $Visitor_i.visited(item)$ is true if visitor saw the item;
- $Visitor_i.numberInGroup$ stores the number of persons in a group of visitors;
- $Visitor_i.position$, $Visitor_i.previousPosition$ store the positions of a visitor;
- $Visitor_i.atItem$ maintains the id of the Item which visitor is currently seeing.

- $Visitor_i.meanTimetoSee$ is the mean time a visitor takes to see and item;
- $Visitor_i.visitDeadline$, $Visitor_i.areaDeadline(Area_i)$, $Visitor_i.itemDeadline(Item_i)$ are the deadlines (if any) to visit the whole museum and each area and item;
- $Visitor_i.previousRoute$ and $Visitor_i.nextRoute$ are the path the visitor has already walked and the scheduled route respectively;

Main Actions proactive agents are able to execute are:

- *Move(from, to* from an area to another or from an item to another;
- *See(item)* an item in an area;
- *Leave(item)* an item in an Area;
- *Enqueue(Area) or (item)* for an area or an Item;
- *Exit(Area) or (Item)* from a queue;

For example, the first action is modeled as follows:

- Precondition: (position==from) AND (to!=from) AND (isAdjacent(from,to) ==true) AND (isFull(to)==false) AND $visitor.time < visitor.visitDeadline$.
- Effects: position=to, preposition=from, from.visitors=from.visitors-1; to.visitors=to.visitors+1.
- Variables: (from,"Sensor","Existential"), (to,"Sensor","Existential")

where obviously *Sensor* is the name of the domain defining all possible sensors in the museum.

Goals we want to reach can be different:

$$G_1 : \{\forall item Visitor.visited(item) == true\}$$

i.e. the goal of visiting all items in the museum (under global deadline) or:

$$G_2 : \{max \sum_i (item_i.weight * item_i.visited)\}$$

i.e. the goal of visiting major items in the museum within deadline, etc.

In order to analyze planner performances, we run simulation for different number of groups where a group is composed by a fixed number of person (in the example a group is made by 15 visitors), considering a single visitor reconsidering its route due to a possible future deadline expiration in visiting a museum with a given $World$ configuration (500 items, 50 areas and 10 items for area; a maximum number of visitor in an area of 25 persons). In addition, we have simulated replanning for both goals in the case $World$ configuration is such that one visitor cannot reach his goal before its deadline expires (i.e. more than half areas are full with half queues full too and full areas associated to the most *important* items that have been disposed in not adjacent areas).

In Fig. 2 average planning time for two goals are reported when number of groups increases. Notice the linear increasing time when considering the first goal and the exponential increases when dealing with a maximization goal. In

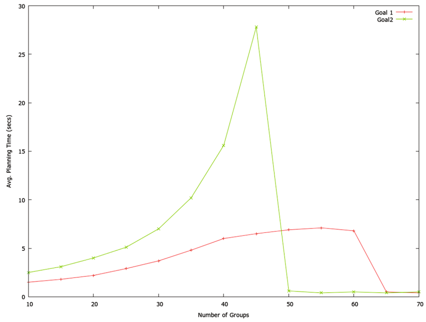


Fig. 2. Average planning time

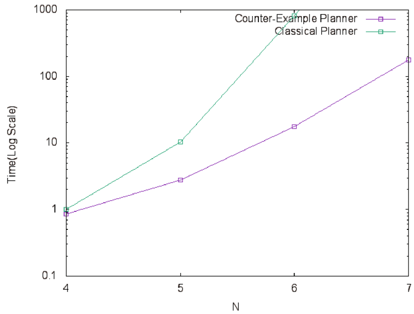


Fig. 3. Time comparison for the $N-$Queens

addition, it is possible to notice that when the number of visitors is too high, the planning execution time is really low. This happens when there exists no route matching deadline requirements (i.e. there is no feasible route within deadlines compliant with the goal defined by the visitor).

In this case, we have that the planner in short time is able to detect that the visitor cannot complete its visit if the environment will remain the same, actually this can be seen as one of the strengths of our approach.

For better understanding of our approach, we consider also a second application. A classical AI problem, the N Queen positioning: given an $N \times N$ chess board, the goal is to position N queens in a way they cannot attack each other. This problem in well known in the AI literature because of its high computational complexity. With our framework, the system environment is be modeled with the following *World* elements, defining initial state, (we'll consider for simplicity the case of $N = 5$):

(pos(q1), "Position",0); (pos(q2), "Position",0);
(pos(q3), "Position",0); (pos(q4), "Position",0);
(pos(q5), "Position",0); (index(q1), "Position",1);
(index(q2), "Position",2); (index(q3), "Position",3);
(index(q4), "Position",4); (index(q5), "Position",5);

In these elements, "Position" is the name of a domain defining the possible values of the $World$ variables: in this case the set of values [0, 1, 2, 3, 4, 5]. pos(q_i) is the row position of the ith queen and index(q_i) is its column position. If $pos(q_i) = 0$ the queen i is not currently positioned. The problem is modeled as a single agent system. Beliefs of the agent are modeled exactly in the same way of $World$ variables. Actions requires the definition of Precondition, Effects and VarInfo elements. We have here a single action defined as follows:

- Precondition: (pos(X)==0) AND (pos(X)!=Y) AND (Y!=pos(Z) AND) (Y!=(pos(Z)-(index(X)-index(Z)))) AND (Y!=(pos(Z)+(index(X)-index(Z))))
- Effects: pos(X)=Y

- Variables: (X, "Queen", "Existential") (Y, "Position, "Existential") (Z, "Queen", "Universal", (index(Z) < index(X)))

So "X" and "Y" are existentially quantified variables, "Z" is a universally quantified variable with the property that index(Z) < index(X), and "Queens" is a domain, in this case of $N = 5$ is the set of values [q1, q2, q3, q4, q 5]. The agent has a single goal, whose condition is, for $N = 5$: (pos(q1)!=0) AND (pos(q2)!=0) AND (pos(q3)!=0) AND (pos(q4)!=0) AND (pos(q5)!=0).

Figure 3 reports response times of the two planners(for the counter example planner the time shown is the sum of the automata construction and planning times).

In this simulation we can notice that the Counter Example Planner clearly overcomes the other planner: for $N = 7$ the planning time is not shown for the Classical Planner since the solution couldn't be found because it consumed all available hardware resources[2].

5 Conclusions and Future Works

In this paper we presented an autonomous planning methodology that is based on multi-agent models. The planning methodology exploits both classical and a new counter-example based approaches in order to build an effective multi-expert system able to face with increasing complexity of these systems [21,22]. We presented a framework able to enact the methodology. The methodology and the framework have been applied to a case study of museum where various visitors and groups of visitors are present. Simulation results showed that the counter-example planner executes faster than classical ones in some cases, motivating the use of a multi expert system in the framework. We also showed the ability of the new planner methodology when dealing with unreachable goals. Future works include the improvement of the counter-example planner algorithm and the extension of the methodology to conflicting interactions among agents and to coordinated execution of plans.

References

1. Culberson, J.C., Schaeffer, J.: Pattern databases. Comput. Intell. **14**(3), 318–334 (1998)
2. Hoffmann, J., Nebel, B.: The FF planning system: fast plan generation through heuristic search. J. Artif. Intell. Res. **14**, 253–302 (2001)
3. Zhou, R., Hansen, E.A.: Breadth-first heuristic search. Artif. Intell. **170**(4), 385–408 (2006)
4. Haslum, P., Geffner, H.: Heuristic planning with time and resources. In: Sixth European Conference on Planning (2014)
5. Baier, J.A., Bacchus, F., McIlraith, S.A.: A heuristic search approach to planning with temporally extended preferences. Artif. Intell. **173**(5), 593–618 (2009)

[2] We performed experiments on a Quad Core i7 4800 CPU with 4 Gbytes RAM.

6. Nishimura, M., Nishino, H., Kagawa, T.: A digital contents management system using a real booklet interface with augmented reality. Int. J. Space-Based Situat. Comput. **4**(3–4), 194–202 (2014)
7. French, T., Bessis, N., Xhafa, F., Maple, C.: Towards a corporate governance trust agent scoring model for collaborative virtual organisations. Int. J. Grid Util. Comput. **2**(2), 98–108 (2011)
8. Higashino, M., Hayakawa, T., Takahashi, K., Kawamura, T., Sugahara, K.: Management of streaming multimedia content using mobile agent technology on pure P2P-based distributed e-learning system. Int. J. Grid Utility Comput. 26 **5**(3), 198–204 (2014)
9. Rintanen, J., Heljanko, K., Niemelä, I.: Planning as satisfiability: parallel plans and algorithms for plan search. Artif. Intell. **170**(12), 1031–1080 (2006)
10. Kolodziej, J., Xhafa, F.: Supporting situated computing with intelligent multi-agent systems. Int. J. Space-Based Situat. Comput. **1**(1), 30–42 (2011)
11. Fikes, R.E., Nilsson, N.J.: STRIPS: a new approach to the application of theorem proving to problem solving. Artif. Intell. **2**(3), 189–208 (1972)
12. Burfoot, D., Pineau, J., Dudek, G.: RRT-plan: a randomized algorithm for strips planning. In: ICAPS, pp. 362–365 (2006)
13. Behrmann, G., David, A., Larsen, K.G., Pettersson, P., Yi, W.: Developing UPPAAL over 15 years. Softw. Pract. Exp. **41**(2), 133–142 (2011)
14. Amato, F., Barbareschi, M., Casola, V., Mazzeo, A.: An FPGA-based smart classifier for decision support systems. Stud. Comput. Intell. **511**, 289–299 (2014)
15. Amato, F., Mazzeo, A., Moscato, V., Picariello, A.: A framework for semantic interoperability over the cloud. In: Proceedings - 27th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2013, pp. 1259–1264 (2013)
16. Wooldridge, M.: Agent-based software engineering. IEE Proc. Softw. Eng. **144**(1), 26–37 (1997)
17. Moscato, F., Amato, F., Amato, A., Aversa, R.: Model-driven engineering of cloud components in metamorp(h)osy. Int. J. Grid Util. Comput. **5**(2), 107–122 (2014)
18. Di Lorenzo, G., Mazzocca, N., Moscato, F., Vittorini, V.: Towards semantics driven generation of executable web services compositions. J. Softw. **2**(5), 1–15 (2007)
19. Di Lorenzo, G., Moscato, F., Mazzocca, N., Vittorini, V.: Automatic analysis of control flow in web services composition processes. In: Proceedings - 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing, PDP 2007, pp. 299–306 (2007)
20. Moscato, F.: Model driven engineering and verification of composite cloud services in metamorp(h)osy. In: Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014, pp. 635–640 (2014)
21. Spaho, E., Barolli, L., Mino, G., Xhafa, F., Kolici, V.: Vanet simulators: a survey on mobility and routing protocols. In: Proceedings - 2011 International Conference on Broadband and Wireless Computing, Communication and Applications, BWCCA 2011, pp. 1–10 (2011)
22. Balzano, W., Murano, A., Vitale, F.: V2V-EN – Vehicle-2-Vehicle elastic network. Procedia Comput. Sci. **58**, 497–502 (2016)

Open Proctor: An Academic Integrity Tool for the Open Learning Environment

Alexander Amigud[1(✉)], Joan Arnedo-Moreno[1], Thanasis Daradoumis[1,2], and Ana-Elena Guerrero-Roldan[1]

[1] Department of Computer Science, Multimedia and Telecommunications, Universitat Oberta de Catalunya (UOC), Barcelona, Spain
{aamigud,jarnedo,adaradoumis,aguerreror}@uoc.edu

[2] Department of Cultural Technology and Communication, University of the Aegean, Mytilini, Greece

Abstract. Managing integrity of continuous and authentic assessments in an open and distance environments is a complex process. The core challenge is to map student identities with their academic work in effective and efficient manners, while preserving privacy, ensuring minimal disruption, minimizing impacts on accessibility and convenience. To address this issue, we have developed a prototype cloud-based application entitled OpenProctor that employs machine-learning techniques to analyze patterns in the learner-produced textual content in order to provide identity and authorship assurance. In contrast to the traditional academic integrity approaches that seek to control the remote learning environment, our method takes advantage of the readily available learner-generated data to analyze how the students go about doing their academic work. We present the framework, the architecture and main functions of the OpenProctor system and discuss the future research directions and plans for the system's future development.

Keywords: Academic integrity · E-assessment · Learning analytics · Machine learning

1 Introduction

Academic course credits are rarely issued by the virtue of participation. Their issuance is generally contingent upon meeting specific objectives following an assessment process that verifies their fulfillment. Assessment activities hold a central role of supporting knowledge and skills acquisition. Integrity of assessment is key to quality assurance and institutional credibility. This entails the need for a robust method of preserving academic integrity in any mode of instruction. In the bricks and mortar institutions this need is traditionally satisfied by invigilating assessment activities. Proctors verify students' identity documents

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_23

and restrict the use of any items (such as cell phones and crib notes) that may jeopardize the assessment integrity. An attempt to deviate from the assessment protocol may result in disqualification. In the learning environments that require physical attendance, turning a class into a proctored assessment session is as easy as adding observers. Students are already in the room, and all one needs to do to make the assessment secure is to look for suspicious behavior. This strategy does not translate well to the distance learning environment where students and instructors are separated by the time and space gap, and use technology to communicate. The traditional integrity strategies hold the open and distance learning back, because the concept of anytime and anyplace learning often excludes the assessment activities. The logistics of managing assessments across different time zones is draining resources away from the learning process. The alternative is to find a suitable compromise (at the institutional level) between integrity and convenience [3]. On the one end of the continuum there sits a complete avoidance of secure assessment [5], on the other end, distance institutions are following the footsteps of the traditional schools [3] which prefer security over convenience. Hence, the question arises: What strategies can help maintain academic integrity in the distance environment while imposing minimal impacts on administrative overheads, accessibility, convenience, and also disruption of learning?

This study is part of an ongoing research program to develop methods and tools that streamline electronic assessments. Our theoretical understanding of the problem—supported by the empirical data—has dovetailed into a pragmatic approach that uses machine-learning techniques to analyze patterns in the learner-produced academic content. We have developed a prototype cloud-based system entitled OpenProctor[1]—a tool that helps validate the integrity of authentic assessments. It provides both identity and authorship assurance in a single-tiered process, and produces a simple to understand report flagging potential cases of academic misconduct. OpenProctor puts the student in the center of the learning process, and does not seek to control the remote learning environment or impose restrictions on the way the students learn.

The rest of the paper is organized as follows: The next section provides the background for this work and discusses theoretical underpinnings. The system architecture and modules are discussed in Sect. 3. We conclude with a discussion of the future research and development directions in Sect. 4.

2 Background

In this section we review the common approaches to providing academic integrity and discuss how our approach differs from the traditional methods. To this end, we summarize the theoretical background of mapping student identities with their academic work through analysis of the student-generated content. In order to keep the scope of discussion manageable we place greater emphasis on the tools for maintaining academic integrity, although other aspects such as the policy, administrative and pedagogical practices play an equally important role and the

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

overall effectiveness of any academic integrity strategy should be examined as a whole.

2.1 Academic Integrity Strategies

The issue of academic integrity has been sporadically covered in the literature and received a variety of proposals much of which are using technology to give a digital facelift to the traditional strategies. These technology-based approaches can be classified into three categories: (a) those that aim to verify student identities, (b) those that validate authorship claims, and (c) those that aim to control the remote learning environment. For example, biometric authentication has been proposed to address the shortcomings of the password-based authentication [15]. Biometric technologies provide a higher level of identity assurance, but depending on modality, may require special hardware. Challenge question authentication extends the concept of the password authentication (what the user knows) by replacing a static data string, with a set of personally identifiable information [14] such as the mother's maiden name and a social security number. The underlying premise is that only the owner of the information will be able to answer the questions correctly. This strategy allows for automation of identity verification; the questions may be presented at random throughout the assessment session and can establish proof of presence. The method is accessible and does not require any special devices, however, when employed continuously this strategy may be disruptive.

The authorship claims are generally verified through plagiarism screening or environmental controls. In both cases, the aim is to collect evidence to refute the authorship claims. For example, schools often screen the submitted work for plagiarism using software tools [13]. Instructors may also run a quick search, using a popular search engine to selectively check parts of the student's work [3]. These tools check for similarity between the submitted text string and the database. If the text is significantly different from that in the database, the work is considered original, which leaves room for blind spots [11].

Exam invigilation [1] remains a popular academic integrity strategy suitable for any assessment task which provides both the identity and authorship assurance through a two-step process of identity verification and environmental control. However, it is resource intensive, logistically challenging, and costly. As such, it may affect the learners' levels of accessibility and convenience. Remote proctoring relieves some of the burdens of its traditional counterpart by abandoning travel requirements in favor of telecommuting [17]. Some services may employ biometric authentication and audio-video monitoring in conjunction with tools for remotely monitoring students' computers—which may be perceived as invasive due to sharing of access to a personal computer with a third party [16]. Instructor validation is another strategy for maintaining academic integrity often employed in the graduate-level courses [20]. It may prove to be valuable in settings where instructors work closely with students and are personally tracking the academic progress. However, this strategy is not readily scalable to a massive course level.

The levels of assurance that academic integrity strategies provide vary, depending on the methods used [3]. For example, invigilation strategies (traditional and human-managed remote proctoring) require verification of student identities prior to entering a monitored assessment session. Because invigilation is a resource-intensive process, it is often employed selectively, targeting the high-stakes assessments that are infrequent. The low-stakes assessments, often do not undergo the same level of scrutiny as their high-stakes counterpart. For example, the weekly online discussions are generally not subject to the authorship validation or identity verification other than the username/password within the learning management system (LMS). Projects and portfolios may undergo plagiarism screening, depending on institutional policy or at the instructor's discretion.

2.2 Theoretical Framework

"The post-traditional learners" crave control over how, where, and when to learn [6], which often comes in conflict with the tendency of the e-learning enterprise to play by the rules of the traditional schools and stipulate what learners can and cannot do. If one was promised the flexibility of anytime-anyplace learning, that promise ought to be kept and applied to the entire academic process, not just access to select services and activities. The main differences between our approach—discussed in the next paragraphs—and the traditional approaches to academic integrity is that the former puts the learner in the center of the learning process, and aims to analyze how the learners are doing the work, as opposed to maintaining control over the learning environment.

Earlier we proposed a behavioral-biometrics based framework for academic integrity that exploits availability of the student-generated textual content [4]. The underlying premise—which allows the mapping of learner identities to their work—is that artifacts produced by the same student are expected to be more stylistically similar to each other, than to that of other students. Writing style is thought to be individual specific [8,9], and by comparing patterns of language use, students' authorship claims can be validated (Fig. 1 depicts the process of analysis of the student work.) At the heart of this approach are computational [10] and stylometric techniques [21], that measure stylistic differences in the student-produced content. The process starts with identifying and measuring individual preferences for putting thoughts into words. Stylistic features—units that represent writing behaviour—need to be extracted from the raw data and quantified. Authorship research has yielded approximately 1,000 stylistic features [9], often organized into five types: character, lexical, syntactic, semantic, and application specific features [21], and any combination of them can be used for authorial discrimination. Once the numerical representation of the writing preferences is obtained for each student, a supervised algorithm for classification is trained on the data. A key advantage of using the machine learning algorithms, is that they learn from data without being explicitly programmed. Text classification is often performed using: Support vector machine-based algorithms, Bayesian-based algorithms, Tree-based algorithms, and Neural Networks, among other algorithms [2]. These algorithms can also be used in ensemble. Factors such as data set size, feature set size and type, number of candidate authors, and classifier parameters, may influence the quality of predictions. The challenge is to find the right combination of style markers and computational techniques, which is validated through experiments.

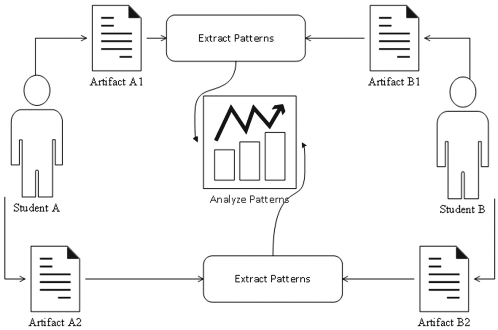


Fig. 1. Analysis of the student-produced content

Because the data is ready available, supervised machine learning techniques are employed and two sets of data are used—one for training the classifier and the other for testing the new data. In other words, a part of the content is used to create a stylistic profile that subsequent content will be compared against. Therefore, this strategy can be posed as a classification problem—where each student is defined as a class—but what is different to the e-learning context is that both the training and testing sets are labeled. The predicted class is expected to be of the student submitting the assignment and claiming its authorship. Let $Y = (y_1...y_i)$ denote class labels of the training set (student profile), and $Z = (z_1...z_i)$ denote class labels of the testing set (submitted assignments). A model is built from a set of labeled data $(x_i, y_i), i = 1...n$ comprised of a feature set X with a class label Y, to learn a function $y = f(x)$ to predict the class label Y for any new values of X. Because the labels of both the training and testing sets are supplied, each case of $y \neq z$ is flagged as a possible academic integrity violation for the instructor to follow up.

3 Architecture

A prototype cloud-based application—entitled OpenProctor—that employs machine-learning techniques to analyze patterns in the learner-produced artifacts was developed. The application has a modular architecture comprised of two components and five modules. The function of the former is to manage the student-generated data and pass it on to the analytics engine. Its two components include: (a) the data management interface and (b) the analytics engine.

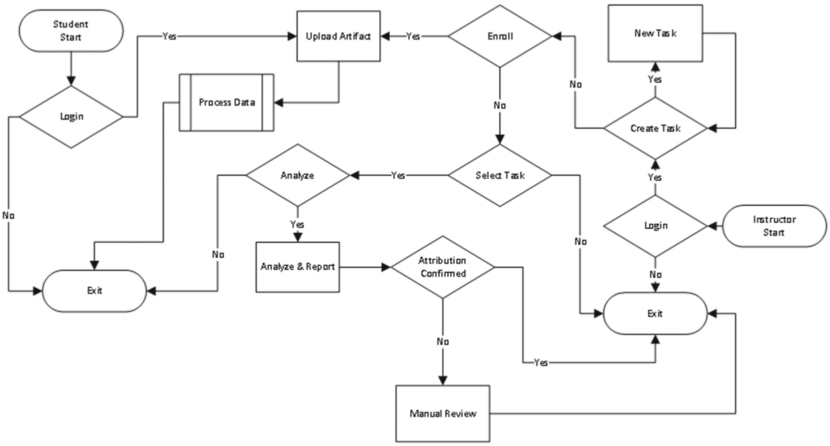


Fig. 2. Process diagram

The function of the latter is to provide the following: (1) user management, (2) learning task management, (3) student profile management, (4) validation of academic integrity, and (5) a messaging service. The function of the analytics engine is to process the data, discover and analyze patterns within, and generate the integrity report. These functions will be discussed in the next sections. The process diagram is depicted in Fig. 2.

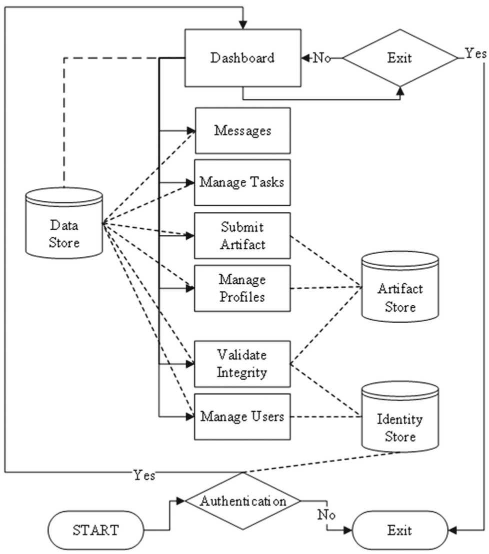


Fig. 3. System architecture

The current version of the OpenProctor has been implemented on a LEMP stack: Linux, Ngnx Web server, MySQL database with the Laravel PHP framework powering the data management functionality, and Python machine learning and natural language libraries [7,18] powering the data processing and analysis modules. A high-level system architecture is depicted in Fig. 3.

3.1 User Management

There are two user types in the OpenProctor system: the standard user (the student) and the administrative user (the instructor). The instructors manage core aspects of the system such as user management, learning task creation, generating reports, among others, while the students can upload their artifacts for each of the learning tasks. User accounts can be created by the instructor: by specifying the student's name, e-mail, password, and the type of the user account; and also by the student: through the account self-registration process, which by default makes all self-registration accounts to be of the standard user type. User passwords can be reset by the instructor or through the password recovery process.

3.2 Learning Task Management

Courses are composed of one or more learning activities, some of which culminate in production of student-generated content such as the written assignments or the online group discussions. These academic artifacts constitute minable data, containing author-specific patterns that can be used to validate the identity and authorship claims. Because not all learning activities bear an equal grade weight, the ones that are a part of the assessment process, need to be identified. For example, the students might be invited to participate in the weekly online discussions throughout the course duration, but only the discussions that address specific issues are counted towards the final grade. Although the entire discussion thread is collected, only a part of it will be used in the assessment process and needs to be defined as such. The task management function provides for doing just that. It is designed to segment the mapping of learner identities with the student-produced artifacts for each activity that is a part of the student assessment process.

Each validation task in the OpenProctor system corresponds to the assessment task in a particular course. If a course is composed of two graded assignments, there will be two tasks and two reports, one for each assessment item. The task has a name and status of being active or inactive. They can be activated, deactivated, renamed, created or deleted at any time during the course. Uploading student assignments is contingent upon task creation by the instructor. Students can only upload content as a response to a specific task, and therefore the assessment task creation by the instructor supersedes the work submission by the student. Once the data is uploaded, students can view, delete and resubmit their work. The student-level analyses are performed on a per-task basis, where one report per assessment task is generated.

3.3 Student Profile Management

In order to validate authorship claims of the submitted artifacts, one needs to have a basis of comparison. The student profile provides just that by enabling the instructor to select the learner-generated content that subsequent student work will be validated against. Student profiles collectively constitute the training set that the supervised machine learning algorithms are trained on, and the student's name is the label. The data is expected to be as close to the ground truth as possible, that is, having high degree of confidence that the content was produced by the student who claims authorship is necessary for maintaining validity. Only the instructor has access to this function, and the data in the profile may be updated as frequently as the instructor deems necessary.

3.4 Integrity Validation

Students access the OpenProctor system and submit assignments for each specific learning task. The assessment process is non-anonymous and requires an identifier such as the student's name which is a part of the user profile. Students are expected to complete and submit their own work, therefore identity and authorship claims are assumed to be of the same entity. The current configuration allows one submission per learning task.

The assignments are grouped together by the learning task and collectively form a testing set which can be validated against the data in the student profile—the training set. The instructor can perform two actions: first, to delete student submission and thus exclude them from the analysis and second, run the analysis which in turn generates a report depicting classification results and highlighting cases that require instructor attention. As students progress through the course materials and submit new content, the instructor performs the analyses for each task and issues the feedback either through the internal messaging system or any other means as necessary.

The instructor receives two types of reports: The dashboard depicted in Fig. 4 provides a general overview including the number and type of the system users, the number of active and inactive learning tasks, the size of the student profiles (training set), and the number of the submitted artifacts (testing set), the most recent files and their GeoIP location. The student-level report is based on the classification analysis that flags cases for the manual review. It identifies the cases of potential academic misconduct, something an instructor should look into further while performing the qualitative assessment (grading) of the submitted work. This functionality is not intended to relieve the instructors from being vigilant of possible cheating, but rather to provide an additional layer of checks and balances that helps to examine the content in quantitative terms. The report is only a suggestion to scrutinize the flagged cases, and it is still at the instructor's discretion to perform the additional checks.

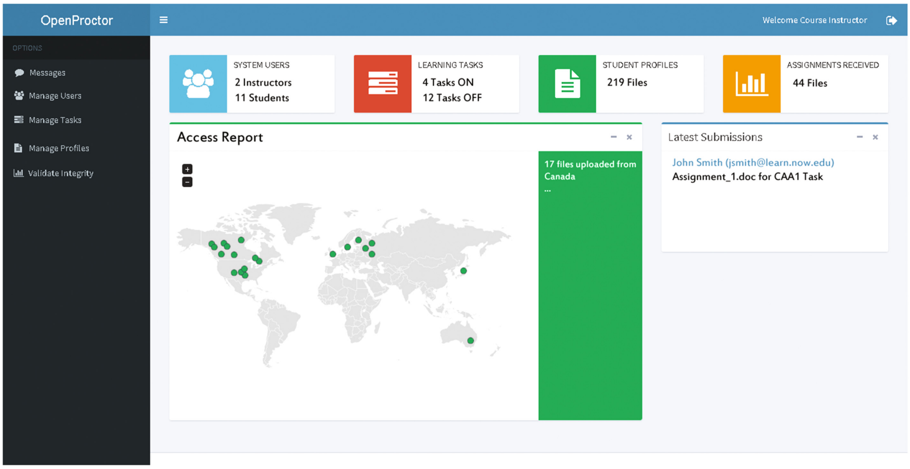


Fig. 4. Dashboard

3.5 Messaging

The messaging component facilitates the exchange of private and public messages among the users. For example, the instructors may post updates, make announcements, or provide students with feedback. Students may communicate their concerns or provide additional details with regards to their submissions. This module can be used alongside the message board of the LMS.

3.6 Analytics Engine

The raw data—the assignments that the students submit—are human readable and disseminated in a variety of file formats. These need to be made machine readable and therefore, formatted according to the input requirements of the machine-learning algorithm. The data undergoes preprocessing which removes the noise, tokenizes text, counts tokens, normalizes token counts, and turns them into numerical vectors. The feature vectors are fit with the classifier which returns a class prediction. Upon the analysis a report is generated, suggesting which artifacts the instructors should pay closer attention to when performing a qualitative review. The report comprises the following information: The number of students/assignments included in the assessment task, a list of students who work require additional review, classification matrix, name of the assessment task, and a timestamp.

OpenProctor has a modular architecture and the analytics engine is one of its interchangeable components that can be updated to include new algorithms and data processing functions. The computational tasks were implemented using Scikit-learn library [18] in Python language. To assess the performance of the analytics engine, we have conducted a series of experiments using real-world data from graduate-level research methods courses. The dataset was composed of academic writings ranging between 1,200 and 5,700 words and online discussion messages ranged between 250 and 450 words by 11 students. The feature set comprised nonconsecutive word pairs with a window size = 10, frequency of occurrence $\geq$5, and counts were TF-IDF normalized. The importance of each feature was computed using the Extra Trees classifier [12] and the feature space was reduced to the top 300 features. Experimental analyses were performed using the majority vote classifier [19] that weighed results from three supervised classifiers: (a) SVC with Linear kernel, (b) Logistic Regression, and (c) Multinomial Naive Bayes. All analyses were conducted using the 10-fold cross-validation method, using the 90:10 data split. The results suggest that the ensemble method outperformed individual classifiers yielding an accuracy rate of 87%. However, this performance metric does not provide relative effectiveness, and therefore additional study to establish a baseline of human performance was conducted. A group of 23 professional instructors—all of whom graded student papers in the past academic year—were invited to read and classify the excerpts from the same dataset, identifying texts written by the same student. The mean accuracy for the group was 12%. The results suggest that computational techniques implemented in the analytics engine of the OpenProctor system outperform experienced instructors in classifying short texts by author.

4 Future Directions and Conclusions

Although OpenProctor was tested on the real-world student data—yielding a high accuracy rate of prediction—it has not yet been on a test drive outside the lab. An array of field tests will be conducted in the future, deploying OpenProctor in a real or simulated course to evaluate its perceived efficacy and usability by the students and teaching staff. The participants' feedback will be incorporated back into the program design to enhance usability and functionality. On the development side, it will be important to further automate the data retrieval process and integrate with some of the popular LMS systems, and also automate the reporting, providing an option for on-event results dissemination via e-mail or instant-messaging. Additionally, we plan on expanding the capabilities of the analytics module by adding new algorithms for validating source code, music scores and visual artwork. This would allow us to provide the assurance in courses in a broader range of disciplines.

In the distance learning environment—where students and instructors are separated by the time and space gap—instructors rely on tools that help facilitate learning and maintain course integrity. In this paper we introduced a prototype cloud-based system designed to provide an accessible, non-invasive approach to identity and authorship assurance. OpenProctor aims to make the assessment process more efficient, as it does not rely on human invigilators to continuously monitor the learning environment, instead it is analyzing patterns in the learner-produced content. Assessments should not be stressful or logistically troublesome for learners or instructors, and the proposed approach strives to deliver just that. The analysis is conducted on the background, and allows the students to learn their way, and use tools they feel comfortable using. Its student-centered approach does not attempt to control the remote learning environment, but aims to provide flexibility and convenience of anytime and anyplace learning.

This study is part of an ongoing research program to develop new academic integrity approaches and empower instructors with tools that streamline student assessments.

Acknowledgment. This work was partly funded by the Spanish Government through projects CO-PRIVACY (TIN2011-27076-C03-02) and SMARTGLACIS (TIN2014-57364-C2-2-R).

References

1. Aceves, P.A., Aceves, R.I.: Student identity and authentication in distance education: a primer for distance learning administrators. Contin. High. Educ. Rev. **73**, 143–152 (2009)
2. Aggarwal, C.C., Zhai, C.: A survey of text classification algorithms. In: Aggarwal, C., Zhai, C. (eds.) Mining Text Data, pp. 163–222. Springer Nature, Boston (2012). doi:10.1007/978-1-4614-3223-4_6
3. Amigud, A.: Institutional level identity control strategies in the distance education environment: a survey of administrative staff. Int. Rev. Res. Open Distrib. Learn. **14**(5), 128–143 (2013)
4. Amigud, A., Arnedo-Moreno, J., Daradoumis, T., Guerrero-Roldan, A.E.: A behavioral biometrics based and machine learning aided framework for academic integrity in e-assessment. In: 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), pp. 255–262. IEEE (2016)
5. Bailie, J.L., Jortberg, M.A.: Distance learning student authentication: verifying the identity of online students. Find. Common Ground Programs Strat. Struct. Support Stud. Success **3**, 65–68 (2008)
6. Bichsel, J.: The state of e-learning in higher education: an eye toward growth and increased access. EDUCAUSE Center for Analysis and Research (2013)
7. Bird, S.: Nltk: the natural language toolkit. In: Proceedings of the COLING/ACL on Interactive Presentation Sessions, pp. 69–72. Association for Computational Linguistics (2006)
8. Brennan, M., Afroz, S., Greenstadt, R.: Adversarial stylometry: circumventing authorship recognition to preserve privacy and anonymity. ACM Trans. Inf. Syst. Secur. **15**(3), 1 (2012)
9. Brocardo, M.L., Traore, I., Saad, S., Woungang, I.: Authorship verification for short messages using stylometry. In: 2013 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–6. IEEE (2013)
10. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
11. Fiedler, R.L., Kaner, C.: Plagiarism-detection services: how well do they actually perform? IEEE Technol. Soc. Mag. **29**(4), 37–43 (2010)
12. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. **63**(1), 3–42 (2006)
13. Heckler, N.C., Rice, M., Hobson Bryan, C.: Turnitin systems: a deterrent to plagiarism in college classrooms. J. Res. Technol. Educ. **45**(3), 229–248 (2013)

14. Jortberg, M.: Experiences verifying the identity of distance learning students. Third annual report on identity in distance learning. http://www.uwex.edu/disted/conference/Resource_library/proceedings
15. Kang, B.H., Kim, H.: Proposal: a design of e-learning user authentication system. Int. J. Secur. Appl. **9**(1), 45–50 (2015). http://dx.doi.org/10.14257/ijsia.2015.9.1.05
16. Lilley, M., et al.: Remote live invigilation: a pilot study. J. Interact. Media Educ. **2016**(1), p. 6 (2016). http://dx.doi.org/10.5334/jime.408. 1-5
17. Mothukuri, U., Jain, S., Muralidharan, V.: Invigilated online assessment: various ways to minimize unauthorized help. In: 2012 IEEE Symposium on E-Learning, E-Management and E-Services (IS3e), pp. 1–4, October 2012
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
19. Raschka, S.: Python Machine Learning. Packt Publishing Ltd., Birmingham (2015)
20. Saunders, G., Wenzel, L., Stivason, C.T.: Internet courses: who is doing the work? J. Coll. Teach. Learn. (TLC) **5**(6), 25–36 (2011)
21. Stamatatos, E.: A survey of modern authorship attribution methods. J. Am. Soc. Inform. Sci. Technol. **60**(3), 538–556 (2009)

k-core Decomposition on Giraph and GraphChi

Xin Hu(✉), Fangming Liu, Venkatesh Srinivasan, and Alex Thomo

University of Victoria, Victoria, Canada
{huxin,jasmine1,srinivas,thomo}@uvic.ca

Abstract. The analysis of characteristics of large-scale graphs has shown tremendous benefits in social networks, spam detection, epidemic disease control, analyzing software systems and so on. However, today, processing graph algorithms on massive datasets is not an easy task not only because of the large data volume, but also the complexity of the graph algorithm. Therefore, a number of large-scale processing platforms have been developed to tackle these problems. GraphChi is a popular system that is capable of executing massive graph datasets on a single PC. Some researchers claim that GraphChi has the same or even better performance, compared with distributed graph-analytics platforms such as the popular Apache Giraph. In this paper, we implement a well-optimized k-core decomposition algorithm on Giraph. Then we provide a comparison of the performance of running the k-core decomposition algorithm in Giraph and GraphChi using various graph datasets.

Keywords: Vertex-centric model · Graph theory · Apache Giraph · GraphChi

1 Introduction

Graphs are widely used as a data structure for representing relationships between different objects or people. In many applications it is of great benefit to discover graph structure and analyse it. For example, advertisement companies utilize social network structure to find targeted communities in order to spread their commercials [6]; by detecting densely connected networks among web links, an organization can facilitate combating spam [24]; as another example, people simulate infectious disease spread by using graph structure in order to be better prepared to stop an epidemic [4]; software engineers use graphs to extract and analyse large-scale software systems so that the complexity of the systems is tackled more effectively [31].

In all these applications, the graphs are grouped into different subgraphs where some are dense and the others are sparse. Intuitively, nodes in a dense network have close ties with each other. So detecting those dense subgraphs in real-life networks is an important task in graph analytics. There are already many early studies on finding dense components such as in [7,15,25]. One popular study is the k-core decomposition which attempts to find all the maximal

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_24

connected subgraphs of a graph so that all vertices within it have at least k neighbors. In another dense graph searching study, [29] proposed a novel density measure that extracts optimal Quasi-Cliques and returns denser subgraphs with smaller diameters.

However today, graphs with millions of nodes and billions of edges are very common, and thus, doing analytics on large graph datasets is challenging due to the sheer size and complexity of graph computations. For example, in April 2017, the Facebook social network graph has over 1.97 billion monthly active users and more than 140 billion friendship connections, followed by Whatsapp and WeChat, with 1.2 billion users and 889 millions users, respectively [27].

Pregel [19], developed by Google in 2010, is a scalable and fault-tolerant platform that provides APIs for supporting large graph processing.

This model provides a vertex-centric computation model which enables users to only focus on programming itself without knowledge of the mechanisms behind it. The Pregel vertex-centric (VC) model has been implemented by several open source projects, for example, Apache Giraph is a framework designed to process iterative graph algorithms that can be parallelized across multiple commodity machines. Giraph became popular after careful engineering by Facebook researchers in 2012 to scale the computation of PageRank to a trillion-edge graph of user interactions using 200 machines [1].

Systems like Apache Giraph and Pregel require a distributed computing cluster to process large scale graph data quickly and effectively. Although distributed computing facilities such as cloud computing clusters are becoming more common and accessible, nevertheless, the question of how to process large scale graph data effectively without distributed commodity computing clusters is an interesting avenue for a data analyst who may need to analyze a large graph dataset but is unable to access a distributed computing cluster. GraphChi proposed by Kyrola and Guestrin [16] is a disk-based, vertex-centric system, which segments a large graph into different partitions. Then, a novel parallel sliding window algorithm is implemented to reduce random access to the data graph. Graphchi can process hundreds to thousands of graph vertices per second. GraphChi became popular, around the same time in 2012, as it made possible to perform intensive graph computations in a single PC in just under 59 min, whereas the distributed systems were taking 400 min using a cluster of about 1,000 computers (as reported by MIT Technology Review [23]). Since then, new versions of GraphChi and Apache Giraph have been released, where new ideas and optimizations have been implemented. Therefore, one needs to validate again the claims made several years ago. In [17], Lu and Thomo present a detailed evaluation of computing PageRank, shortest-paths, and weakly-connected-components on Giraph and GraphChi. In this work, we embark in computing k-core decomposition using a vertex-centric algorithm on Giraph and GraphChi. We adapt for this the algorithm of Montresor et al. [22]. The latter was implemented and optimized for GraphChi by Khaouid et al. [12].

2 Graphs and Cores

We consider undirected and unweighted graphs. We denote a graph by $G = (V, E)$, where V denotes the set of n vertices and E denotes the set of m edges linking the vertices. The neighbors of a vertex v are denoted by $N_G(v)$, and the degree of this vertex is denoted by $d_G(v)$.

Let $K \subseteq V$ be a subset of vertices of a graph $G = (V, E)$. We have the following definitions.

Definition 1. *Graph $G(K) = (K, E_K)$, where $E_K = \{(u, v) \in E : u, v \in K\}$ is called* the subgraph of G induced by K.

Definition 2. *(k-core): Graph G(K) is a k-core if and only if for each $v \in$ G(K), $d_{G(K)}(v) \geq k$, and G(K) is a maximum size subgraph with this property.*

The process of finding the k-core of a graph is to recursively prune all vertices that have degree less than k until converging to a subgraph in which all the vertices have degrees greater or equal to k.

Definition 3. *(Coreness): A vertex has a coreness of k if it is in the k-core but not in the $k+1$-core.*

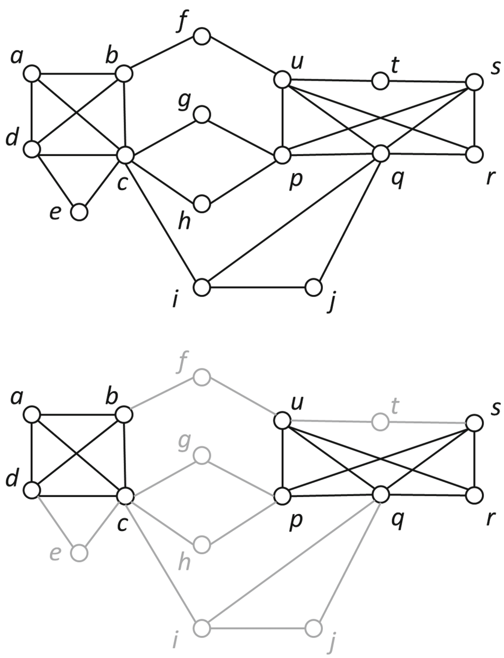


Fig. 1. Example graph (top). The 3-core of the graph (bottom) [12].

For an example see Fig. 1 (from [12]). By definition cores are nested, meaning 3-core also belongs to 2-core and 1-core. Each of the vertices in the graph is in 2-core and no vertices have coreness greater than 4. Even though vertices such as b, c, and q have a degree of 4 or more, their neighbours have degree less than 4, thus they do not belong in 4-core.

2.1 Pseudocode

[12] implements and optimizes an algorithm from [22] on GraphChi. It takes advantage of the vertex-centric programming model proposed in Pregel [19].

This model requires programmers to "think like a vertex" that can send messages to its neighbours by writing on to outgoing edges and receive messages from incoming edges, that is, in a graph, when computing an update function on a vertex, it enables the value on the vertex to be sent to its adjacent vertices. The computations in the update function go on iteratively until there are no more messages sent by vertices.

Computing the k-core decomposition becomes very natural in the vertex-centric model. First all vertices store an estimation of their coreness number in their vertex value, which initially is their degree. Messages can be used to propagate the estimation from the vertex itself to its neighbours using outgoing edges. Algorithms 1 and 2 show the flow of the computation in the vertex-centric Giraph model. They are adaptations of corresponding GraphChi algorithms in [12]. Even though both Giraph and GraphChi follow a vertex-centric model, there are differences in the way operations are expressed in each of them.

In Algorithm 1, the first superstep is a special case, where each vertex initializes its vertex value with its degree number which equals the number of its out-going edges. Then it sends this value to all its neighbours. Any vertex that is not being halted will be rescheduled to the next superstep (Lines 2–5).

In the next superstep, a function that computes an upper bound of the coreness of the vertex is assigned to a local estimate called *localEstimate*. The vertex value will be updated to the local estimate if the vertex value is greater than the upper bound of the vertex. In such a case, this new vertex value will also be sent to all the neighbours of the vertex (Lines 7–10).

The vertex will have a chance to lower its core estimate if any of its neighbours has a lower coreness value, then this vertex will be scheduled to the next superstep. Otherwise, this vertex will not be scheduled, and then it can switch to inactive state by voting to halt (Lines 11–20).

Algorithm 2 displays the details of the computation of a tighter upper bound of the vertex. It uses a *count array* which is indexed by the value of the upper bound of its neighbours. The value of each element is the number of the neighbours which have an estimate equal to the index. The largest index of the count array is the value of the vertex's the current coreness estimate. Any neighbour that has an upper bound greater than the current vertex value will be added to the last element of the array (Lines 5–8).

Algorithm 1. Update function running at a vertex

```
function UPDATE(Vertex vertex, Iterable messages)
    if superstep = 0 then
        vertex.value ← vertex.numOutEdges
        sendMessageToAllEdges(vertex, vertex.value)
    else
        localEstimate ←computeUpperBound(vertex, messages)
        if localEstimate < vertex.value then
            vertex.value ← localEstimate
            sendMessageToAllEdges(vertex, vertex.value)
        end if
        halt ← true
        for all message in messages do
            if vertex.value ≥ message then
                halt ← false
                break
            end if
        end for
        if halt = true then
            vertex.voteToHalt()
        end if
    end if
end function
```

Algorithm 2. ComputeUpperBound function for a vertex

```
1: function COMPUTEUPPERBOUND(Vertex vertex, Iterable messages)
2:     for all i ← 1 to vertex.value do
3:         c[i] ← 0
4:     end for
5:     for all message in messages do
6:         j ←min(message, vertex.value)
7:         c[j] + +
8:     end for
9:     cumul ← 0
10:    for all i ← vertex.value to 2 do
11:        cumul ← cumul + c[i]
12:        if cumul ≥ i then
13:            return i
14:        end if
15:    end for
16: end function
```

In order to compute the new upper bound of the vertex, a loop is used to add the array elements beginning from the largest index down to 2 until the summation of the array elements is greater or equal to the corresponding index. This index is the new coreness estimate for the vertex in the superstep (Lines 9–15).

3 Experimental Evaluation

All the experiments are conducted on Amazon Web Services (AWS) using the Amazon Elastic Compute Cloud (EC2) platform. We configured twenty-one virtual machines, with one master machine and twenty slaves. All of the virtual machines have two cores, Intel Xeon Family, 2.4 GHz CPU with 8 GB RAM running Ubuntu Linux System.

To explore the relationship between the number of the slaves and the running time, we respectively use two, five, ten, fifteen, and twenty slaves to handle different datasets. The datasets we used in this experiment were chosen from Stanford Large Network Dataset Collection. They are Astro Physics (ca-AstroPh), Gnutella P2P network (p2p-Gnutella31), Amazon product co-purchasing network (amazon0601), California road network (roadNet-CA), and Live-Journal social network (soc-LiveJournal1). The detailed information of these datasets is described in Table 1. From the table, we can observe that the first two datasets are small; they only have few thousands of nodes and a hundred thousands of edges. The medium sized datasets are amazon0601 and roadNet-CA with around three million edges. The largest dataset is soc-LiveJournal1, which has 4.8 million nodes and approximately 69 million edges.

Table 1. Datasets used for the experiments

Dataset name	Numbers of nodes	Numbers of edges
ca-AstroPh	18,772	198,110
p2p-Gnutella31	62,586	147,892
amazon0601	403,394	3,387,388
roadNet-CA	1,965,206	2,766,607
soc-LiveJournal1	4,847,571	68,993,773

Results for the Giraph implementation are shown in Table 2. Column "Sent Messages" gives the total numbers of messages that were sent during the whole computation. Column "Update Times" gives the average vertex update times for each dataset. "K-Max" and "K-Ave" are the maximum and average k-core numbers for each dataset. From the table, we can observe that the number of sent messages and vertex update times are not only dependent on the size of the datasets, but also on K-Max and K-Ave. The larger the latter numbers are, the more frequent the message sending and vertex updates will be.

Figure 2(left) shows the number of iterations executed on Giraph and GraphChi. The reason why the iteration numbers for the same dataset are different is that we cannot control the order of running each vertex in distributed cluster-based Giraph. However, because of the selective scheduling feature, the order is fixed when running GraphChi on a single machine. Except for the largest dataset and the smallest dataset, Giraph needs fewer iterations than GraphChi

Table 2. Results for the Giraph implementation

Dataset name	$\lvert V\rvert$	$\lvert E\rvert$	Sent messages	Update times (ms)	K-max	K-ave
ca-AstroPh	18.7 K	198.1 K	5,104,983	5414	17	2.01
p2p-Gnutella31	62.6 K	147.9 K	322,906	280	50	1.143
amazon0601	0.4 M	2.4 M	12,122,458	2284	10	2.51
roadNet-CA	2.0 M	2.8 M	11,035,492	785	6	1.999
soc-LiveJournal1	4.8 M	43.1 M	888,141,866	3507	434	1.689

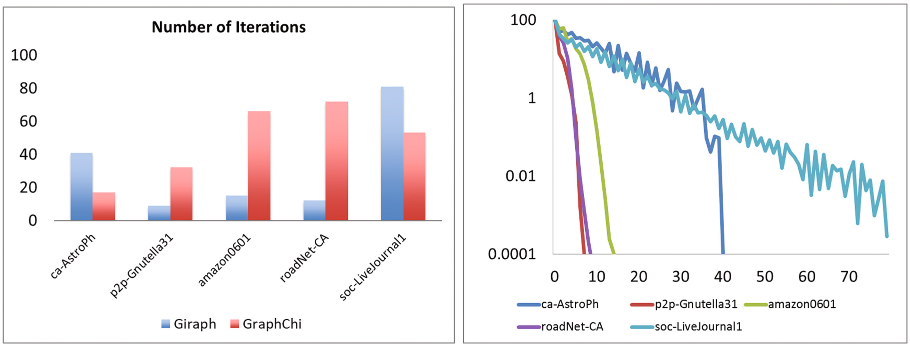


Fig. 2. Number of iterations (left). Percentage of updated vertices in Giraph vs. number of iterations (right).

with its advantage of running on multiple machines. The percentage of updated nodes over several iterations is shown in Fig. 2(right).

Figure 3, left and right, shows the running time (in milliseconds) of Giraph versus the number of slave machines used. In the left, we see that with the increase in the number of machines, the running time also increased. The more machines we have, the fewer the number of tasks assigned to each one are. However, the more machines we have, the more the amount of time spent on communication is.

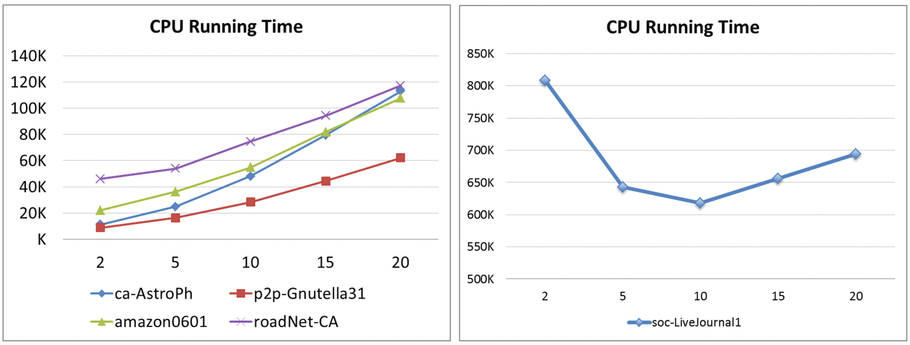


Fig. 3. Running time (ms) in Giraph vs. number of machines.

That is why the running time does not decrease when we configure more machines for Giraph. For the largest dataset shown in the right, we can notice that Giraph with two slave machines needs the most running time for the computation, which is around 800 s. On the contrary, it takes the least running time with ten machines, which can finish the computation within around 700 s.

To compare the running time with Giraph and GraphChi, we select the least running time with the proper number of machines for each dataset for Giraph. Figure 4 shows the running time comparisons between Giraph and GraphChi. Giraph spent more time than GraphChi on running the algorithm for all datasets. To be specific, the running time of Giraph and GraphChi are very close when dealing with medium data amazon0601 and roadNet-CA. However, GraphChi is more efficient in computing k-core for the small size and large size datasets on a single machine than Giraph with multiple machines.

However, the conclusion is that the performance of Giraph, with a relatively small number of machines, is quite close to the performance of GraphChi. This is in contrast to what was reported in 2012 in [23], where the situation was quite different. Then, a cluster of 1000 machines was an order of magnitude slower than GraphChi running on a single machine.

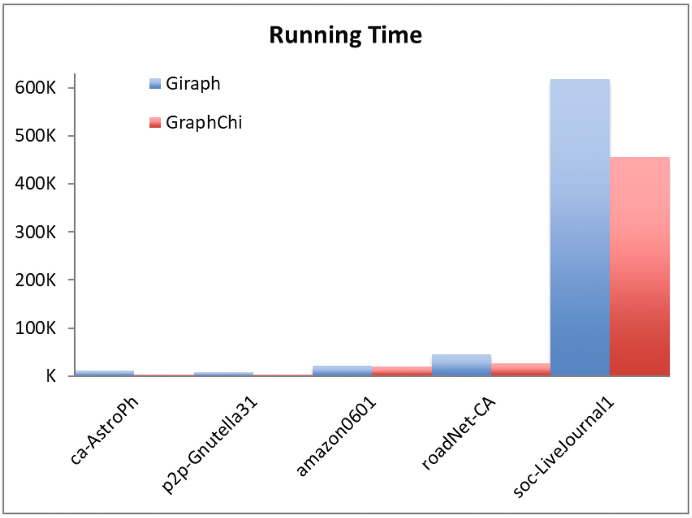


Fig. 4. Running time (ms) in Giraph compared to GraphChi.

4 Related Work

The Pregel distributed graph processing framework was introduced by Malewicz et al. [19]. Apache Giraph (http://giraph.apache.org) is an open source implementation of Pregel based on Hadoop. An excellent reference on Giraph is the recent book by Martella et al. [20].

GraphChi was created by Kyrola et al. [16]. Its excellent speed compared to distributed vertex-centric systems at the time (2012) was commented with awe at MIT Technology Review [23].

Around the same time, a group of Facebook researchers introduced several optimizations to Giraph [1]. These and other optimizations to Giraph are described in a recent paper by Ching et al. [2].

Thorough analysis of distributed vertex-centric systems have been presented by Han et al. [10] and Lu et al. [18]. A recent survey of vertex-centric frameworks is by McCune et al. [21].

5 Conclusions

From the experiments for k-core computation on Giraph and GraphChi, we observe that Giraph is suitable for analyzing medium- and large-size data since it can synchronously implement the computation by assigning the tasks to each slave. We observe that the performance of Giraph for computing k-core using a relatively small number of machines is in the same range as the performance of GraphChi for the same vertex-centric algorithm. As such, this is in contrast to the situation described in [23], where a cluster of 1000 machines was slower by an order of magnitude than GraphChi running on a single machine.

As future work, we would like to analyze more specialized graphs with their edges being labeled and/or weighted (c.f. [8,9]). It will be interesting to see how to devise vertex-centric algorithms for computing k-core on such graphs. Also, adapting the algorithms for environments with many machines failures (c.f. [26,28]) is another avenue to explore. Finally, we would like to explore the behaviour of Giraph vs. single machine systems for computing other complex graph analytics, such those for trust propagation and probabilistic graph summarization (c.f. [3,11,14]), as well as, complex analytics for special kind of graphs, e.g. user-item bipartite graphs for recommendation systems (c.f. [5,13,30]).

References

1. Scaling apache Giraph to a trillion edges. http://bit.ly/1TomAkh. Accessed 22 May 2016
2. Ching, A., Edunov, S., Kabiljo, M., Logothetis, D., Muthukrishnan, S.: One trillion edges: graph processing at Facebook-scale. PVLDB **8**(12), 1804–1815 (2015)
3. Chowdhury, M., Thomo, A., Wadge, W.W.: Trust-based infinitesimals for enhanced collaborative filtering. In: COMAD (2009)
4. Cui, W., Xiao, Y., Wang, H., Wang, W.: Local search of communities in large graphs. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, pp. 991–1002. ACM (2014)
5. Ebrahimi, S., Villegas, N.M., Müller, H.A., Thomo, A.: Smarterdeals: a context-aware deal recommendation system based on the smartercontext engine. In: Proceedings of the 2012 Conference of the Center for Advanced Studies on Collaborative Research, pp. 116–130. IBM Corporation (2012)
6. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)

7. Goldberg, A.V.: Finding a Maximum Density Subgraph. University of California, Berkeley (1984)
8. Grahne, G., Thomo, A.: Algebraic rewritings for optimizing regular path queries. Theor. Comput. Sci. **296**(3), 453–471 (2003)
9. Grahne, G., Thomo, A., Wadge, W.: Preferentially annotated regular path queries. In: Schwentick, T., Suciu, D. (eds.) ICDT 2007. LNCS, vol. 4353, pp. 314–328. Springer, Heidelberg (2006). doi:10.1007/11965893_22
10. Han, M., Daudjee, K., Ammar, K., Özsu, M.T., Wang, X., Jin, T.: An experimental comparison of pregel-like graph processing systems. Proc. VLDB Endow. **7**(12), 1047–1058 (2014)
11. Hassanlou, N., Shoaran, M., Thomo, A.: Probabilistic graph summarization. In: Wang, J., Xiong, H., Ishikawa, Y., Xu, J., Zhou, J. (eds.) WAIM 2013. LNCS, vol. 7923, pp. 545–556. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38562-9_55
12. Khaouid, W., Barsky, M., Srinivasan, V., Thomo, A.: K-core decomposition of large networks on a single PC. Proc. VLDB Endow. **9**(1), 13–23 (2015)
13. Khezrzadeh, M., Thomo, A., Wadge, W.W.: Harnessing the power of favorites lists for recommendation systems. In: Proceedings of the third ACM Conference on Recommender Systems, pp. 289–292. ACM (2009)
14. Korovaiko, N., Thomo, A.: Trust prediction from user-item ratings. Soc. Netw. Anal. Mining **3**(3), 749–759 (2013)
15. Kortsarz, G., Peleg, D.: Generating sparse 2-spanners. J. Algorithms **17**(2), 222–236 (1994)
16. Kyrola, A., Blelloch, G., Guestrin, C.: Graphchi: large-scale graph computation on just a PC. In: USENIX OSDI, pp. 31–46 (2012)
17. Lu, J., Thomo, A.: An experimental evaluation of Giraph and Graphchi. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 993–996. IEEE (2016)
18. Lu, Y., Cheng, J., Yan, D., Wu, H.: Large-scale distributed graph computing systems: an experimental evaluation. PVLDB **8**(3), 281–292 (2014)
19. Malewicz, G., Austern, M.H., Bik, A.J., Dehnert, J.C., Horn, I., Leiser, N., Czajkowski, G.: Pregel: a system for large-scale graph processing. In: SIGMOD. ACM (2010)
20. Martella, C., Shaposhnik, R., Logothetis, D.: Practical Graph Analytics with Apache Giraph. Apress, Berkeley, CA (2015). doi:10.1007/978-1-4842-1251-6
21. McCune, R.R., Weninger, T., Madey, G.: Thinking like a vertex: a survey of vertex-centric frameworks for large-scale distributed graph processing. ACM Comput. Surv. (CSUR) **48**(2), 25 (2015)
22. Montresor, A., De Pellegrini, F., Miorandi, D.: Distributed k-core decomposition. IEEE Trans. Parallel Distrib. Syst. **24**(2), 288–300 (2013)
23. Pavlus, J.: Your laptop can now analyze big data. MIT Technol. Rev. **1**, 1 (2014)
24. Saito, H., Toyoda, M., Kitsuregawa, M., Aihara, K.: A large-scale study of link spam detection by graph algorithms. In: Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web, pp. 45–48. ACM (2007)
25. Seidman, S.B.: Network structure and minimum degree. Soc. Netw. **5**(3), 269–287 (1983)
26. Shoaran, M., Thomo, A.: Fault-tolerant computation of distributed regular path queries. Theoret. Comput. Sci. **410**(1), 62–77 (2009)
27. Statista. Leading social networks worldwide, ranked by number of active users (in millions) (2017). https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

28. Stefanescu, D.C., Thomo, A., Thomo, L.: Distributed evaluation of generalized path queries. In: SAC. ACM (2005)
29. Tsourakakis, C., Bonchi, F., Gionis, A., Gullo, F., Tsiarli, M.: Denser than the densest subgraph: extracting optimal quasi-cliques with quality guarantees. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 104–112. ACM (2013)
30. Yazdanfar, N., Thomo, A.: Link recommender: collaborative-filtering for recommending urls to Twitter users. Procedia Comput. Sci. **19**, 412–419 (2013)
31. Zhang, H., Zhao, H., Cai, W., Liu, J., Zhou, W.: Using the k-core decomposition to analyze the static structure of large-scale software systems. J. Supercomput. **53**(2), 352–369 (2010)

A Bayesian Network for Improving Organizational Regulations Effectiveness: Concurrent Modeling of Organizational Resilience Engineering and Macro-Ergonomics Indicators

A. Azadeh[1,2], M. Partovi[1], M. Saberi[2(✉)], Elizabeth Chang[2], and Omar Hussain[2]

[1] School of Industrial Engineering, University of Tehran, Tehran, Iran
[2] School of Business, UNSW Canberra, Canberra, Australia
m.saberi.post@gmail.com

Abstract. This study presents a novel Bayesian Network (BN) based model for improving organizational regulations effectiveness (OREF) through concurrent modelling of organizational resilience engineering (ORE) and macro-ergonomics (ME) indicators in an organization. Six indicators namely teamwork, preparedness, fault tolerance, flexibility, redundancy and self-organization are considered as representatives of ORE. The macro-ergonomics indicators considered in this study are redesign, decision-making pace and information flow. The construction of the proposed model is composed of five steps. The BN is then used to track the status of OREF by considering ORE and ME indicators. Since these indicators represent the nodes of the Bayesian Network, in the first step they have been selected and confirmed by experts to be further considered in the model. The causal relationships between the nodes are acquired through aggregating experts' opinions by using the Dempster-Shafer theory.

Keywords: Resilience Engineering (RE) · Macro-Ergonomics (ME) · Bayesian Network (BN) · Dempster-Shafer theory

1 Introduction

For years, human errors and individual component failure were considered the basic reasons for many accidents [1, 15, 21–24]. It has been lately argued that human error was a causal factor in the occurrence of many serious accidents in the railway system, both in the UK [2] and across Europe [3].

Many organizations are looking forward to reduce human errors to increase efficiency. For example, a petrochemical unit is one of the places that has the potentiality prone to accidents. In other words, in these places human mistakes impose at great costs to the organization and hence reduce efficiency.

In this regard in the recent years a large number of articles have been published which aimed to reduced human errors and consequently increase organizational efficiency. Madigan et al. [4] examined the relationship between active and latent factors and found a solution for these types of incidents by using human factor analysis and classification system (HFACS) to examine rail industry incident reports.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_25

As mentioned earlier, human error and safety management are the subjects which many organizations face and in this regard, several methods and approaches have been suggested which aim to increase organizations, ability regarding them. In another study, Azadeh and Zarrin [5] have evaluated the efficiency and effectiveness of a large petrochemical plant's staff by considering three concepts namely Resilience Engineering (RE), motivational factors in the work environment and Health, Safety Environment and Ergonomics (HSEE).

Malgorzata Pecillo [6] checked out whether the resilience engineering concept is related to the implementation of occupational safety and health management systems (OSH MSs) and to safety levels in Polish enterprises of different sizes and activities. In the literature, two concepts and approaches namely Resilience Engineering and Integrated Resilience Engineering have been proposed to reduce accidents and increase efficiency.

If we were to break down Resilience Engineering from the engineering point of view, that would be the practice of identifying and detecting the instabilities, variations, interferences or any sort of surprises and actually try to change, revise or tackle the identified issues [7]. During the recent years, a strong interest has been shown in organizational resilience in both theory and application [8].

Nowadays the resilience engineering approach appears to be a practical idea, considering the fact that it can extensively increase the safety in complex systems since it has the ability to incorporate both safety and performance together in a functional way without causing any conflict in terms of performance and safety [9].

Resilience engineering is the ability of organizations for unforeseen events and adapts to the potential for accidents [10]. In fact, RE either improves safety of systems in the process industry or as a new method can control incidents and limit their outcome [11].

Azade and Zarrin [5] showed that five factors from Resilience Engineering, two from Health, Safety, Environment and Ergonomics (HSEE) and three factors from work motivational factors (WMFs) have significant positive effects on effectiveness. Assessment and studies over Resiliency factors (RFs) in an unpredictable environment shows that this approach has significant impacts on the management of enterprises.

Resilience engineering and macro-ergonomic focus on reducing risks and increasing efficiency within organizations. In this paper ten factors that were selected from both the fields of resilience engineering and macro-ergonomics were introduced and the relationship between them analyzed.

The purpose of this analysis is to identify the mutual influence of factors among each other as well as to determine the degree of this effect. All this analysis has been performed by a Bayesian Network modeling approach. Accordingly, the aim of this paper is to analyze the sensitivity between the factors that are selected from the fields of resilience engineering and macro-ergonomic. In other words, the aim of this work is to investigate the achievable level efficiency by considering resilience engineering and macro-ergonomics.

The paper has been organized as follows. Section 2 presents the methodology and result of the work is discussed in Sect. 3. The paper is finally concluded in Sect. 4.

2 Methodology

This study aims to find the relationship and also the level of importance of ORE and ME indicators on organizational regulations effectiveness (OREF). The six resilience engineering factors are as follows: *teamwork*, *preparedness*, *flexibility*, *fault- tolerance*, *redundancy* and *self-organization* and the three factors selected from macro-ergonomics are: *redesign*, *DM speed ability* and finally *information flow*. In this study, these nine factors are systematically analyzed and evaluated through a Bayesian network model. The effect of these factors on each other as well as OREF are investigated. Results and Sensitivity analysis are finally represented in the next section. The Bayesian network modelling in our study includes the five following step (Fig. 1).

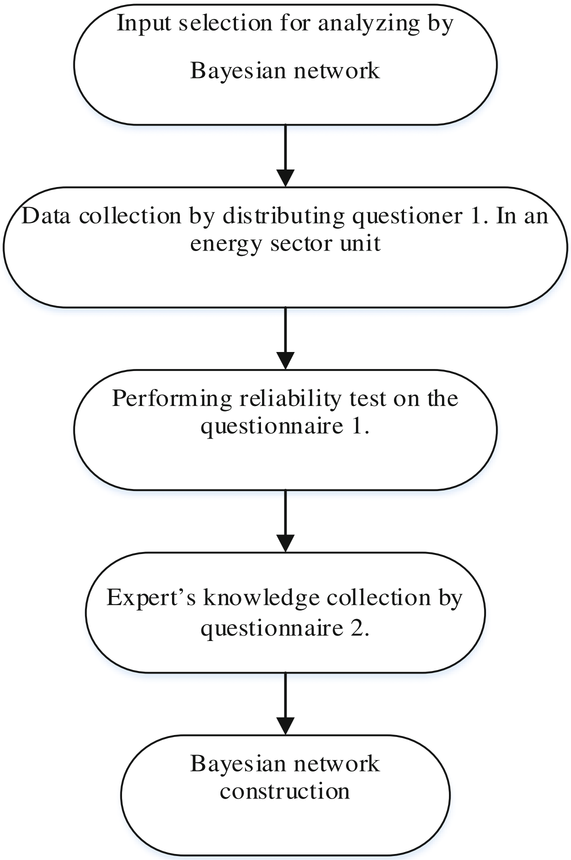


Fig. 1. Flow chart for Bayesian network construction by considering ten factors selected from resilience engineering and macro-ergonomic.

2.1 Bayesian Networks

Bayesian networks (BNs) are useful tools for demonstrating the causal relationships and interdependence between a number of factors [12, 13]. BNs are a well-established graphical formalism for the demonstration of conditional probabilistic relationships among indecisive variables [14]. Since BNs are based on experts' knowledge they are a very powerful tools which can be utilized in many different fields for the sake of modeling. Especially the causal interrelationships among some variables can be easily modeled using BNs [12, 13]. Moreover, BNs are a graphical structure that can consider and analyze a large number of variables in state of uncertainty. Nodes in a BNs represent variables that may be either connected together directly by an arc or connected indirectly.

3 Case Study and Computational Results

A large energy sector unit consisting of 150 operators, was selected for this study. The first part of the required data was obtained by distributing questionnaire (Questionnaire 1) among the operators located in control rooms, who constituted the target sample. The questionnaire consisted of 80 questions, scoring of each question ranged from 1 to 20. By using the questionnaire, the basic demographic information of operators is also extracted. Standard items of the Questionnaire 1 were adopted from reliable resources [15–17]. The other required data that were necessary to build the Bayesian network were collected through Questionnaire 2.

3.1 Input Determination

As mentioned before, the aim of this study is to improve organizational regulations effectiveness (OREF) through concurrent modelling of organizational resilience engineering (ORE) and macro-ergonomics (ME) indicators in an organization. Six indicators namely teamwork, preparedness, fault tolerance, flexibility, redundancy and self-organization are considered as representatives of ORE. The macro-ergonomics indicators considered in this study are redesign, decision-making pace and information flow. We leverage the power of BNs to improve OREF through concurrent modelling of ORE and ME. The nine factors act as the nodes of the Bayesian network, hence in the first step they are selected and confirmed by experts to be further considered in this study.

3.2 Collecting Information by Questionnaire 1

In this step, standard questions related to the nine factors of both organizational resilience engineering (ORE) and macro-ergonomics (ME) along with organizational regulations effectiveness (OREF) were considered in designing Questionnaire 1. This questionnaire is then distributed among the operators. They can chose answers to the questions range from 1 to 20 where 1 is considered the worst score and 20 as the best score. The operators also responded to the items of the questionnaire regarding demographic information.

3.3 Testing the Reliability of Results Obtained by Questionnaire 1

Another important task before lunching BN is to test the reliability of the designed questionnaire. In this regard, Cronbach's alpha is calculated and used to estimate the reliability of the questionnaire:

$$Cronbach's alpha = \frac{n}{n-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right) \quad (1)$$

Where:

n indicates the number of questions and, s_i *and* s_t *represent the standard deviation of the* i^{th} *question and total standard deviation respectively [18].*

As a rule of thumb, we required a value of 0.6 or higher for Cronbach's alpha before using the questionnaire. Table 1 shows that the all ten factors achieve this value which allow us to use the designed questionnaire.

Table 1. Reliability of questionnaire 1 by Cronbach's alpha

Factor number	Factor	Cronbach's alpha
1	Information flow	0.855
2	Teamwork	0.617
3	Redesign	0.770
4	Self-organization	0.832
5	Fault-tolerance	0.896
6	Flexibility	0.828
7	Preparedness	0.818
8	Organizational regulations effectiveness	0.904
9	DM speed and control ability	0.844
10	Redundancy	0.741

Thus according to the value for alpha presented in Table 1, the reliability of the first questionnaire was accepted.

3.4 Collecting Experts' Knowledge by Questionnaire 2

In this step, Questionnaire 2 has been designed for collecting experts' knowledge to obtain information required for building the Bayesian Network.

The first step in conducting a BN analysis is to build a causal graphical model in the form of a directed acyclic graph (DAG) that represents interrelationships among all desired variables [20]. There is not a unique method in the literature to build such a network. We utilize Dempster-Shafer theory for soliciting the domain knowledge of experts. Dempster-Shafer was used by Mohammadfam et al. in the process of BN construction [20]. To this end, the Questionnaire 2 was designed and three experts with the appropriate domain knowledge of resilience engineering complete the Questionnaire 2. This questionnaire demonstrates three possible relationships between each two factors which then the experts were asked to reveal their opinions on what type of relationship they mostly think is true and should be formed between those two factors. For example, there could be three possible relationships between two variables, namely A and B. A affects B (A → B), B affects B (A ← B) or no relationship exists between A and B (A ↑ B) [20].

3.5 Bayesian Network Construction

In this step, according to the information collected in the previous section and by using the Dempster-Shafer theory the BN was constructed. The Dempster-Shafer theory is a useful method to reduce inconsistencies when information is collected from various sources [20]:

$$m(a) = \frac{1}{1-k} \sum_{A_1 \cap A_2 \cap A_3 \ldots \cap A_n = A}^{n} m(A_1).m(A_2).m(A_3) \ldots ..m(A_n) \tag{2}$$

$$K = \sum_{A_1 \cap A_2 \cap A_3 \ldots \cap A_n = \emptyset}^{n} m(A_1).m(A_2).m(A_3) \ldots ..m(A_n) \tag{3}$$

In these equations, k represents the amount of discrepancies between the source of information and $(1 - k)$ is the normalizing factor. To clarify this approach better, an example is given below:

Example: Assuming that two experts have offered the following probability values for the three possible relationship between the two factors information flow and preparedness;

Expert 1 = (information flow → preparedness = 0.6, information flow ← preparedness = 0.3, information flow ↑ preparedness = 0.1).
Expert 2 = (information flow → preparedness = 0.3, information flow ← preparedness = 0.5, information flow ↑ preparedness = 0.2).

Table 2. Calculating the integrated mass probability for each relationship between information flow and teamwork based on the opinions elicited from expert 1 and expert 2 (a = Information flow → Preparedness, b = Information flow ← Preparedness, c = Information flow ↑ Preparedness)

	$M_1(a) = 0.6$	$M_1(\mathrm{b}) = 0.3$	$M_1(c) = 0.1$
$M_2(a) = 0.3$	$m_1(a) \times m_2(a) = 0.18$ $\{a\} \cap \{a\} = a$	$m_1(b) \times m_2(a) = 0.09$ $\{b\} \cap \{a\} = \phi$	$m_1(c) \times m_2(a) = 0.03$ $\{c\} \cap \{a\} = \phi$
$M_2(b) = 0.5$	$m_1(a) \times m_2(a) = 0.3$ $\{a\} \cap \{b\} = \phi$	$m_1(b) \times m_2(b) = 0.15$ $\{b\} \cap \{b\} = b$	$m_1(c) \times m_2(b) = 0.05$ $\{c\} \cap \{b\} = \phi$
$M_2(c) = 0.2$	$m_1(a) \times m_2(c) = 0.12$ $\{a\} \cap \{c\} = \phi$	$m_1(a) \times m_2(a) = 0.06$ $\{b\} \cap \{c\} = \phi$	$m_1(c) \times m_2(c) = 0.02$ $\{c\} \cap \{c\} = c$

In the first step, Table 2 was formed. The steps to do this have been completely explained by (Sentz and Ferson 2002); Moreover, one of the principles that need to be fulfilled in structuring a Bayesian network is to avoid the creation of any cycle.

According to the information represented in Table 2, the integrated mass probability for each relationship is calculated by using Eq. (2):

$$K = 0.09 + 0.03 + 0.3 + 0.05 + 0.12 + 0.06 = 0.65, 1 - k = 1\text{–}0.65 = 0.35$$
$$m_{1-2}(a) = 0.18/0.35 = 0.514$$
$$m_{1-2}(b) = 0.15/0.35 = 0.428$$
$$m_{1-2}(c) = 0.02/0.35 = 0.057$$

According to the results, since the amount of $(m_{1-2}(a))$ is greater than the other two values, therefore the final relationship between these two factors based on the opinions of two experts is: a = information flow → preparedness. Which means Information flow affects Preparedness and not the other way. Consequently, according to the information collected from three experts and by using the Dempster-Shafer theory, the Bayesian network was constructed as in Fig. 2.

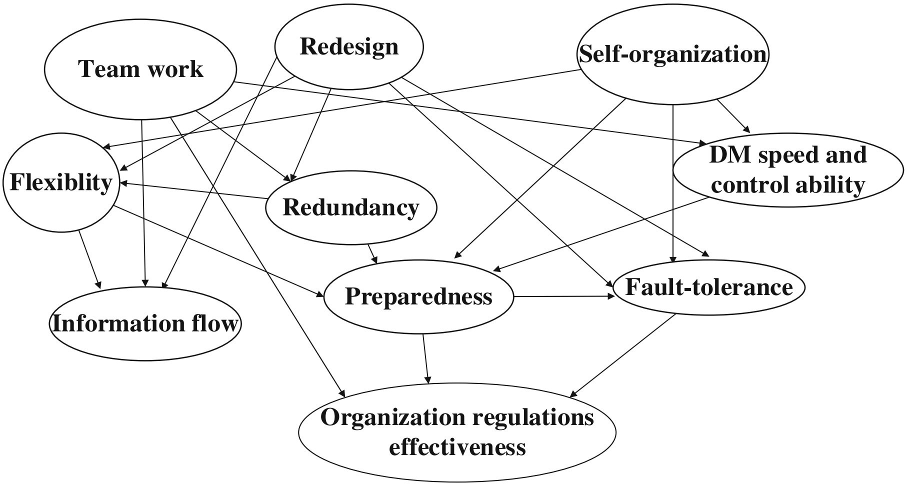


Fig. 2. The Bayesian network

3.6 Conditional Probabilities

Once the Bayesian Network is constructed, the conditional probability distribution for each node should be calculated. In this regard, two states namely poor and good have been defined for all of ten factors. As previously mentioned, in the range of 1 to 20, 1 indicates the worst and 20 indicates the best score. If the score assigned by the experts, ranges between 1 to 12, then the state for a factor is assumed poor and if the score is in the range between 12 to 20, then the factor is in a good state. Thus the conditional probability distribution for each node in the Bayesian Network is represented in a Conditional Probability Table (CPT). In a CPT, the state of each node is dependent and conditioned to its parent nodes' states while it is independent to other nodes. For example, the CPT for the redundancy node, with Teamwork node and Redesign node as its parents, is represented as follows (Table 3):

Table 3. CPT table for redundancy node.

Parents nodes		Child node	
		Redundancy	
Team work	Redesign	Poor	Good
Poor	Good		
Poor	Good		
Poor	Good		
Poor	Good		

3.7 Results

In the constructed BN, there are three root nodes including *teamwork*, *redesign* and *self-organization*. By changing these three factors, other factors including redundancy, information flow, fault tolerance, flexibility, Preparedness, DM speed and control ability and finally organizational regulations' effectiveness will be changed. Results of sensitivity analysis indicated that by a variation in each of the three root nodes, a significant change can be observed in other factors and finally on OREF. Moreover, based on the results, teamwork factor is more effective than redesign and self-organization. The effect of root node (teamwork, self-organization and redesign) on OREF is represented in Table 4.

Table 4. The obtained results which indicate the effect of teamwork, redesign and self-organization on organizational regulations effectiveness.

Root nodes						Organizational regulations effectiveness (OREF)	
Team work		Self-organization		Redesign			
Poor	Good	Poor	Good	Poor	Good	Poor	Good
0.5	0.5	0.5	0.5	0.5	0.5	0.45	0.55
0.5	0.5	0.5	0.5	0	1	0.3	0.7
0.5	0.5	0	1	0.5	0.5	0.35	0.65
0	1	0.5	0.5	0.5	0.5	0.11	0.89

The results that are represented in Table 4 shows the impact of three root nodes including teamwork, redesign and self-organization on OREF.

3.8 Validation

In this section, we analysis by which extend the change of teamwork, self-organizational and redesign which are root nodes in the BN affect other nodes. The results presented in the table showed that, with the increasing the 3 indicators, the other indices significantly increased, therefore, on this basis, the ORE and ME by increasing the ten indicators factors, has improved. This results are represented in Table 5. It is also obvious that OREF has been improved by 41%.

Table 5. The results of changes in factors according to suggested for improvement team work, self-organization and redesign

Factors	Factor average before of changes	Factor average after of changes	Percent improvement (100%)
Team work	12.3	16	70
Redesign	9	14	94
Self-organization	11.7	17	1
Redundancy	13.1	16	54
DM speed and control ability	15	16.8	33
Flexibility	10	14	75
Information flow	12.9	14.6	32
Preparedness	13.9	16	40
Fault-tolerance	14	17	56
Organizational regulations effectiveness	12.8	15	41

The results according to Table 5 demonstrate improvement in factors after changes in three root node factors that includes: teamwork, self-organization and redesign.

4 Conclusion

This paper aimed to optimize and improve organizational regulations effectiveness (OREF) through concurrent modelling of organizational resilience engineering (ORE) and macro-ergonomics (ME) indicators in an organization. Therefor six indicators, namely, teamwork, preparedness, fault tolerance, flexibility, redundancy, self-organization and organizational regulations effectiveness are considered for ORE. The macro-ergonomic indicators are redesign, decision-making speed and information flow. The combined factors are considered and analyzed through Bayesian Network (BN). Three factors that including teamwork, self-organization and redesign were known as the factors which have the greatest impact on OREF.

References

1. Qureshi, Z.H.: A review of accident modelling approaches for complex socio-technical systems. In: Proceedings of the Twelfth Australian Workshop on Safety Critical Systems and Software and Safety-Related Programmable Systems, vol. 86. Australian Computer Society, Inc. (2007)

2. French, S., Cope, J.: A review of human factors identified in investigations by Rail Accident Investigation Branch (RAIB). In: International Railway Safety Conference, London, UK (2012)
3. Kyriakidis, M., Happee, R., de Winter, J.C.: Public opinion on automated driving: results of an international questionnaire among 5000 respondents. Transp. Res. Part F Traffic Psychol. Behav. **32**, 127–140 (2015)
4. Madigan, S., et al.: Representational and Questionnaire Measures of Attachment: A Meta-Analysis of Relations to Child Internalizing and Externalizing Problems. American Psychological Association (2016)
5. Azadeh, A., Zarrin, M.: An intelligent framework for productivity assessment and analysis of human resource from resilience engineering, motivational factors, HSE and ergonomics perspectives. Saf. Sci. **89**, 55–71 (2016)
6. Pęciłło, M.: The resilience engineering concept in enterprises with and without occupational safety and health management systems. Safety Sci. **82**, 190–198 (2016)
7. Hollnagel, E., Resilience: The Challenge of the Unstable (2006)
8. Henry, D., Ramirez-Marquez, J.E.: Generic metrics and quantitative approaches for system resilience as a function of time. Reliab. Eng. Syst. Saf. **99**, 114–122 (2012)
9. Morel, G., Amalberti, R., Chauvin, C.: How good micro/macro ergonomics may improve resilience, but not necessarily safety. Saf. Sci. **47**(2), 285–294 (2009)
10. Woods, D.D., Hollnagel, E.: Prologue: resilience engineering concepts. In: Resilience Engineering: Concepts and Precepts, pp. 1–16 (2006)
11. Azadeh, A., et al.: Assessment of resilience engineering factors in high-risk environments by fuzzy cognitive maps: a petrochemical plant. Saf. Sci. **68**, 99–107 (2014)
12. Nielsen, T.D., Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer, New York (2009)
13. Neapolitan, R.E.: Learning Bayesian Networks, vol. 38. Pearson Prentice Hall, Upper Saddle River (2004)
14. Constantinou, A.C., et al.: From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. Artif. Intell. Med. **67**, 75–93 (2016)
15. Azadeh, A., et al.: An adaptive neural network algorithm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: the case of a gas refinery. J. Loss Prev. Process Ind. **24**(4), 361–370 (2011)
16. Cooper, D.R., Schindler, P.S., Sun, J.: Business Research Methods. McGraw-Hill, New York (2003)
17. Woodson, W.E., Tillman, B., Tillman, P.: Human Factors Design Handbook: Information and Guidelines for the Design of Systems, Facilities, Equipment, and Products for Human Use. McGraw-Hill, New York (1992)
18. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika **16**(3), 297–334 (1951)
19. Hsu, S.H., et al.: The influence of organizational factors on safety in Taiwanese high-risk industries. J. Loss Prev. Process Ind. **23**(5), 646–653 (2010)
20. Mohammadfam, I., et al.: Constructing a Bayesian network model for improving safety behavior of employees at workplaces. Appl. Ergon. **58**, 35–47 (2017)
21. Azadeh, A., et al.: An intelligent algorithm for performance evaluation of job stress and HSE factors in petrochemical plants with noise and uncertainty. J. Loss Prev. Process Ind. **26**(1), 140–152 (2013)
22. Azadeh, A., et al.: Improved prediction of mental workload versus HSE and ergonomics factors by an adaptive intelligent algorithm. Saf. Sci. **58**, 59–75 (2013)

23. Azadeh, A., et al.: A neuro-fuzzy algorithm for assessment of health, safety, environment and ergonomics in a large petrochemical plant. J. Loss Prev. Process Ind. **34**, 100–114 (2015)
24. Azadeh, A., et al.: Performance evaluation of integrated resilience engineering factors by data envelopment analysis: the case of a petrochemical plant. Process Saf. Environ. Prot. **92**(3), 231–241 (2014)

Behavioral Keys in Cryptography and Security Systems

Marek R. Ogiela[1(✉)] and Lidia Ogiela[2]

[1] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, 30 Mickiewicza Ave., 30-059 Krakow, Poland
mogiela@agh.edu.pl

[2] Cryptography and Cognitive Informatics Research Group, AGH University of Science and Technology, 30 Mickiewicza Ave., 30-059 Krakow, Poland
logiela@agh.edu.pl

Abstract. In this paper will be presented new ways of using some behavioral features and habits for security purposes and cryptography. In particular several different solutions will be described, which present possible application of selected behavioral patterns, characteristic for particular users. Such behavioral patterns can be extracted thanks to the application of new generation cognitive vision systems. Obtained personal feature, can be next use for security reasons, as well as in cognitive cryptographic protocols.

1 Introduction

In modern cryptographic technologies and security protocols an important role may play application of some, very specific and individual human characteristics, including behavioral features. In many situations we need to use some protocols, which should be oriented on particular user or be assigned to user of communication protocol. To create such human oriented cryptographic procedures we can use specific or very unique personal features, including behavioral patterns. Of course involving personal features into security procedures should allow providing required security level, and be connected with particular user. In this paper will be described selected approaches, based on application of personal and behavioral patterns, in creation of such security protocols. Extraction or registration some specific behavioral features are usually very difficult, but application of cognitive vision systems allow us to evaluate some common behavioral patterns, like hand or finger movements, specific human body motions or unusual personal gesture. Having such specific behavioral patterns or personal gestures, we can find many interesting application in creation protocols for securing strategic or classified data, guarantee secure transmission, remote services management or application in the Cloud environment and many others [1–3]. Everything shows that in near future such technologies may play important role in ambient cities, pervasive computing and IoT [4–6].

The main goal of this paper will be presentation of possible application of personal and behavioral features for security purposes and cryptographic protocols. Presented solutions will be based on evaluation of behavioral feature vectors, which next can be

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_26

involved into the encryption process [6, 7]. Such solutions are very promising for development of modern security technologies based on vision systems, which can monitor personal habits, movement characteristics or other human motion actions. Such new procedures may allow extending existing cryptographic methodologies towards a new branch of cognitive cryptography [8, 9].

2 Acquisition of Behavioral Patterns

For security application it is possible to use different motion and behavioral patterns. From our point of view, the most interesting is acquisition of nonstandard personal features in the form of behavioral patterns, which may be applicable for security purposes. Conducted research showed that for such purposes are very useful cognitive vision systems [5, 10].

In general cognitive systems try to imitate the mental functions and thinking processes, and are based on one of the known model of human visual perception, named knowledge based perception [4, 11]. In this model cognitive resonance processes are performed, which allow to understand observed situation or patterns, and allow evaluating the semantic meaning. Using cognitive information systems we tried to show some new opportunities in application of personal and behavioral features in new security protocols and cryptographic procedures. Such features may be used in creation of strong encryption keys or creation of a special kind i.e. personalized behavioral lock. It is worth underline that creation of such cryptographic solutions is inspired by biological models and finally allows to create a new branch called cognitive cryptography defined in [7, 12]. This new computational paradigm combines techniques, which are using to guarantee data confidentiality, with some personal or behavioral information describing particular user and extracted using cognitive information systems. Such compilation of security approaches with cognitive processes may be very promising for future development of security solutions.

Acquisition of personal features or behavioral characteristics can be performed with application of cognitive vision systems in connection with other motion sensor devices like Leap Motion technologies, Kinect or MoCap devices.

We can consider following different behavioral patterns, which can be used for security application:

- Simple finger or palm movements. For registration of such simple patterns we can use Leap Motion or Kinect devices. Personal feature vector may be next created using cognitive vision systems, which allow extracting informative parameters registered by these recording devices.
- More complex motion patterns performed using hands or other human body parts may be analyses by advanced motion capture devices connected with cognitive vision systems.
- The most complex patterns (specific exercises, sport techniques) may be analyzed using dedicated recording equipment with cognitive features.

Presented methodology allows analyzing different types of behavioral patterns, and create personal feature vector, which next may be used in creation personalized security solutions. Among the most important areas of application of behavioral features for security purposes we can consider:

- Visual cryptography and secret sharing protocols.
- Personally oriented information management procedures.
- Fuzzy vault and steganography which depend on personal feature [2].

3 Hand Gesture Features for Security Applications

For security applications we can consider different specific behavioral patterns, also these connected with gesture or movement activities.

Application of motion sensors or Kinect devices, which allow trace and register finger or palm movements, allow also to extract some specific motion features and use them for security purposes. The simplest solution is using hand or single finger movements or gesture for personal authentication procedures, during which we can extract very specific personal features, which next may be used as behavioral lock or authentication keys. Finger motion analysis seems to be one of the most natural and simplest, so it may be focused on tracing a fingertip positions changes during making particular gesture. Analysis of simple gestures has many advantages like noninvasive data acquisition, possibility of analysis of different gesture types like user determined, fixed movement patterns, or natural gesture. It allows considering also the motion dynamics and acceleration of performed gesture, and may be performed in real-time.

In real security applications we should consider only enough complex gesture or movements, which will be sufficient for obtaining the distinctive features for particular user, which are not similar to other persons. To register such complex patterns we can analyze movements performed using one or more fingertips or, if necessary, for using the palms or hands.

To register unique personal gesture features we can use any motion sensor (Kinect, MoCap etc.), which is able to register hand position changes during the time. As a result we can create a personal motion feature vector, which contain very unique parameters describing performed gesture like direction, velocity and acceleration.

In conducted research we have implemented such procedures, which involve cognitive systems for hand and finger analysis. In the first stage of such analysis it is necessary to create a learning set, which contain particular number of well-defined motion patterns. The second stage allows classifying a new pattern by comparing it with elements stored in learning set. Such comparison is made by calculation any similarity measure between new registered motion pattern and elements stored in learning set. Base on a feature vector and classification function, it is possible to determine the type of hand or finger movement, but also it is possible to use such feature vector as a specific behavioral key or in other security applications like personalized encryption key generation, multi secret steganography, fuzzy vault etc. The security of such behavioral keys is strongly depended on complexity of analyzed movements, so for stronger cryptographic solutions should be considered more complex or very specific gestures.

4 Complex Movements for Security Applications

Beside simple behavioral features describing hand or finger movements it is also possible to consider more complex individual movements performed while walking, dancing or practicing sport. In general analysis of such movements is more difficult and should be connected with registration of motion sequences, but in some special cases it may also be done towards extraction of personal feature vector for security application. The reason is very simple, and connected with different abilities of performing very complex movement by different persons, which are strongly dependent on personal habits, physical conditions, age, and skills. For example persons acknowledgeable with infrequent dances can perform it in his personal manner, and such performance allows evaluating some specific personal features, which may be next used for security purposed.

In the same way we can consider some sport and gym activities and extract specific features from recording presenting very high sport skills or techniques e.g. acrobatics, martial arts etc. (Fig. 1).

Fig. 1. Extraction of personal features from sport and acrobatic techniques.

Analysis of several sport activities towards evaluating very specific motion feature can be done in the same way as in [12]. Having such specific motion features we can use them for security purposes in the same manner as described previously with connection to finger and palm gestures.

5 Conclusions

In this paper we have presented different possibilities of using personal features and unique characteristics for security application and cryptography. Personal information extracted from motion sequences and representing behavioral features may be used in security protocols dedicated for information sharing, fuzzy vaults and multi-secret steganography. Besides traditional simple finger and palm movements, for security application we can consider more complex gesture patterns representing personal behavioral features. Extraction of personal or behavioral characteristics is possible thanks to the application of cognitive information systems, which allow evaluate unique parameters from nonstandard personal patterns or specific human body movements. Application of such specific and unique parameters proved that personal features and cognitive systems can be used in development of advanced cryptography procedures for strong key generation, secret management, visual cryptography, and creation of behavioral lock.

Acknowledgments. This work has been supported by the AGH University of Science and Technology research Grant No 11.11.120.329.

References

1. Cox, I.J., Miller, M.L., Bloom, J., Fridrich, J., Kalker, J.: Digital Watermarking and Steganography. Morgan Kaufmann Publishers, Burlington (2008)
2. Jin, Z., Teoh, A.B.J., Goi, B.-M., Tay, Y.-H.: Biometric cryptosystems: a new biometric key binding and its implementation for fingerprint minutiae-based representation. Pattern Recogn. **56**, 50–62 (2016)
3. Ogiela, L.: Computational intelligence in cognitive healthcare information systems. Stud. Comput. Intell. **309**, 347–369 (2010)
4. Ogiela, L.: Cognitive informatics in image semantics description, identification and automatic pattern understanding. Neurocomputing **122**, 58–69 (2013)
5. Ogiela, L., Ogiela, M.R.: Beginnings of cognitive science. In: Advances in Cognitive Information Systems. Cognitive Systems Monographs, vol. 17, pp. 1–18 (2012)
6. Ogiela, L., Ogiela, M.R.: Data mining and semantic inference in cognitive systems. In: Xhafa, F., Barolli, L., Palmieri, F., et al. (eds.) 2014 International Conference on Intelligent Networking and Collaborative Systems (IEEE INCoS 2014), Salerno, Italy, pp. 257–261, 10–12 September 2014
7. Ogiela, L., Ogiela, M.R.: Management information systems. In: Park, J., Pan, Y., Chao, H.C., Yi, G. (eds.) Ubiquitous Computing Application and Wireless Sensor. LNEE, vol. 331, pp. 449–456. Springer, Dordrecht (2015)
8. Ogiela, M.R., Ogiela, U.: Linguistic approach to cryptographic data sharing. In: FGCN 2008 – The 2nd International Conference on Future Generation Communication and Networking, Hainan Island, China, vol. 1, pp. 377–380, 13–15 December 2008
9. Ogiela, M.R., Ogiela, U.: Security of linguistic threshold schemes in multimedia systems. Stud. Comput. Intell. **226**, 13–20 (2009)
10. Ogiela, M.R., Ogiela, U.: Shadow generation protocol in linguistic threshold schemes. In: Ślęzak, D., Kim, T., Fang, W.C., Arnett, K.P. (eds.) Security Technology. CCIS, vol. 58, pp. 35–42. Springer, Heidelberg (2009)
11. Ogiela, M.R., Ogiela, U.: Grammar encoding in DNA-like secret sharing infrastructure. In: Kim, T.-h., Adeli, H. (eds.) ACN/AST/ISA/UCMA -2010. LNCS, vol. 6059, pp. 175–182. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13577-4_15
12. Ogiela, M.R., Ogiela, U.: Secure Information Management Using Linguistic Threshold Approach. Advanced Information and Knowledge Processing. Springer, London (2014)

Enhancing Ant Brood Clustering with Adaptive Radius of Perception and Non-parametric Estimation on Multi-core Architectures

Mohmmed Qasem, Ying Ying Liu, Ziyue Wang, Parimala Thulasiraman(✉), and Ruppa K. Thulasiram

University of Manitoba, Winnipeg, MB, Canada
thulasir@cs.umanitoba.ca
http://www.cs.umanitoba.ca/∼thulasir

Abstract. Clustering is an important problem in complex networks. Exact algorithmic approaches to clustering is not affordable for many real world instances, requiring innovative, approximation algorithms. Among them are meta-heuristics such as nature-inspired techniques. One of the existing techniques inspired by real ants in nature, is called *ant brood clustering algorithm* (ACA). In this paper, we present Ant Clustering Algorithm with Adaptive Radius (ACA-AR). Unlike existing ACA Models, ACA-AR utilizes Kernel Density Estimation (KDE) to measure average dissimilarity of data objects in ant's neighborhood, and it allows ants to adapt the radius of perception so they can avoid the convergence to a local-optimum. We also present a parallel counterpart of the algorithm on the Graphics Processing Unit (GPU) using NVIDIA CUDA and on multi-core CPU cores using OpenMP. Our results on benchmark datasets show that ACA-AR gains substantial clustering accuracy, and the parallel version executes up to 39x faster whilst preserving the quality of the retrieved clusters.

Keywords: Ant clustering · GPU · CUDA · OpenMP · Adaptive radius · Kernel density estimation

1 Introduction

Bio-logically inspired algorithms such as ant colonies, bird flocking, animal herding, bacterial growth, and fish schooling are inspired by animal behavior. They are suitable for designing parallel and distributed algorithms because, the organisms (agents) modeled in the system, work independently (parallelism) but co-operate/co-ordinate through stigmergic communication (distributed and minimize global communication)[1], self-organize (fault tolerant) and work asynchronously (reduce fine-grained synchronization) within their local environments (locality). Clustering is an important problem in complex networks. Exact algorithmic approaches to clustering is not affordable for many real world instances.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_27

In the literature, ant colony optimization (ACO) has been used to solve the clustering problem in community detection [2]. However, the algorithm does not work well for large networks. In this paper, we consider another ant-based algorithm called *ant brood clustering* (ACA)[3] algorithm. Unlike the traditional clustering methods, ACA does not make any prior assumptions on the number or the shape of the clusters. This algorithm is therefore, suitable for dynamic networks [4].

1.1 Ant Brood Clustering Algorithm (ACA)

Ant brood clustering was inspired from the observation that some species of ants have the ability to sort large corpses and eggs into clusters. This behavior was modeled by Deneubourg et al. [5] to enable real-world robots perform certain clustering tasks. In the basic model, the data objects are scattered on a two-dimensional grid and the ants randomly move withing this grid. Each ant decides to either *pick up* a free object or *drop off* a loaded object depending on the average similarity of the object with its neighboring objects.

The basic model is enhanced by Lumer and Faieta [3] to work with multidimensional data that are comparable according to a measure of dissimilarity (LF Model). The average dissimilarity is estimated using a local density function as shown in Eq. (1). In this function, $d(o_i, o_j)$ is a similarity function that measures the distance between the object o_i and a neighboring object o_j (e.g. Euclidean distance or Cosine) scaled by a constant $\alpha \in [0, 1]$. r is the radius of perception of the ant, and $(2r)^2$ is the area of ant's neighborhood, which is a normalization factor that represents the square area surrounding the object o_i. As shown in the pickup Eq. (2), the more dissimilar objects there are in the ant neighborhood, the more likely the objects will be picked up. Conversely, the more similar objects there are in the ant neighborhood, the more likely the objects will be dropped off, as given in Eq. (3). k_p and k_d are two parameters for adjusting pickup and drop-off probabilities, respectively. After several iterations, data clusters emerge from the collective and collaborative activities of the ants.

The LF model is considered as a standard ant clustering algorithm in mining data. In 2002, Handl and Meyer implemented ACA for classifying on-line documents based on their cosine similarity [6]. Wu et al. [7] combined ACA with k-means algorithm to achieve more accurate clustering results.

$$f(o_i) = \min\left(\frac{1}{(2r)^2} \sum_{o_j \in N(o_i)} \left(1 - \frac{d(o_i, o_j)}{\alpha}\right), 0\right) \tag{1}$$

$$P_{pick}(o_i) = \left(\frac{k_p}{k_p + f(o_i)}\right)^2 \tag{2}$$

$$P_{drop}(o_i) = \left(\frac{f(o_i)}{k_d + f(o_i)}\right)^2 \tag{3}$$

1.2 Shortcomings of LF Model

The LF model suffers from three shortcomings: (i) three initial parameters (α, k_p, k_d) in Eqs. (1), (2) and (3) have to be "experimentally" fine-tuned; (ii) the ant's radius of perception is constant - narrowing an ant's visibility, consequently, converging to a local optimum; (iii) no communication between ants prevents the ants from dropping the object in the "best" location. As a result, ants spend lots of time randomly moving on the grid until they find a location that satisfies the object dropping criteria which is computationally intensive and sometime redundant.

Contribution: We propose three major modification to the LF model to alleviate the aforementioned shortcomings and present a parallel implementation of the modified algorithm on multicore architectures.

2 Enhancements to Ant Brood Clustering

2.1 Applying Kernel Density Estimation to the Ant Neighborhood

Equation (1) estimates the average similarity of ant's neighborhood. We replace this by Kernel Density Estimation (KDE), a non-parametric statistical model that estimates the probability density function of a random variable. KDE is also known as the Parzen-Rosenblatt window method. Given the kernel function, K and bandwidth parameter, h, KDE estimates the probability density of a particular object o_i with respect to each neighboring object $o_j \in N(o_i)$, as shown in Eq. (4). Given an object o_i at grid location (x, y), we define $N(o_i)$ as the set that includes n objects within the area $[x \pm r, y \pm r]$. Aggregating these probability densities gives an overall picture of the underlying structure of the data and its density function.

$$KDE(o_i) = \frac{1}{nh^d} \sum_{o_j \in N(o_i)} K\left(\frac{||o_i - o_j||^2}{h}\right) \tag{4}$$

Using KDE in ACA is more convenient than the average similarity function (Eq. (1)) not only because it eliminates the need for the initial parameters, but also because it gives the choice of using different kernel types (e.g. Gaussian, Linear, Polynomial, etc.) for different clustering purposes. In this work, we adopt using Gaussian kernel $K(o_i, o_j) = exp(\frac{-||o_i - o_j||^2}{2\sigma^2})$. Substituting the Gaussian kernel in (4) gives:

$$KDE(o_i) = \frac{1}{n(\sigma\sqrt{2\pi})^d} \sum_{o_j \in N(o_i)} e^{\frac{-||o_i - o_j||^2}{2\sigma^2}} \tag{5}$$

where the standard deviation parameter σ works as the bandwidth parameter, h. If an ant decides to pick up or drop off an object o_i, it computes $KDE(o_i)$ to determine whether the objects at the neighborhood of o_i are actually neighbors to that object. $KDE \in [0,1]$, with 1 being the maximum similarity. The pickup and drop-off probabilities are then computed by substituting the KDE value into the logistic function as shown in Eqs. 6 and 7. In both equations, c is an integer constant that controls the convergence of algorithm. The lower the KDE value, the greater the probability of picking up an object o_i and the higher the KDE values, the greater the probability of dropping off an object o_i (Fig. 1).

$$Pr_{pickup}(o_i) = 1 - \frac{1 - exp(-cKDE(o_i))}{1 + exp(-cKDE(o_i))} \tag{6}$$

$$Pr_{drop}(o_i) = \frac{1 - exp(-cKDE(o_i))}{1 + exp(-cKDE(o_i))} \tag{7}$$

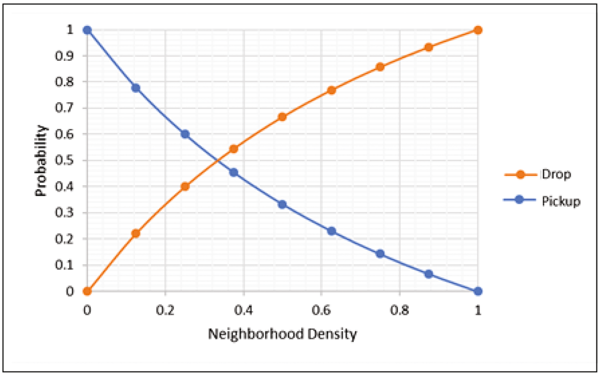


Fig. 1. Pickup and drop-off probabilities using Gaussian KDE for neighborhood density

2.2 Adaptive Radius-Based Neighborhood Function

The ant radius of perception (r) determines the area in which the ant can explore the nearby objects in order to decide the appropriate action. The ant's action can be either picking up a free object available at its current position on the grid, dropping off a loaded object if the ant position is empty, or moving to another random grid position. Such area is determined by a neighborhood function that can be as simple as a radius-based area or as complex as copula. Regardless of the complexity of this neighborhood function, however, the accuracy of ACA is highly sensitive to the value of ant radius of perception. This is because the latter significantly affects KDE, which in turn, affects the convergence rate and the number of the final clusters. For instance, setting the radius to a low value (e.g. $r = 1, r = 2$) results in forming a larger number of (more accurate) clusters than what originally exists in the data. In contrast, high radius values (eg. $r \geq \frac{1}{4}\sqrt{grid\ area}$) yields a smaller number of clusters than what originally exists in the data. To identify the exact number of clusters in a given dataset, we implement ACA with adaptive ant radius where each ant can increase or decrease its radius of perception each time it becomes unable to perform pickups in a certain range of iterations.

In more detail, initially, we set the radius of all the ants as $r = 1$. The ants randomly walk on the grid until they are completely unable to pickup objects. At this point, the outcome will be a large number of small dense clusters (local-minima solution). To avoid such convergence to a local minima, each ant gradually increases its radius of perception by one. Increasing ant's radius increases the probability of picking up because the ant can recognize more dissimilar objects in its larger neighborhood. The radius increment occurs each time the ant becomes unable to perform pickups until the radius reaches a maximum threshold. When the radius threshold is reached, the process is reversed (i.e. the radius is gradually decreased by one). The gradual decrement of ant radius results in compacting the clusters and hence increasing their spatial locality. The ant is terminated when the value its radius reaches becomes zero.

Applying ACA with adaptive radius of perception has many advantages over ACA with constant radius: (i) it allows the ants to search a significantly larger number of candidate clustering solutions and avoid local-optimum solution; (ii) it is more capable of detecting data outliers and finding the exact number of clusters in the data; (iii) it substantially improves the spatial separation of clusters on the grid which is an essential requirement to retrieve the clusters.

2.3 Termination Condition

In the existing ACA models, the algorithm is terminated when the ants reach a predefined number of maximum steps (iterations). This number is determined by trial and error. In this work, we formulate the termination condition based on the number of pickups that are accomplished by an ant on a certain range of steps. If this number is zero, the ant increases its radius of perception by one as long as the radius is smaller than a given threshold. When the ant's radius reaches the maximum threshold, the ant decreases its radius by one till it reaches zero. The gradual decrease of the radius results in compacting the clusters so the spatial separation among them increases. The ant is terminated when its radius is zero.

A high-level description of ACA including the above modifications is illustrated in Algorithm 1.

3 Parallel Implementation

The algorithm is implemented on the multicore GPU ans CPU machines using CUDA and OpenMP, respectively. On the GPU, the algorithm starts from the CPU. The CPU reads in the data file and creates two arrays: ants and objects, which stores their location on the grid. The grid is created as an array (integer) initialized to -1. Then, the ants and objects are randomly placed on the grid. The arrays and grid are moved from the CPU to the GPU global memory to

Algorithm 1. ACA with Adaptive Radius of Perception

Input: feature vectors $\{x^i\}_{i=1}^n$ $x^i \in R^m$ N_{ants} $grid_{(h\times w)}$ max radius r M_{steps}
Output: grid coordinates $\forall x^i \in \{x^i\}_{i=1}^n$

$\forall x^i \in \{x^i\}_{i=1}^n$, assign x^i to a random grid location
$\forall$ $ant_i \in N_{ants}$, assign ant_i to a random grid location
for $ant \in N_{ants}$ in parallel **do** {main loop}
  $ant.radius \leftarrow 1$ {initialize ant radius of perception}
  while $ant.radius > 0$ **do**
    $ant.pickups \leftarrow 0$ {this is to count the number of pickups that are performed successfully by the ant}
    for $step \in \{1, ..., M_{steps}\}$ **do**
      if $ant.location \neq null$ **then** {there is object at ant's location}
        if *ant is unladen* **then** {ant does not carry object}
          obj ← grid(ant.location) {get the object using ant position}
          draw a random number $R \in [0, 1]$
          compute $P_{pickup}(obj)$, Eq. (6)
          if $P_{pickup}(obj) \geq R$ **then** {perform pick up}
            ant.obj = grid(ant.location) {ant carries object}
            grid(ant.location) ← null {remove object from grid}
            $ant.pickups \leftarrow ant.pickups + 1$
          end if
        else{ant is loaded with object so it continue wandering}
          move ant to a random grid location
          continue {go to while loop}
        end if
      else{there is no object at ant's location}
        if *ant is carrying object* **then**
          obj← ant.carrying
          draw a random number $R \in [0, 1]$
          compute $P_{drop}(obj)$, Eq. (7)
          if $P_{drop}(obj) \geq R$ **then** {perform drop off}
            grid(ant.location) ← ant.obj {store ant's object in the grid}
            ant.obj = null {ant becomes unladen}
          end if
        else{ant is unladen, and grid location is empty so ant continue wandering}
          move ant to a random grid location
          continue {go to while loop}
        end if
      end if
    end for
    increase ant radius by 1 as long as ant radius does not reach max radius r
    decrease ant radius by 1 if ant radius r reaches max radius r
  end while
end for
$\forall x^i \in \{x^i\}_{i=1}^n$ retrieve grid coordinates of x^i
identify clusters using cluster retrieval algorithm 2

Algorithm 2. Cluster Retrieval

```
Input: grid coordinates ∀x^i ∈ {x^i}_{i=1}^n
Output: cluster label l_i ∀x^i ∈ {x^i}_{i=1}^n
initialize undirected graph G(V, E) V = {x^i}_{i=1}^n
for x_i ∈ {x_i}_{i=1}^n do
    for x_j ∈ neighbors(x_i) do
        E = E ∪ {(x_i, x_j)}
    end for
end for
identify connected components in G using depth or breadth first search
assign a unique cluster label for each subset of vertices that is identified as con-
nected component
```

avoid communication latency between the CPU and GPU. The CPU also calculates the distances between each objects using the Euclidean distance (in Eq. 4) which is also passed to the GPU. Algorithm 1 is executed on the GPU. Each ant is a block in the GPU. We restricted each block to one thread since there was no performance gain when more ants were added, creating unnecessary synchronization latency and race conditions. The computations of each ant, a_i, is very fine-grained. There are three operations performed by each ant: pick-up, drop-off and move. Each ant can be categorized as either loaded or not loaded. If the ant is not loaded, the ant considers its current location: (i) if empty it moves randomly to the adjacent cells; (ii) otherwise, the ant applies Eq. 6 to determine whether to pick up the object. The number computed by Eq. 6 is compared to a random number (*rand*) between 0 and 1 generated by the ant. If the pick-up probability determined by the Equation is larger than *rand*, the ant picks up the object and continues to wander, otherwise, it moves to another grid location. On the other hand, if the ant is loaded, the ant calculates the next drop- off probability using Eq. 7. Again, a random number is generated and compared to the drop-off probability. If the drop-off probability is greater than the generated random number the ant drops the object. This implies that the objects are all similar. Otherwise, the ant moves to another adjacent cell. Note that, only one ant may be located in each grid location. The termination condition is the total number of iterations. The ants radius of perception is increased to cover up to 1/4th of the grid area. Each ant starts with radius one. This results in forming a large number of small dense clusters because ants become unable to perform pick-ups. To merge the small clusters into larger dense ones, the ant gradually increases its radius of perception to recognize more dissimilarity, which, in turn stimulates pick-ups. When the ant covers up to 1/4th of grid area, it gradually decreases its radius to compact the large clusters. The final result is sent back from the GPU to the CPU. The algorithm is very similar on the CPU, with the exception of sending data between CPU and GPU.

4 Cluster Retrieval

When the ACA converges to a solution, a cluster retrieval algorithm is needed to assign a cluster label to each data point. The grid is actually edgeless (torus) which means that data points at the edges are adjacent. To retrieve the clusters, we construct an undirected graph $G(V, E)$ such that $V = \{x^i \in x^n_{i=1}\}$. To set the connectivity of G, each data point x^i at position (x, y) is connected to its neighbors at positions $(x \pm 1, y \pm 1)$. As a result, the data points that form a blob on the edgeless grid form a connected component in G.

Formally, let $G = (V, E)$ be an undirected graph. Given a set A such that $A \subset V$, and its complement $\overline{A} = V \backslash A$. A is *connected* if for any pair of vertices $(v_i, v_j) \in A$, there exists a path that connects v_i and v_j, and all intermediate vertices that lies in the path are in A. A is a *connected component* if and only if A is connected and there are no connections between A and $\overline{A}$.

To label the clusters, we perform connected components labeling in G which is straightforward and can be computed using either breadth-first search or depth-first search in linear time. Algorithm 2 illustrates the cluster retrieval process.

5 Evaluation

We evaluate the clustering quality of ACA-AR on three benchmark datasets [8] as listed in Table 1: Iris, Yeast, and a subset of the 20-Newsgroups.

Table 1. Datasets

Dataset	Instances	Features	Classes	Distance function
Iris	150	3	3	Euclidean
Yeast	1484	8	10	Euclidean
20-Newsgroups	1500	5000	5	Cosine

To evaluate the different clustering quality of ACA with adaptive radius, we adopt two external evaluation metrics: V-measure [9] and Adjusted Random Index (ARI) [10] as well as one evaluation internal metric which is the Silhouette Coefficient (SC) [11] We compare the clustering quality of ACA-AR with the two benchmark clustering algorithms: K-Means (KM) and Mean Shift (MS) [12].

5.1 ACA with Adaptive Radius

Table 1 shows the comparative results of Iris clustering solutions obtained by K-Means, Mean Shift, ACA with fixed radius, and ACA-AR. As the metrics indicate, ACA-AR outperforms all other clustering algorithms in identifying the Iris clusters. Mean Shift identifies only 2 out of the 3 existing clusters because it fails to separate the overlapped Iris classes. ACA and ACCA-AR, by contrast, identify all of the three classes. In addition, ACA-AR gains 18% increase in ARI, as indicated by its V-measure score (82%), compare to ACA. This increase is due to the improvement of the completeness and homogeneity of clusters.

Table 2. Cluster validity - Iris

	KM	MS	ACA	ACA-AR
V-measure	0.66	0.73	0.64	0.82
ARI	0.62	0.57	0.62	0.71
SC	0.46	0.58	0.43	0.54
Number of clusters	3.00	2.00	3.00	3.00

Table 3. Cluster validity - Yeast

	KM	MS	ACA	ACA-AR
V-measure	0.53	0.73	0.77	0.82
ARI	0.28	0.57	0.65	0.69
SC	0.35	0.58	0.55	0.57
Number of clusters	10.00	2.00	3.00	5.00

We also evaluate the clustering quality of ACA-AR on the Yeast dataset, which contains a larger number of classes (10) with highly unbalanced class distribution (463, 429, 244, 163, 51, 44, 37, 30, 20, 5). Table 2 shows the experimental results on this dataset. k-means converges to a local-minima solution with ADI of 28% due to the poor selection of initial centroids. Although Mean Shift achieves higher scores for ARI and V-measure, it fails to detect classes with small sizes. ACA and ACA-AR, by contrast, outperform Mean Shift because they are more capable of detecting the small classes.

Clustering 20-Newsgroups documents is challenging because of the "curse of dimensionality", which the result of the fact documents are represented using sparse high-dimensional feature vectors (e.g. 10000 features). To obtain the results using the cosine distance, we use the Spherical k-means, which is a k-means variant for text clustering. To handle the cosine similarity in MS, ACA and ACCA-AR, We adapt Eq. (4) to $N_h(o_i) = \{o_j : ||cosine(\boldsymbol{o_i}, \boldsymbol{o_j})|| \geq h\}$ as proposed by [13]. Reviewing the results in Table 3, KM obtains the worst results because of class non-linearity and initial selection of cluster centroids. Both MS and ACA identify only three out of the five existing clusters because they fail to separate some of the overlapped classes. ACCA-AR, by contrast, identify all of the five classes (Table 4).

Table 4. Cluster validity - 20-Newsgroups

	KM	MS	ACA	ACA-AR
V-measure	0.21	0.57	0.55	0.61
ARI	0.18	0.49	0.55	0.59
SC	0.04	0.51	0.55	0.57
Number of clusters	5.00	3.00	3.00	5.00

5.2 Parallel Implementation Results

Although all ants wander simultaneously on the grid, only those ants with non-overlapped neighborhoods can compute a pick-up or drop-off probability at the same time. This is because each ant considers its neighborhood as a critical section when it computes a pick-up or a drop-off probability. To highlight the impact of such synchronization mechanism on the performance, we report speedups of both parallel implementation strategies (OpenMP and GPU) by varying radius of perception from 1 to 6. In all experiments, we take the number of ants (threads) as 1% of data size N. Figure 2 shows speedup results for OpenMP implementation. Speedup increases with respect to the data size for small values of radius of perception ($r = 1, 2$). This result is due to the fact that larger number of threads can be executed simultaneously. However, speedup decreases rapidly as the radius increases because of race condition. In addition, the increase in radius results in collecting a larger data sample within ant's neighborhood. This, in turn, makes computing KDE, pick-up and drop-off more computationally-intensive.

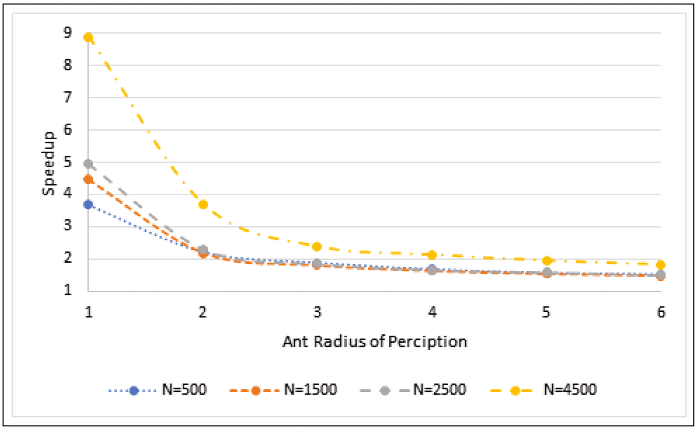


Fig. 2. OpenMP-based results

Unlike OpenMP results, The gained speedup on GPU implementation increases up to 39x compared to the sequential implementation as shown in Fig. 3. Since the most expensive computation in our algorithm is due to the KDE, we consider a parallelization strategy that offload KDE computation by each ant to block-level at GPU. In GPU implementation, each ant is mapped to a block so that independent ants can run at the same time. Since the major task of each ant is to compute KDE using the objects within its neighborhood, we partition ant task into finer sub-tasks that handle KDE computation (i.e., summation of kernels of probability density function of for an object x_i). Notice that a significant speedup can be gained when KDE is implemented in a data-level parallel manner in each GPU block. This is because sequential KDE computation for n data objects of dimension d requires $O(n^2d)$ [14]. In addition to that, parallel implementation of multivariate KDE can entirely get the benefit of data-level parallelism because it's is embarrassingly parallelizable (straightforward vectorization). This parallel strategy reduces the impact of ant synchronization mechanism, especially for ants with large radius of perception. To reduce latency of global memory access, each ant transfers the data object to be picked-up or dropped-off to the register level, and the rest of the objects within its neighborhood to the shared memory level. This significantly reduces memory global accesses because such objects are the most frequently accessed data object in KDE computation. This strategy enable us to obtain a scalable GPU-based implementation with an increase in speedup up to 39%.

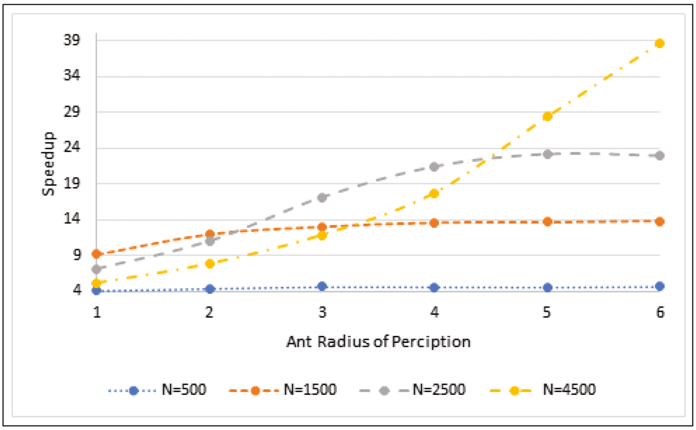


Fig. 3. GPU based results

6 Conclusion

In this paper, we presented ant brood clustering algorithm with adaptive radius (ACA-AR), a new variant of ACA that involves two major enhancements. ACA-AR employs Kernel density estimation and Sigmoid function to estimate ant's pick-up and drop-off probabilities. It also allows ants to adjust radius of perception to improve ants' pick-up capability. We have experimentally evaluated ACA-AR on three benchmark data sets that present different clustering challenges. Our results have shown that ACA-AR outperforms ACA, Mean Shift and k-means algorithms in terms of clustering accuracy, completeness, and homogeneity. Our results also show that a speedup of up to 39x can be obtained compared to the sequential counterpart using GPU. Extending ACA-AR to a semi-supervised clustering model and applying ACA-AR to industrial applications such as Sentiment Analysis is the primary avenue for future research.

References

1. Navlakha, S., Bar-Joseph, Z.: Distributed information processing in biological and computational systems. Commun. ACM **58**(1), 94–102 (2015)
2. Honghao, C., Zuren, F., Zhigang, R.: Community detection using ant colony optimization. In: CEC, pp. 3072–3078. IEEE (2013)
3. Lumer, E.D., Faieta, B.: Diversity and adaptation in populations of clustering ants. In: Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3: From Animals to Animats 3, pp. 501–508. MIT Press (1994)
4. Wang, J., Osagie, E., Thulasiraman, P., Thulasiram, R.: HOPNET: a hybrid ant colony optimization routing algorithm for mobile ad hoc network. Ad Hoc Netw. **7**(4), 690–705 (2009)
5. Deneubourg, J.L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., Chrétien, L.: The dynamics of collective sorting robot-like ants and ant-like robots. In: Proceedings of the First International Conference on Simulation of Adaptive Behavior on From Animals to Animats, pp. 356–363 (1991)
6. Handl, J., Meyer, B.: Improved ant-based clustering and sorting in a document retrieval interface. In: Guervós, J.J.M., Adamidis, P., Beyer, H.-G., Schwefel, H.-P., Fernández-Villacañas, J.-L. (eds.) PPSN 2002. LNCS, vol. 2439, pp. 913–923. Springer, Heidelberg (2002). doi:10.1007/3-540-45712-7_88
7. Wu, B., Shi, Z.Z.: An ant colony algorithm based partition algorithm for TSP. Chin. J. Comput.-Chin. Edit. **24**(12), 1328–1333 (2001)
8. http://archive.ics.uci.edu/ml/
9. Hirschberg, J.B., Rosenberg, A.: V-measure: a conditional entropy-based external cluster evaluation. In: Proceedings of EMNLP (2007)
10. Hubert, L., Arabie, P.: Comparing partitions. J. Classif. **2**(1), 193–218 (1985)
11. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
12. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory **21**(1), 32–40 (1975)
13. Senoussaoui, M., Kenny, P., Dumouchel, P., Stafylakis, T.: Efficient iterative mean shift based cosine dissimilarity for multi-recording speaker clustering. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7712–7715. IEEE (2013)
14. Michailidis, P.D., Margaritis, K.G.: Accelerating Kernel density estimation on the GPU using the CUDA framework. Appl. Math. Sci. **7**(30), 1447–1476 (2013)

Designing an Efficient Self-adaptive Parallel Algorithm Using Math Oracles

Gabriel Luque(✉) and Enrique Alba

Departamento de Lenguajes y Ciencias de la Computación, Andalucía Tech,
Universidad de Málaga, Málaga, Spain
{gabriel,eat}@lcc.uma.es

Abstract. In this paper we present a methodology to develop self-* methods at a low computational cost. Instead of going purely ad-hoc we define several simple steps to include theoretical models as additional information in our algorithm. Our idea is to incorporate the predictive information (future behavior) provided by well-known mathematical models or other prediction systems (the *oracle*) to build enhanced methods. We show the main steps which should be considered to include this new kind of information in any algorithm. We actually test the idea on a specific algorithm, a parallel genetic algorithm (GA). Experiments show that our proposal is able to obtain similar, or even better, results when it is compared to the traditional algorithm. We also show the benefits in terms of saving time and a complexity of parameter settings.

1 Introduction

The development of self-* algorithms which adapt their behavior to the specific characteristics of the problem is currently a very hot topic [1]. One of the goals of these self-* methods is to ease the utilization of the proposed algorithms by unspecialized users by freeing them from the need of having additional knowledge due to the ability of the technique to self-adjust its parameters online.

In a revision of the state of the art, we can find, on the one hand, that most of the proposed techniques in this domain change their dynamics (modifying the variation operators in some way, mainly) using the current status of the method and some historical data [2]. On the other hand, there exists a number of mathematical studies on convergence [3], run-time analysis [4], or landscape analysis [5] which could provide us some information about the expected future behavior of the method. Using this kind of theoretical information in practice is our key approach here to decide how to self-adapt the algorithm. However, most mathematical models make several assumptions that cannot be met in real scenarios where the method is to be applied. Therefore, we need some guidelines about how to use this information in our algorithms when solving actual problems.

We define a mathematical oracle as a mathematical description of the behavior of a search technique that allows a prediction on the near future steps of the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_28

technique. We assume that this prediction is done in the oracle in an approximate way, what needs continuous corrections during the search of a solution. We also assume that this oracle can be used in a black-box form, so that we input in it some data on the past search and get from it the expected behavior.

The main goal of this paper is to propose a methodology to use the information provided by a mathematical tool allowing to predict (in an approximate way) the behavior of the search, what will allow us to build enhanced versions of them. Later, we show an illustrative case study in which we show how this methodology can be used to design a new method.

The contributions of this work are: (i) a new way of using theory in practical design of algorithms (ii) an efficient modification of existing metaheuristics to convert them into self-* techniques (iii) a source for new research lines that builds self-* that are actually numerically competitive, and not just "good ideas".

This paper is organized as follows. Section 2 presents the methodology to include the information provided by the mathematical oracle in any optimization algorithm. In Sect. 3, we show an example of applying this methodology to design a self-adaptive genetic algorithm. Later, we describe the benchmark used to study the behavior of this theoretically guided self-* algorithm (Sect. 4) and analyze the results in Sect. 5. In the last section we summarize the conclusions and give some hints on the future work.

2 A Methodology to Use Mathematical Oracles

A general optimization technique is an iterated process that finishes when it meets a stopping criterion. The input received by the algorithm is the parameter settings and the information of the problem to solve. The output of the optimization procedure is the best solution ever found. This process can be represented by the pseudo-code in Algorithm 1.

Data: Problem data and algorithm settings
Result: best solution ever found
algorithm initialization;
while *termination condition is not met* **do**
  perform a step of the algorithm;
end

Algorithm 1: General scheme for a general iterative optimization technique

As we said in the previous section, a *mathematical oracles* is (for us) a theoretical tool (usually a mathematical model) which is able to provide some knowledge about the behavior of the method in the future steps using the current status and other historical data of the technique. For example, a theoretical model for takeover time [3] can be considered as such a mathematical tool: it receives the information on the evolution of the current population and returns the expected convergence time (see Subsect. 3.2). From the point of view of our methodology, this mathematical oracle will be considered as a black box system, and it is not necessary to know the concrete details about how it works (see Fig. 1).

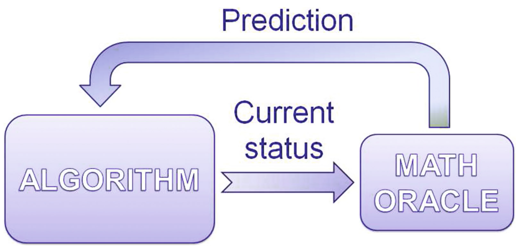


Fig. 1. Interaction between the algorithm and the mathematical oracle

Another important clarification is about the precision of the prediction of these mathematical tools. Most probably, any useful theoretical tool to be used as an oracle will make several assumptions that cannot be met in real scenarios where the method is to be applied, and therefore, we have to do some corrections on the oracle in order to get it nearer the actual behavior of the algorithm. Of course, the better the kind of prediction of the oracle the better the expected results. In general, we will assume that an iterative on-line learning is also performed in the oracle to adjust its own predictions.

Data: Problem data and algorithm settings
Result: best solution found ever
algorithm initialization;
info = $\emptyset$;
while *termination condition is not met* **do**
  perform a step of the algorithm;
  info = info $\cup$ current status;
  if *condition to use mathematical oracle is met* **then**
    predictive info = mathematical_oracle (info);
    if *modification condition is reached* **then**
      apply actions according the predictive information;
    end
  end
end

Algorithm 2: General scheme for building self-adaptive techniques with an oracle

After this first description of the components of the self-* technique, we present our proposed methodology in Algorithm 2 in which we have included some new phases (not included in the base algorithm) to consider the information coming from the mathematical oracle. The main decisions are:

1. *Offering to the oracle*: what information is required by the oracle and how it is gathered.
2. *Praying the oracle*: when are we going to apply the mathematical oracle.
3. *Oracle's response*: what information is provided as output by the oracle.
4. *Pray and work*: according to that predictive information, when and how are we going to modify the dynamics of the technique.

In the next paragraphs, we will go explaining these decisions.

Offering to the Oracle
First, we have to gather the input data latter offered to fed to the oracle. This information depends on the mathematical oracle which we are using. Generally, it is related to the evolution of the search such as improvement speed, number of steps performed, ..., or in the case of population-based techniques some statistical indicator of the population (such as diversity, entropy, standard deviation, ...). In some cases, the exact information required by the oracle cannot be obtained since the actual combination of algorithm-plus-problem does not fulfill all the assumptions of the mathematical model. Then, if not exactly, we have to provide some information similar to the expected one. We will see an example of this case in the Sect. 3.

Praying the Oracle
Now, since we have the data needed by the mathematical oracle (got in the previous phase), we could apply it to obtain the new predictive information. The next decision is when the mathematical oracle is to be applied, e.g. we could apply it in every step of the algorithm. However, this might not be an adequate value since it is needed some algorithmic steps to see if the last change is positive or not. Also, some mathematical oracles need to gather the information of several steps to provide accurate results. This value can be defined as the number of algorithmic steps between two consecutive applications of the oracle or we could use a more complex condition according to some properties of the search procedure.

Oracle's Response
When we use the mathematical oracle, we get some prediction about the expected behavior of the technique in the next steps (expected convergence, diversity, movement in search space, ...). As we said in the first paragraphs of this section, this information has some inaccuracy since our algorithm does not fulfill the assumptions required by the mathematical model and it should be used carefully. In fact, we recommend to use this value (the oracle's response) as a trend information since the exact value maybe could be inaccurate.

Pray and Work
Finally, maybe the most important issue is how to use this information to change our algorithms. This part is very related to the algorithm used and the kind of information provided by the oracle. Specific mathematical models from theory in metaheuristics can be used in practice to provide some relation between the expected behavior and same parameters of the methods (probability of mutation, selection pressure, ...), and therefore, the way of performing the change is quite straightforward [6]. But this is not the general case, and usually we will have arbitrary potential oracles which provide information only about the convergence time, the improvement speed, or similar. We have to decide if this trend is the adequate one for the future run of our algorithm and otherwise, take some actions (change parameters or operators of the method systematically) in order to approach this prediction to the expected one.

In the next section, we detail all the steps for a concrete example, showing several alternatives to deal with each decision.

3 Case Study: A Self-adaptive Parallel Genetic Algorithm

This section shows how this general methodology can be instantiated for a concrete algorithm and a specific mathematical oracle. We design a family of self-adaptive distributed genetic algorithms using some predictive information coming from a mathematical model of takeover time. We first give a brief background of the distributed Genetic Algorithms (Subsect. 3.1) and about the mathematical oracle used (Subsect. 3.2). Later, in Subsect. 3.3 we describe how to develop a new enhanced method.

3.1 Distributed Genetic Algorithms

A genetic algorithm (GA) is a population based technique. A fitness function assigns a value to each individual indicating its suitability to the problem. This method iteratively applies stochastic operators such as selection, crossover and mutation on a pool of tentative solutions (population) to find a satisfactory optimal solution to the problem. In distributed GAs (dGAs), the population is structured into smaller subpopulations relatively isolated one from the others. The principal feature of this kind of algorithm is that copies of individuals within a particular sub-population (island) can occasionally migrate to another one [7].

The dGA model requires the identification of a suitable migration policy, and this is often done at hand based in the experience of the researcher and running a set of preliminary experiments whose overhead (time, effort, and the knowledge gained in it) is often not reported in scientific articles. The main parameters of the migration policy include the following ones: *migration period* (determines how many iterations occur between consecutive migrations), *migration size* (defines the number of solutions that migrate), *selection/replacement* (decides how to select emigrant solutions, and which solutions have to be replaced by the immigrants), and *topology* (defines the neighbor of each island, i.e., the islands that a concrete sub-population can send individuals to, or receive from.

In general, as we said before, the choices of a migration policy are made by preliminary experimental studies. The idea of this work is the utilization of the results of some theoretical studies (see next subsection) which characterize the parameters to self-tuning some of them. In particular, in this paper we focus on the migration period, which is considered the most important one.

3.2 Mathematical Oracle: Takeover Time

A common analytical approach to study the selection pressure of an EA is to characterize its takeover time [3], i.e., the number of generations it takes for the best individual in the initial population to fill the entire population under selection only. The growth curves are another important mathematical tool to analyze the dynamics of population-based methods. These growth curves are functions that associate the number of generations of the algorithm with the proportion of the best individual in the whole population. Now we give a mathematical definition of these concepts. Let us start by formally defining what the growth curve is.

Definition (Growth Curve). Given a population-based algorithm, under selection only, with an initial population containing exactly one individual of the best fitness class; *growth curve* is defined as a function $P_{sel} : \mathbb{N} \rightarrow [0,1]$ that maps the proportion of copies of the best individual in the population to each generation step. ■

As a result of applying selection only in an population-based method (without variation operators), at every generation the number of copies of the best individual potentially grows up. The number of generations it takes for the algorithm to completely fill the population is what is called the takeover time, formally defined as follows:

Definition (Takeover Time). Let P_{sel} be the function defining the growth curve induced by a selection method, the value $t_{sel} = min\{t : P_{sel}(t) = 1\}$ is called the *takeover time* associated to the given selection method. ■

Several models have been proposed to estimate the takeover time and growth curves for canonical and structured GAs in the past. In this paper we use a model presented in [8] for $(\mu + \mu)-$dEA, where μ is the total population size:

$$t^* = period \cdot d(\triangle) - \frac{1}{b} \cdot ln \left(\frac{1}{a} \cdot \frac{\varepsilon}{N - d(\triangle) - \varepsilon \cdot N} \right), \tag{1}$$

where t^* is the takeover time, *period* is the migration period, N is the number of islands, $\triangle$ is the topology, $d(\triangle)$ is the diameter of the topology (length of the longest shortest path between any two islands), ε is the expected level of accuracy (a small value near to zero) and a and b are adjustable parameters.

3.3 Designing a Self-adaptive GA

In this subsection, we are going to show how the previous mathematical model can be used as an oracle inside a distributed genetic algorithm. We are going to propose several alternatives to each step of this methodology to illustrate the many interesting future research lines that opens from our idea.

Offering to the Oracle

Our theoretical oracle can estimate some values related to the convergence of the dGA. To calculate these values it needs the proportion of the optimal solution ($P(t)$) in the populations of the last steps of the algorithm. Here, we face the first challenge, since this value is always 0 (a real method has got no copies of the optimal value until maybe its last steps). In order to use $P(t)$ in an actual algorithm we propose three alternative definitions of this concept.

The first proposal to obtain a usable value of $P(t)$ is to calculate the proportion of all values greater than $\alpha \cdot f_{optimum}$ in generation t:

$$P(t) = \frac{\sum_{i=1}^{N} \sum_{f \geq f'} n_f^i(t)}{\mu}, \tag{2}$$

where $f^{'} = \alpha \cdot f_{optimum}$ for some $\alpha \in (0..1]$, and the $n_f^i(t)$ indicates the number of solutions with fitness value equal or higher than f in the island i at generation t. Clearly, for $\alpha = 1$ we obtain the original expression of $P(t)$. We shall refer to this expression as the αOpt.

The second proposal, which we call *bestSF*, is based on the best solution found in the current generation ($f_{bestfound}$). This proposal consists in calculating the proportion of individuals with this fitness value at generation t:

$$P(t) = \frac{\sum_{i=1}^{N} n_{f_{bestfound}}^i(t)}{\mu}. \tag{3}$$

The third proposal is based on the diversity of the population. To do this we consider the standard deviation of the fitness values at generation t:

$$\sigma(t) = \sqrt{\frac{1}{\mu} \cdot \sum_{i=1}^{\mu} \left(\bar{f}(t) - f_i(t)\right)^2}, \tag{4}$$

where $f_i(t)$, with $i = 1 \ldots \mu$, are all the fitness values of μ individuals at generation t and $\bar{f}(t)$ is the mean of these values. Then we can calculate $P(t)$ as:

$$P(t) = \max\left(1 - \frac{\sigma(t)}{\sigma(0)}, 0\right), \tag{5}$$

where $\sigma(0)$ is the standard deviation of de initial population fitness distribution. We shall refer to this variant as *stdDev*.

Our last proposal is one that was used in [6] and is also based on diversity. In this case the diversity is calculated doing some modifications to the expression of the standard deviation (4). This is done introducing a threshold fitness value (f^{ϕ}), so that all fitness values under f^{ϕ} are considered equal to it. Besides, the standard deviation calculation will be made with respect to the objective fitness value instead of the mean:

$$\sigma_{f_{objective}, f^{\phi}} = \sqrt{\frac{1}{\mu} \cdot \sum_{i=1}^{\mu} (f_{objective} - \max(f_i, f^{\phi}))^2}, \tag{6}$$

then the expression of $P(t)$ is thus (model *stdDevObj*):

$$P(t) = \left(1 - \frac{\sigma_{f_{objective}, f^{\phi}}}{f_{objective} - f^{\phi}}\right). \tag{7}$$

Praying the Oracle
As we will see in the *Pray and work* step, the main goal of this new self-* algorithm is adapting the period to the current evolution of the population. The oracle only needs information on the generations since the last migration on how may "best" individuals existed in them (a cheap counting at every generation) to calculate the actual takeover time.

Oracle's Response
In the previous section we stated that the oracle can facilitate two types of information: the actual takeover time and the theoretical expected takeover time. When a difference between these values is detected (see Fig. 2), the algorithm adapts its parameters (switches the algorithm configuration).

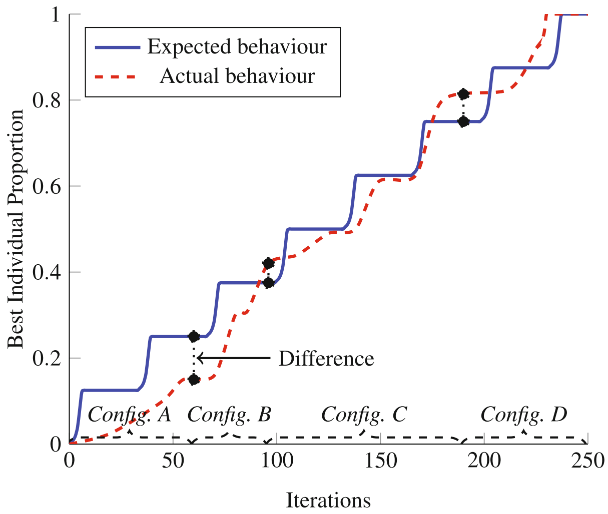


Fig. 2. Theoretical convergence model used to self-adapt the configuration of our algorithm during the search.

Pray and Work
Let us see in detail how to the migration period is adjusted. The Eq. 1 provides, given a migration policy, the takeover time. We have proceeded in the opposite direction: given a target takeover time (the complete execution time), how do we set our migration schedule in order to achieve that the algorithm will converge towards the execution end? The proposed method to self-adjust the migration period follows the next equation:

$$period = \frac{t_{remaining} + K}{d(\triangle) - \left(\frac{P(t)}{1/N}\right)}, \text{ where}$$
$$K = \frac{1}{b} \cdot \ln\left(\frac{1}{a} \cdot \frac{\varepsilon}{N - d(\triangle) - \varepsilon \cdot N}\right), \tag{8}$$

and $P(t)$ is the proportion of the best individual at the generation t, $t_{remaining}$ is the number of remaining iterations for the end of execution, a is equal to the size of a sub-population hosted by an island (μ/N), $b = 0.4$ and the tolerance parameter $\varepsilon = 0.1$. This mathematical expression is derived from the takeover time equation (Eq. 1) which was validated to real cases in [8]. Here we will use it as an oracle telling us what the best migration period is to be used for the next steps of the distributed algorithm.

4 Empirical Validation

To test our self-adaptive techniques we chose the well-known problem MAX-SAT problem [9]. This problem consists in finding an assignment for a set of variables which maximizes the number of clauses satisfied on a given Boolean formula. In the experiments we used nine different instances of this problem with $l = 50$, 75, 100, 125, 150, 175, 200, 225, and 250 variables. These instances are in the phase transition (difficult ones, although we are not seeking at solving them, but at to show that our self-* approach works in practice), where the ratio between the number of clauses and the number of variables is approximately 4.3 [10].

In these experiments we used a total population of 400 individuals ($\mu = 400$), distributed in 8 islands ($N = 8$) each of which hosted a population of 50 individuals. We use a directed ring topology, so that $d(\triangle) = N - 1$. Our migration policy only sends a single solution in each exchange, and the selections for migration will be elitist: the best solution of the source island is transmitted, the worst solution of the receiving island is always replaced. We test a wide set of values for the migration period ranging from 1 (constant communication among islands) to 5000 (complete isolation). Each result was obtained from conducting of 50 independent executions, in order to obtain reliable statistical results. The Table 1 shows the values of configuration parameters for the dGA. Two different configurations for the selection parameters have been studied since they have a high influence on the intensity and diversity. The first one (*Random* selection) emphasizes diversity, combining a random selection of the parents and

Table 1. Set of configuration parameters for the dGA

Parameter	Value
Number of iterations	5000
Population size	400
Mutation type	Bit–flip
Mutation probability	$1/l$
Crossover type	Two–points
Crossover probability	0.6
Selection	Random/binary tournament
Replacement	Binary tournament/elitist

a binary-tournament selection of the next generation. The second one (*Elitist* selection) has a special stress on intensification, combining binary-tournament for the parents, and elitist selection of the next generation.

5 Analysis of the Results

In order to analyze whether the results are statistically reliable, we applied the Wilcoxon test [11], with a confidence level of 95%. Table 2 shows the results of applying the Wilcoxon test confronting each of our algorithm variant (different way to calculate $P(t)$) against the 180 different configurations of the traditional dGA (9 instances $\times$ 10 standard migration periods $\times$ 2 selection procedures). The column ▲ indicates the number of times that the our techniques produced statistically better results, column $-$ indicates the number of times that there were no statistically significant differences, and column $\triangledown$ indicates the number of times that the our method produced significantly worse results compared against the executions with constant migration period. The columns tagged with $\%>$ and $\%\geq$ contain the percentage of times that our self-* model improved and improved or equaled the results with permanent migration periods, respectively. Finally, the $\%<$ indicates the percentage in which a fixed value of the migration period parameter is better than our proposal.

Table 2. Results of applying the Wilcoxon test

	▲	$-$	$\triangledown$	$\%>$	$\%\geq$	$\%<$
$\alpha_1 Opt$	49	130	1	27.22	**99.44**	**0.66**
$\alpha_{0.985} Opt$	38	135	7	21.11	96.11	3.88
$\alpha_{0.95} Opt$	23	123	34	12.78	81.11	18.88
bestSF	47	132	1	26.11	**99.44**	**0.66**
stdDev	55	123	2	**30.56**	98.89	1.11
stdDevObj	42	129	9	23.33	95.00	5.00

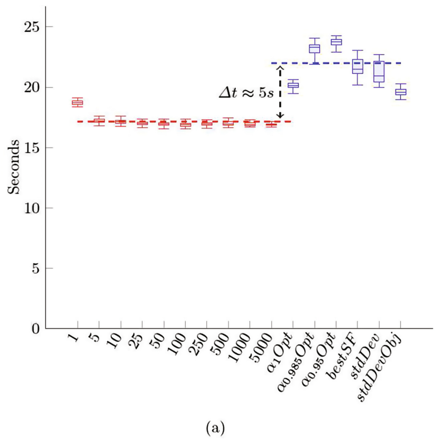
Several conclusions can be obtained from Table 2. First, $\alpha_1 Opt$ and *bestSF* variants are the best ones. They obtain very accurate results: they allow to improve the results of handmade tuning (done by a researcher) in 55 scenarios out of the 180 testing ones, and they are only worse in a single one. These two models share the same feature: they are very strict about what solutions are considered in the calculation of $P(t)$, which is very beneficial for this problem. In fact, when the number of solutions considered as "good" ones by the model is increased (i.e., it increases the value of $P(t)$ quite fast), the result of the model is worse. This can be observed when we change α value in αOpt model: the lower the α value is (and then $P(t)$ grows faster), the worse the results are.

The algorithms based on a diversity metric, *stdDev* and *stdDevObj*, also obtain quite accurate results but their accuracy is sightly worse than $\alpha_1 Opt$ and *bestSF* variants. In any case, all our self-adapt alternatives (with the exception $\alpha_{0.95}Opt$) achieved better results than the dGA (the values of the column %> is significantly greater than the ones of the column %<).

Now, we focus on analyzing the overhead provoked by the calculations needed by the different strategies to their self-adaptation. The use of self-tuning technique means an increase in the computation time of the algorithm. This is because the overhead of sending all fitness values, after each migration to the master process for the calculation of the new migration period.

Figure 3a shows a boxplot graph with runtimes of all executions carried out to the 250 variables instance using Elitist selection. As it can be easily seen, the executions with constant migration periods usually take about 17 s, while using our adaptive migration takes between 20 and 24 s. This represents an increase of between 3 and 7 s (depending of the model used) in the computation time for each run in case of using the adaptive technique. Although this sounds a bit negative at a first glance, this increase can be offset by the global time saved in finding an appropriate configuration for the migration period.

The benefit of using the adaptive technique is showed in Fig. 3b. In that figure, we can see that the time required by the different proposed method against the time required to test different values for constant periods. If we only test one constant period, it can be observed that our models represents an small overhead, but this situation is not realistic, since we need to analyze several periods to get the best one, as usually done by researchers. When two or more values are analyzed, our proposal allows to reduce the computation time significantly. In fact, this reduction is around 86% when 10 different values are tested. This means that using our algorithms is not only interesting because they are theoretically grounded and provide better results, but that the overall time that a researcher needs to employ for his/her scientific analysis is globally reduced in a meaningful manner. This is good



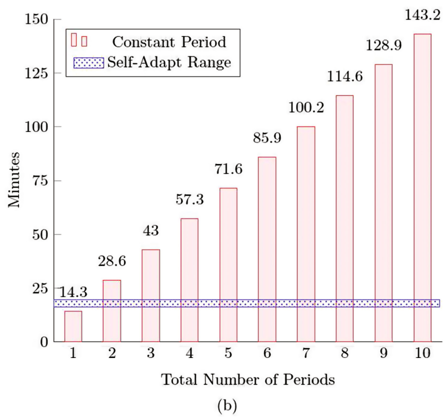


Fig. 3. Execution time: (a) Boxplot with the runtimes of the dGAs, and (b) the global computation time using different numbers of migration periods.

from different points of views: larger instances can be solved in the same time, wider analysis could be included in future papers, or even higher productivity in publication in conferences and journals since results are got in a reduced time.

6 Conclusions

In this work, we have presented a general methodology to include a mathematical prediction model inside a regular algorithm to create a self-*. This idea is general, accurate, efficient, and based in (present or future) theory on EAs or metaheuristics in general. This last creates a new bridge between theory and practice, much needed long time ago in this field.

As a example, we have shown here how a very simple mathematical (takeover time analysis) can be used as an oracle to improve the behavior of a dGA. We tested the proposal in a wide set of instances of a hard optimization task (MAXSAT problem). The results showed that the resulting self-adaptive algorithms can outperform the classical ones, with a negligible overhead.

As future work, we plan to apply this methodology to design new general self-* algorithms which will be able to outperform state of the art algorithms in different problems. Basically, we will start to apply more sophisticated theoretical tools as the oracle; our expectations are high since the used in this paper is really simple and still proved to be competitive at negligible costs. We will also try to keep using a rigorous statistical analysis, several problems in the validation phase, and an explicit overhead discussion. All these practices are very necessary in this domain of research but unfortunately still not in use (all together) in a single paper. This is also a future consideration to do in everyone's work with self-* techniques.

Acknowledgements. Authors acknowledge funding from the Spanish Ministry of Sciences and Innovation European FEDER under contract TIN2014-57341-R (MoveOn project available at URL http://moveon.lcc.uma.es).

References

1. Burke, E., Kendall, G., Newall, J., Hart, E., Ross, P., Schulenburg, S.: Hyper-heuristics: an emerging direction in modern search technology. In: Glover, F., Kochenberger, G.A. (eds.) Handbook of Metaheuristics, pp. 457–474. Springer, Boston (2003)
2. Burke, E.K., Curtois, T., Hyde, M., Kendall, G., Ochoa, G., Petrovic, S., Vazquez-Rodriguez, J.A., Gendreau, M.: Iterated local search vs. hyper-heuristics: towards general-purpose search algorithms. In: IEEE Congress on Evolutionary Computation (CEC 2010), pp. 3073–3080 (2010)
3. Goldberg, D.E., Deb, K.: A comparative analysis of selection schemes used in genetic algorithms. In: Rawlins, G.J. (ed.) Foundations of Genetic Algorithms, pp. 69–93. Morgan Kaufmann, San Mateo (1991)
4. Jansen, T., Wegener, I.: The analysis of evolutionary algorithms-a proof that crossover really can help. Algorithmica **34**(1), 47–66 (2002)

5. Chicano, F., Whitley, L.D., Alba, E.: A methodology to find the elementary landscape decomposition of combinatorial optimization problems. Evol. Comput. **19**(4), 597–637 (2011)
6. Osorio, K., Luque, G., Alba, E.: Distributed evolutionary algorithms with adaptive migration period. In: 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 259–264, November 2011
7. Alba, E., Tomassini, M.: Parallelism and evolutionary algorithms. IEEE Trans. Evol. Comput. **6**, 443–462 (2002)
8. Alba, E., Luque, G.: Theoretical models of selection pressure for dEAs: topology influence. In: The 2005 IEEE Congress on Evolutionary Computation, 2005, vol. 1, pp. 214– 221 (2005)
9. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman & Co., New York (1990)
10. Cheeseman, P., Kanefsky, B., Taylor, W.M.: Where the Really Hard Problems Are, pp. 331–337. Morgan Kaufmann, San Francisco (1991)
11. Demsar, J.: Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. **5**, 1–30 (2006)

Wind Speed Prediction with Genetic Algorithm

Michal Prilepok(✉)

FEECS, Department of Computer Science, VŠB - Technical University of Ostrava, 17. Listopadu 15/2172, 708 33 Ostrava - Poruba, Czech Republic
michal.prilepok@vsb.cz

Abstract. Nowadays trends pay attention to used renewable energy sources, e.g. wind – wind energy or sun irradiance – solar energy, as a source of electrical power. This kind of energy sources are very unstable and inconstancy (nonstationary) over the time. The proper and accurate wind speed or sun irradiance prediction is necessary to control the power grid. This paper presents short time wind prediction algorithm with genetic column subset selection problem. It uses multiple weather data sources, genetics algorithm for features selection, and the prediction is done by a neural network. The genetic algorithm chooses the most important features for the prediction algorithm.

1 Introduction

Nowadays renewable energies sources like as wind energy and solar energy are used to generate electrical power. These energy sources are nonstable and inconstancy (nonstationary) over the time. The generated power depends on the current weather conditions. The weather prediction play the big role in this issue. In on grid system, an electric power grid consists of a steady power source, such as thermal power plant, and a renewable energy source. The renewable energy source generates electricity only when the weather conditions are positive. Therefore the whole generated power by grid must be consumed by users economically. It means that the generated energy must be equal to the consumed energy approximately. For this reason, we must know how much energy will be generated by renewable sources. These sources are unstable, so the accurate prediction plays a key role in the estimation of a stable power production source.

In the present time the wind and solar energy source are widely used around the world. These energy sources are clean and free energy sources. This energy cannot be drained, is renewed by nature. Renewable energy plants supply and contribute the power grid. It reduces environmental pollution. The simple scheme of on grid power system is shown in Fig. 1. It contains both stable and renewable sources of energy – wind and solar.

A wind turbine is simple electrical generator which converts the wind power into electrical power. A solar photovoltaic panel converts sunlight into electricity which independent on light and radiance of the sun. In some specific cases, a

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_29

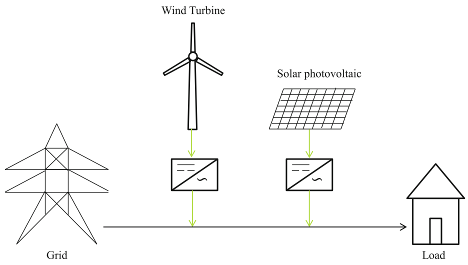


Fig. 1. A simple diagram of the power grid with renewable energy sources.

customer connection to the power grid is expensive. Therefore we use renewable energy plants separately, and it is called off grid system.

The paper is organized as follows. In Sect. 2, we discuss the previous studies in solar and wind energy prediction. The details of the proposed are described in Sect. 3. The Sect. 4 describe the used weather data and experiment. Section 5 evaluates the obtained results, compares model performance. The last Sect. 6 concludes this study.

2 Related Works

In the last years there have published several studies focused in the field of unstable energy source and weather prediction. As an unstable energy source can be considered any kind of power energy which depends on the nature, for example solar, wind, and water energy. The proper prediction plays a big role in the power grid management and control. The presented related works are divided into two main groups. The first group deals with prediction of input power plant variables - weather data, such as wind speed and direction or global irradiation. The second group of works focuses on the output power plant variables such as photovoltaic power prediction.

2.1 Wind Speed Prediction

To predict the wind speed researches have been used wide range of methods based on artificial neural networks (ANN), support vector machines (SVM), and genetic algorithms (GA). Zhao et al. applied SVM for wind speed prediction in [1]. The predicted result of the proposed model outperforms backpropagation neural networks. The SVM had the minimal value of mean absolute error and mean square error in comparison to backpropagation neural networks. Bhaskar et al. in [2] presented a prediction model for wind generated power based two stages. The first stage uses wavelet decomposition and adaptive wavelet neural network (AWNN) to predict wind speed. The second stage utilizes a feedforward neural network (FFNN) to convert wind speed into wind generated power. The results of the predicted wind power confirmed the efficiency of proposed method. Liu et al. [3]

have proposed a hybrid wind speed prediction model. The model combines wavelet transform (WT), SVM and GA. WT decomposes the original wind speed signal - measured wind speeds, GA evaluates and adjusts the optimal weights of SVM, and SVM predicts wind speed. The presented method has been compared with another method such as SMV with GA and was more accurate for wind speed prediction. Azad et al. [4] combined statistical model with a ANN to predict hourly wind speed in long-term. They compared hybridization of different optimization approaches. The results demonstrate that the proposed model overcomed the other compared prediction models for long-term wind speed prediction. By comparing the actual and predicted WSD, it can be seen that the hybrid technique can follow actual series closely. Wanga et al. [5] combined Empirical mode decomposition (EMD) with Elman neural network (ENN) for wind speed prediction. Te authors compared the presented model with the persistent model, backpropagation neural network, and ENN. The simulation results show that the proposed EMD-ENN model consistently has the minimum statistical error regarding the mean absolute error, mean square error, and mean absolute percentage error. In the next publication [6] the authors Wang et al. designed hybrid system to predict wind speed. This model is constructed of improved EMD and Genetic Algorithm-BP neural network. The simulation results demonstrate that the designed system was better than standard GA-BP neural network presented in previous publication. It shows that the proposed method based on hybrid EMD and GA-BP neural network performs well in wind speed prediction, and is suitable for ultra-short term (10 min) and short-term (1 h) wind speed prediction. A short-term wind speed prediction model for intervals between 1-hour and 5 h based on wavelet packet decomposition, crisscross optimization algorithm, and ANN was proposed by Meng et al. in [7]. Wavelet decomposing used to decompose wind speed, and ANN optimized by crisscross optimization algorithm used for predict wind speed. The proposed system achieved minimal mean absolute percentage error when it was compared with other hybrid methods.

2.2 Photovoltaic Power Prediction

The researcher applied many methods for photovoltaic power prediction, for example support vector machines and regression, fuzzy logic, and neutral networks. Shi et al. [8] proposed power output prediction algorithm for photovoltaic systems based upon weather classification and SVM. The weather conditions are divided into four types. Four SVM models are set up according to SVM algorithm. The obtained results show promising application in prediction of photovoltaic power output. Xu et al. [9] used weighted support vector machine (WSVM) to predict short-term photovoltaic power plant output. The simulation results of their model show the accuracy of the model is better than model based on ANN and is more practicable. Mandal et al. [10] combined WT with radial basis function neural network (RBFNN) to predict power output of photovoltaic based on irradiance and temperature. The experiments results proved the accuracy and efficiency of the proposed model in comparison to other tested alternatives. Prokop et al. [11] utilized genetic programming with Fuzzy Logic in they proposed method.

The main goal of the propsoed method was to predict the predict power output of a photovoltaic power plant. The proposed method was evaluated on they own solar data collected from the Czech Republic. The authors mentioned that the result for time ahead prediction was reassured. Zeng et al. [12] used Least Square SVM for solar prediction based atmospheric weather data - humidity, wind speed, and sky cover. The simulation results show the proposed model was better than others such as Autoregressive (AR) model and Radial Basis Function Neural Network (RBFNN) model. Prediction of energy production of a photovoltaic power plant for short times, e.g. 15 min, 1 h and 24 h ahead averaged power output PV power plant was presented by Li et al. [13]. The authors are using AAN and support vector regression (SVR). The proposed approach has been evaluated using statistical errors. The simulation results showed the proposed model exceed other compared methods. Dolara et al. [14] proposed Physical Hybrid Artificial Neural Network (PHANN) for ahead predicting of the photovoltaic system output. The obtained results of the proposed PHANN were compared with standard ANN which. The results proved the accuracy of proposed method than ANN. Teo et al. [15] utilized ANN with Extreme Learning training algorithm to predict the photovoltaic power output. The experimental results on various simulation showed that the proposed system with logistic function could predict power photovoltaic output with high efficiency.

3 Proposed Method

The proposed method is based on work Kromer el al. [16]. The authors proposed a column subset selection problem to chose the most significant features from the input data and to reduce the dimension of the data. After reducing the dimensionality of the data a ANN is train to predict the wind speed in short time. In this paper a short time period is considered ad a time interval of 24, 36, and 48 h.

3.1 Column Subset Selection Problem

The Kromer's genetic algorithm for the column subset selection problem (CSSP) is used for the feature selection of the input data. The input data are organized in a matrix $A^{m \times n}$, so that the m represents the number of input data samples, and n is the number of features in the dataset. The goal of CSSP is to select k columns so that $k < n$. It creates a new matrix $C \in R^{m \times k}$ such that the

$$||A - P_C A||_\epsilon \tag{1}$$

is minimized over all possible $\binom{n}{k}$ column choices for the matrix C. Here, $P_C = CC^\dagger$ denotes the projection onto the k-dimensional space spanned by the columns of C, $\epsilon = 2, F$ denotes the spectral norm and Frobenius norm respectively [17], and $C^\dagger$ stands for the Moore-Penrose pseudoinverse [18] of C. The goal of CSSP is to find the best subset containing exactly k columns selected from all n columns of A which have a minimum error.

In our experiment, we utilized this algorithm to select the best combination of input features - weather variables from neighboring airports. The input data are organized in a matrix, where the columns represent weather data from selected locations and rows measured samples in same time.

Genetic algorithms (GA) form a family of well known population-based meta-heuristic soft optimization methods [19]. GAs solve complex optimization problems by the programmatic evolution of an encoded population of candidate problem solutions. The solutions are ranked using a problem-specific fitness function.

The GA input data are encoded in a binary vector V - chromosome. The V length is equal to the number of columns of the matrix A. Each item v_i of V can have values 1, this means that the column is present in the output solution of GA, or 0, the columns will be skipped in the output solution of GA.

The algorithm uses 1-point crossover and uniform mutation operations. The crossover operation selects a random position i in parent chromosomes c_1 and c_2, which is inspected for suitability to be a crossover point. This operation crossover genes of selected chromosomes. The new chromosome consists of genes of c_1 on position from first position to $i-1$ position and genes from c_2 on positions i to kth position. The operation of mutation replaces ith gene c_i in chromosome c with an inverse value. If the $c_i = 0$ than the new value of $c_i = 1$ and otherwise.

The GA operations are applied on all data columns. For more data description see Sect. 4. In same cases can occur a situation in which some weather data from the wind power plant will be skipped. The newly selected combination of input variables is used to learn Backpropagation neural network as a prediction model of the wind speed.

3.2 Artificial Neural Networks

Artificial neural network (ANN) is a computation process which tries to imitate biological nervous systems that can learn from examples. ANN is constructed of a large number of neuron connected in a way to solve a specific problem such as pattern recognition, classification, prediction, and so on. These neurons organized in layers - input layer, hidden layer(s), and the output layer. The neurons connected to each other via weights, in learning phase the ANN try to modify theses weights to minimize the error between the target outputs and the inputs. The weights are modified till the network learned all the training samples. More details about ANN can found in [20]. In our experiment, we utilized Feedforward neural network (FfNet) and Function fitting neural network (FitNet) from Matlab Neural Network Toolbox.

4 Experiment and Data

This experiment has been done to predict wind speed based weather conditions using ANN and genetic algorithm, see Sect. 3. The aim is to compare the different selected features for model created by ANN. The experiment was run many of times with the various settings of selected method. The model performance was evaluated using mean squared error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2, \tag{2}$$

where and mean absolute percentage error (MAPE)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right|. \tag{3}$$

where A_i is the actual value – actual wind speed, F_i is predicted value – predicted wind speed, n is the number of evaluated predicted values. MSE measures the average of the squares of the errors or deviations, the difference between the estimator and what is estimated. MAPE expresses accuracy as a percentage.

The dataset which we used was taken from the Center for Solar Energy Research and Studies (CSRS) Tripoli Tajura[1]. The captured data has been recorded for the whole month November 2015 in a one-minute interval. From the recorded data we choose following values wind direction (Wd) and speed (Ws), air temperature (T), air humidity (Rh), global radiation (Gr), and air pressure (P).

To increase the number of input features, we added data from nearby airports from Tunis - Monastir Habib Bourguiba International Airport (DTMB), Remada Air Base (DTTD), Gabs - Matmata International Airport (DTTG), Djerba - Zarzis International Airport (DTTJ), Sfax - Thyna International Airport (DTTX). All neighboring airports are situated in the west direction from CSRS in Tunis. We could not find any reliable weather data source in other direction or closer to CSRS or in Libya.

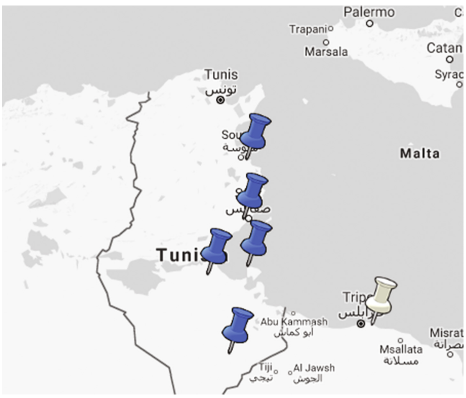


Fig. 2. Weather data source map. (Color figure online)

[1] http://www.csers.ly/en/.

Every nearby airport had data for air temperature, relative humidity, air pressure, and wind speed and direction. The positions of airports and CSRS are depicted on Fig. 2. The Blue pushpins represent weather data obtained from the airports, and the white pushpin is CSRS in Libya.

After collecting and cleaning all necessary data, the input matrix had 31 columns, 6 columns for CSRS - Wd, Ws, T, Rh, Gr, P, and 5 columns for each neighboring airport - Wd, Ws, T, Rh, P. The columns were ordered in following order CSRS, DTMB, DTTD, DTTG, DTTJ, and DTTX. The data were recorded in 1 h in November 2015. The input matrix has 31 columns and 720 rows.

5 Results

The genetic CSSP algorithm was run several times to get different solutions. The CSSP generated 15 combinations of selected features from the input data. For every obtained solution from the CSSP and ANN was created and trained with same data - same input data with selected columns according to the obtained solution. As a reference model, a prediction model with all 31 features was created.

The prediction performance was evaluated using MSE, Eq. 2, and MAPE, Eq. 3. The obtained results are summed up in Table 1, where the prediction results for solutions generated be genetic CSSP and wind speed prediction with all input features. The prediction results were calculated for 24, 36 and 48 h prediction. In all experiments, the same ANN configuration was used. The ANN has three hidden layers with 10, 4, and two neurons with Conjugate Gradient with Powell/Beale Restarts training function.

In average the best prediction results were obtained for 36 h prediction interval. In this experiment, the lowest MAPE 34.6027 (MSE 67.8861) was obtained. The biggest MAPE 50.1062 (MSE 86.7971) was obtained for 48 h prediction interval. In average the generated solution by genetic CSSP had 24 features. In comparison to experiment with all features, the average MAPE was higher. The lowest MAPE for all used features was 35.9368 and MSE 50.2349 for 24 h prediction interval.

For 24 h prediction interval, the MAPE varied between 28.3490 (Solution 9) and 46.270 (Solution 5). The average MAPE in this experiment reached 37.7354. The obtained MSE was in range 34.1339 and 94.0268, with average value 55.0167.

The best results can we see in the experiment with 36 h prediction interval. This experiment has lowest average MAPE value 34.6027. The lowest MAPE 27.8052 was obtained by Solution 4 and the highest 51.0918 by Solution 11. The Solution 4 reached the lowest MAPE (27.8052) value in all experiments. The MSE varied between 37.3323 (Solution 6) and 101.2771 (Solution 3).

The experiment with 48 h prediction interval gave us MAPE equal to 50.1062. The lowest MAPE was generated 37.6175 by Solution 14 and the highest 89.4342 by Solution 3. The MSE varied in range 52.1463 and 174.4476.

Table 1. Experiment results.

Solution	Features	24 h		36 h		48 h	
		MSE	MAPE	MSE	MAPE	MSE	MAPE
1	25	41.1009	29.4123	75.1478	37.8350	71.1637	44.4767
2	25	85.2391	35.4713	99.2258	36.7478	87.9598	62.1109
3	26	52.8699	39.2153	101.2771	37.7870	174.4476	89.4342
4	25	64.0718	34.6016	76.0505	**27.8052**	170.4476	51.6399
5	23	55.0049	46.2703	70.2675	30.0510	59.4679	45.0971
6	23	46.6921	29.1149	37.3323	30.5428	53.5575	41.8518
7	25	45.2233	41.7794	47.5252	29.5093	88.4567	44.2901
8	22	65.1780	30.5553	76.6844	34.1415	100.7582	48.5050
9	24	34.1339	**28.3490**	42.3561	30.0745	95.0537	54.4543
10	24	58.5557	38.6247	50.0489	34.4465	114.7980	49.5355
11	27	48.3963	46.1076	77.1747	51.0918	92.4238	54.8267
12	21	45.3949	46.1076	109.104	38.6007	74.9933	47.6774
13	25	44.1453	37.4104	51.3612	32.0474	52.1463	40.1434
14	24	45.2170	44.5723	48.4116	36.8322	57.4674	**37.6175**
15	27	94.0268	38.4380	56.3235	31.5275	59.1215	39.9324
Average	24	55.0167	37.7354	67.8861	34.6027	86.7971	50.1062
All features	31	50.2349	35.9368	80.4748	36.2525	116.7692	53.3014

6 Conclusion

The primary purpose of designing the prediction models for wind speed or photovoltaic power is to create an intelligence system to effective power grid control. In previous studies as listed in related works section, there are numerous of articles which are focused on wind speed and solar power prediction. Researching in this field is still open topic, due to the difficulty of weather modeling and prediction. We need to take into account not only the measured values, such as air temperature, wind speed and direction and lot of other properties of the environment, but also the characteristics of the power plat where is situated.

This study introduces a combination of a genetic algorithm for column subset selection problem. This Algorithm is used to select the best combination of input features, which are in further step used to train a prediction model. As a prediction model, an artificial neural network was used.

As the Sect. 5 shows. The best prediction performance was obtained in 36 h interval. In this experiment setup, the mean absolute percentage error reaches the lowest value. The applied genetic algorithm for column subset selection problem chooses the most important features of the input data. These reduced data are used to train the prediction model based on an artificial neural network. This combination is suitable to predict the wind speed for wind power plant with multiple nearby weather data sources. As nearby weather sources, we used weather data provided by airports in the region.

In future studies and articles, we will focus on the prediction of wind speed, solar power, and consumption and generation of power, with more different input features for improved prediction of the results.

Acknowledgment. This work was supported by the Czech Science Foundation under the grant no. GACR GJ16-25694Y and in part by Grant of SGS No. SP2017/85 VŠB - Technical University of Ostrava, Czech Republic.

References

1. Zhao, P., Xia, J., Dai, Y., He, J.: Wind speed prediction using support vector regression. In: 2010 5th IEEE Conference on Industrial Electronics and Applications, pp. 882–886, June 2010
2. Bhaskar, K., Singh, S.: Awnn-assisted wind power forecasting using feed-forward neural network. IEEE Trans. Sustain. Energ. **3**(2), 306–315 (2012)
3. Liu, D., Niu, D., Wang, H., Fan, L.: Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew. Energy **62**, 592–597 (2014)
4. Azad, H., Mekhilef, S., Ganapathy, V.: Long-term wind speed forecasting and general pattern recognition using neural networks. IEEE Transact. Sustain. Energ. **5**(2), 546–553 (2014)
5. Wang, J., Zhang, W., Li, Y., Wang, J., Dang, Z.: Forecasting wind speed using empirical mode decomposition and elman neural network. Appl. Soft Comput. J. **23**, 452–459 (2014)
6. Wang, S., Zhang, N., Wu, L., Wang, Y.: Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew. Energy **94**, 629–636 (2016)
7. Meng, A., Ge, J., Yin, H., Chen, S.: Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers. Manage. **114**, 75–88 (2016)
8. Shi, J., Lee, W.J., Liu, Y., Yang, Y., Wang, P.: Forecasting power output of photovoltaic systems based on weather classification and support vector machines. IEEE Trans. Ind. Appl. **48**(3), 1064–1069 (2012)
9. Xu, R., Chen, H., Sun, X.: Short-term photovoltaic power forecasting with weighted support vector machine. In: 2012 IEEE International Conference on Automation and Logistics, Zhengzhou, pp. 248–253 (2012). doi:10.1109/ICAL.2012.6308206
10. Mandal, P., Madhira, S., Ul haque, A., Meng, J., Pineda, R.: Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Procedia Comput. Sci. **12**, 332–337 (2012)
11. Prokop, L., Misak, S., Snasel, V., Krmer, P., Platos, J.: Photovoltaic power plant power output prediction using fuzzy rules. Przeglad Elektrotechniczny **89**(11), 77–80 (2013)
12. Zeng, J., Qiao, W.: Short-term solar power prediction using a support vector machine. Renew. Energ. **52**, 118–127 (2013)
13. Li, Z., Mahbobur Rahman, S., Vega, R., Dong, B.: A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. Energies **9**(1), 55 (2016)

14. Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliari, E.: A physical hybrid artificial neural network for short term forecasting of PV plant power output. Energies **8**(2), 1138–1153 (2015)
15. Teo, T.T., Logenthiran, T., Woo, W.L.: Forecasting of photovoltaic power using extreme learning machine. In: 2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA), Bangkok, pp. 1–6 (2015). doi:10.1109/ISGT-Asia.2015.7387113
16. Kromer, P., Platos, J., Snasel, V.: Genetic algorithm for the column subset selection problem. In: 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems, pp. 16–22, July 2014
17. Boutsidis, C., Mahoney, M.W., Drineas, P.: An improved approximation algorithm for the column subset selection problem. In: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2009, Philadelphia, PA, USA, pp. 968–977. Society for Industrial and Applied Mathematics (2009)
18. Friedberg, S.: Linear Algebra, 4th edn. Prentice-Hall Of India Pvt. Limited (2003)
19. Mitchell, M.: An Introduction to Genetic Algorithms. MIT Press, Cambridge (1998)
20. Zurada, J.: Introduction to Artificial Neural Systems. West Publishing Co., St. Paul (1992)

Evaluation of Pseudorandom Number Generators Based on Residue Arithmetic in Differential Evolution

Pavel Krömer(✉), Jan Platoš, and Václav Snášel

Department of Computer Science, VŠB Technical University of Ostrava, Ostrava, Czech Republic
{pavel.kromer,jan.platos,vaclav.snasel}@vsb.cz

Abstract. Sequences of pseudorandom numbers, produced by various pseudorandom number generators (PRNGs), are essential for the implementation of stochastic algorithms, simulations, and methods including nature–inspired computation. The non–determinism, observed in nature, is emulated with the help of pseudorandom values, predefined probabilities and thresholds, and the properties of generated sequences may have significant impact on the computation. They affect both, the ability of the algorithms to find high–quality problem solutions and the speed of the computation. Residue number systems provide means for alternative representation of integers that is suitable for parallel systems. In this work, several PRNGs based on residue arithmetic and residue number systems are empirically investigated as a source of stochasticity for the differential evolution algorithm.

Keywords: Differential evolution · Pseudorandom numbers · Residue arithmetic · Inversive congruential generator

1 Introduction

Non–determinism and stochasticity are important properties of natural systems from which evolutionary and swarm methods draw inspiration. In practical implementations of nature–inspired algorithms, stochasticity is obtained by series of pseudorandom numbers produced by different types of hardware and/or software generators. The properties of the pseudorandom sequences (value distribution, probability density, period) have significant impact on the algorithms.

There are many different PRNGs with different properties. Some of them provide strong security and pass complex statistical randomness tests. They are usually used for cryptographic applications where resilience to cryptanalysis and various types of sophisticated attacks is a must. Nature–inspired algorithms, on the other hand, do not pose the same requirements for pseudorandom number sequences. Pseudorandom number sequences used in nature–inspired methods do not need to be cryptographically secure, completely unpredictable, or to

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_30

have extremely long periods. The primary high–level requirement for pseudorandom sequences in nature–inspired computation is that they *drive the algorithm towards globally or at least locally optimal problem solutions* and *enable efficient exploration of problem (solution) space*. The secondary requirement is that they can be *produced at low time and space costs*. Because of that, it is often not necessary to use complex cryptographically secure PRNGs in nature–inspired algorithms. Instead, lightweight generators based e.g. on deterministic chaos [12] or other principles can be successfully employed. However, the suitability of a particular PRNG is often algorithm and problem (application) bound [3,4,13]. On the other hand, some experiments suggest that certain stochastic algorithms might be able to solve simple (e.g. very low–dimensional) problems regardless the used PRNG [24].

Residue number systems (RNS) provide an alternative representation of integers as remainders after integer division by a specially designed set of divisiors called moduli. The remainders are collectively known as residues and form a residue set representation of an integer. The residues are much smaller than the corresponding integers and arithmetic operations including addition, subtraction, and multiplication, are executed over the residue sets [21]. These operations are in a RNS data parallel because such systems are non–positional and there is no carry–propagation. Although the realization of some operations is in RNS more complex than in traditional number systems, they generally allow fast implementation of large integer arithmetic on parallel systems.

There are several algorithms for pseudorandom number generation based on residue number systems [9] or, more generally, on the underlying congruence relation [6–8,11]. It is well–known that they are neither cryptographically secure nor suitable for all types of applications [5,10]. However, the pseudorandom sequences they generate have interesting properties (e.g. provably maximum–length period [7]) and may be useful for nature–inspired algorithms. In this work, we implement and empirically evaluate five RNS–based pseudorandom number generators. They are used as source of stochasticity for the differential evolution algorithm solving problems from the 2017 edition of the well–known CEC real–parameter optimization benchmark [1]. The rest of the work is organized in the following way: Sect. 2 summarizes recent work on PRNGs in nature–inspired computation. Section 3 introduces RNS and elementary operations in residue arithmetic. The principles of PRNGs based on the congruence relation and on RNS and are introduced in Sect. 4. Differential evolution is briefly described in Sect. 5 and the computational experiments conducted in order to empirically evaluate the generators are described and discussed in Sect. 6. Finally, the work is concluded in Sect. 7.

2 Stochasticity in Evolutionary Computation

The impact of different pseudorandom number generators and other sources of stochasticity on the performance of evolutionary and in general stochastic optimization methods has been studied intensively [2–4,14,16–19,24–26].

The research has focused especially on the ability of nature–inspired algorithms powered by different sources of stochasticity to find high–quality problem solutions and to quickly converge under different conditions.

Several works [17–19] have analyzed the effect different pseudo–random number generation strategies have on Genetic Algorithms (GAs). They have shown that a high–quality PRNG is not necessary to facilitate a successful evolutionary search and conversely, even a poor PRNG can drive a well–performing simple GA [19]. Another study [3] revealed more complex relationship between PRNGs and the performance of simple GA. It has shown that different stages of the algorithm have different sensitivity to pseudo–random number quality. The generation of the initial population was most sensitive to PRNG quality while the rest of the algorithm is more robust with regard to the choice of the PRNG. The study also presented that the selection of PRNG parameters and initial seeds has significant impact on the performance of the GA. Random seeds were found as a rather bad choice for PRNG initialization. It has also shown that the behavior of GA in relation to the PRNG is not the same for all types of problems (e.g. for all test functions) [3].

The effect of PRNG quality on different variants of GA and Simulated Annealing algorithm solving the Travelling Salesman Problem (i.e. a combinatorial optimization problem) was assessed in [16]. The experiments suggested that period length is an important property of a PRNG and that population–based algorithms such as GA and DE have different sensitivity to PRNG properties than trajectory–based methods including the Simulated Annealing algorithm. Anyway, the work concluded that the use of diverse PRNGs contributes to better results in general. The sensitivity of four evolutionary algorithms, including GA and DE, to the change of PRNG was investigated in [4]. Mersenne Twister [15] and GCC rand were used by the investigated algorithms to find optima of several test functions and the results showed that both GA and DE *are* affected by the choice of PRNG. However, no clear correlation between PRNG quality and algorithm performance was found. This observation was confirmed in [26] where the authors confirmed that the performance of the algorithm is affected by the choice of PRNG but a lower-quality PRNG can yield a DE with good performance.

Several studies investigated recently the potential of alternative sources of stochasticity for evolutionary and in general nature–inspired computation [12–14,25]. For example, a PRNG with Lévy probability distribution was considered for the use with Evolutionary Programming [14] and Deterministic Chaos was investigated in context of GA and DE [12,13,25]. This overview of related work does confirm that the type and quality of stochasticity source does not correlate with the performance of GA and DE in a general case and the computational costs required to obtain high–quality PRNG might be better used at another stage of the algorithms. Together, they provide a body of evidence that the relation between PRNGs, their properties, and algorithm performance is a complex algorithm and application dependent problem.

3 Residue Number Systems

Residue number systems [20,21] provide means for an alternative representation of integers that does not rely on the correspondence between sequences of digits and numbers in the usual sense [21]. RNSs are based on the congruence relation and represent numbers by sets of remainders (residues) after integer division by a fixed set of divisors (moduli). Suppose q and r are the quotient and remainder of integer division and $a = q \cdot m + r$. Then r is the residue of a with respect to the modulus m and we write $r = |a|_m$. A RNS is a number representation and arithmetic system defined by a set of two or more moduli [20,21].

Definition 1 (Residue number system). *Consider a set of positive integers (moduli),* $\mathbf{M} = \{m_1, m_2, \ldots, m_N\}$, *for which it holds that* $\forall j, k \in N : j \neq k \Rightarrow m_j$ *and* m_k *have no common divisor greater than 1. Every two moduli in* $\mathbf{M}$ *are relatively prime (co–prime) to each other and any number,* x, *can be represented by a set of residues,* $\mathbf{r} = \{|x|_{m_i} : 1 \leq i \leq N\}$. *Let* $\mathcal{DR} = \prod_{i \in N}(m_i)$ *be the dynamic range of the RNS. Every* $x < \mathcal{DR}$ *has an unique representation in the RNS defined by* $\mathbf{M}$. *A RNS can represent positive integers from the range* $[0, \mathcal{DR} - 1]$. *We adopt the notation from [21] and write*

$$x \cong \langle x_1, x_2, \ldots, x_N \rangle \tag{1}$$

to indicate that x *is a number uniquely represented by the residue set* $\{|x|_{m_1}, \ldots, |x|_{m_1}\}$ *in the RNS defined by the moduli set* $\{m_1, m_2, \ldots m_N\}$.

Basic arithmetic operations are in RNSs defined on residue sets [20,21]. Addition of two numbers, $x \cong \langle x_1, x_2, \ldots, x_N \rangle$ and $y \cong \langle y_1, y_2, \ldots, y_N \rangle$, is in a RNS defined by

$$\begin{aligned} x + y &\cong \langle x_1, x_2, \ldots, x_N \rangle + \langle y_1, y_2, \ldots, y_N \rangle \\ &= \langle |x_1 + y_1|_{m_1}, |x_2 + y_2|_{m_2}, \ldots, |x_N + y_N|_{m_N} \rangle. \end{aligned} \tag{2}$$

Multiplication of x and y is defined in an analogous way

$$\begin{aligned} x \times y &\cong \langle x_1, x_2, \ldots, x_N \rangle \times \langle y_1, y_2, \ldots, y_N \rangle \\ &= \langle |x_1 \times y_1|_{m_1}, |x_2 \times y_2|_{m_2}, \ldots, |x_N \times y_N|_{m_N} \rangle. \end{aligned} \tag{3}$$

In the simplest case, RNSs work with unsigned non–negative integers. However, they can be easily extended on negative integers and subtraction operation can be defined. The realization of other arithmetic operations such as comparison and division are in general problematic in RNSs [21].

Although RNSs are generally more complex than traditional number systems, the definitions of addition and multiplication clearly show why are they attractive for an efficient realization of fast arithmetics. The computations with potentially very large numbers are split into N less complex independent tasks in a divide–and–conquer manner [20] and can be executed concurrently. They are performed over individual residues that have significantly smaller range than the

RNS as a whole. Because of that and because there is no carry–propagation for addition and multiplication in RNSs, the operations can be realized in hardware fast [20] and at low–power costs [21].

Residue number systems are interesting for pseudorandom number generation. The congruence relation, upon which they are based, has been used as a fundamental concept of a family of PRNGs suitable for parallel pseudorandom number generation [6,7,11]. Although Inversive Congruential Generators (ICG) have good theoretical and practical properties and are useful for massive Monte Carlo computations [11], they have to be used with care and generally with regard to the intended application domain [10]. Recently, RNSs have been used explicitly as a part of a hybrid PRNG design suitable for an efficient hardware implementation [9].

4 Inversive Congruential Generators

Inversive Congruential Generators [6,7,11] form a family of PRNGs that use the congruence relation. They use the congruence

$$y_{n+1} \equiv a \cdot y_n^{-1} + b \pmod{p}, \tag{4}$$

where $p \geq 5$ is a prime, $a \in \mathbb{Z}_p \backslash \{0\}$ and $b \in \mathbb{Z}_p$ are integers such that $|a|_p \neq 0$ and $|b|_p \neq 0$, and y_n^{-1} is multiplicative inverse of y_n modulo p, to generate an *inversive congruential sequence*. In the previous, $\mathbb{Z}_p$ is a finite field of order p. Although the sequence is purely periodic, an ICG with a prime modulus has no lattice structure [11] and it is known that a proper choice of parameters a and b leads to the maximum period length [7]. Real values from the range $[0, 1)$ can be obtained from Eq. 4 by the normalization [11]

$$x_n = \frac{y_n}{p}. \tag{5}$$

Explicit Inversive Congruential Generators (EICG) [8] are based on a modified congruence in the form [8]

$$y_n \equiv (a \cdot n + b)^{-1} \pmod{p} \tag{6}$$

which allows generation of an *inversive congruential sequence* without recursion in a completely data parallel way. EICGs have been analyzed and found able to generate sequences similar to those produced by ICGs. An ICG and EICG can be characterized by a condensed inline notation in which (E)ICG(p, a, b) denotes an (Explicit) Inversive Congruential Generator with modulus p and parameters a and b.

The disadvantages of inversive congruential generators are twofold. First, in order to produce sequences with sufficiently long period, a very large prime p has to be chosen and big integer arithmetics is required for its computational realization. Second, the computation of modular multiplicative inverse is required for every number of the sequence. In this step holds that the larger modulus, the more time is required for for the computation of the multiplicative inverse [5].

An approach that addresses both problems at the same time is the *compound method* for inversive congruential sequence generation [6]. Informally, it combines the outputs of several inversive congruential generators together to generate a sequence of integers with very large period while using only relatively small moduli [6,11]. Consider a set of N distinct primes $p_1, p_2, \dots p_N$ and the maximum length sequences generated by the corresponding ICGs or EICGs, $\{y_n^{(i)} : 0 \leq n \leq p_i, 1 \leq i \leq N\} = \mathbb{Z}_{p_i}$. Then, a sequence of *compound (explicit) inversive pseudorandom numbers* is defined by

$$x_n \equiv \sum_{i=1}^{N} \frac{y_n^{(i)}}{p_i} \pmod 1, \quad n \geq 0. \tag{7}$$

The period of the sequence defined by Eq. 7 is equal to $\prod_{i=1}^{N} p_i$ [8,11]. Obviously, all computationally expensive integer operations are performed in parallel and in modulo $p_1, p_2, \dots p_N$ (i.e. in $\mathbb{Z}_{p_1}, \mathbb{Z}_{p_2}, \dots \mathbb{Z}_{p_N}$).

Intuitively, it is clear that compound ICG and EICG use the same strategy to represent very large numbers and perform arithmetic operations over large integers as RNSs. A divide–and–conquer strategy is used to split the operation into several independent streams that operate in smaller precision and can be efficiently implemented by modern microprocessors and circuits (field programmable gate arrays [9]). Compound inversive congruential generators, however, avoid the most expensive part of RNS computations – exact transformation from RNS to native number representation.

A pseudorandom number generator design explicitly combining RNS and ICG was proposed and evaluated in [9]. The authors split the computation of a sequence of pseudorandom numbers to several streams (channels), each of them implementing an ICG in $\mathbb{Z}_{p_i}$. They introduced several modifications in order to enable a seamless FPGA implementation while maintaining sufficient randomness of the generated sequences. A pseudorandom number, x_n is in this approach defined by

$$x_n \cong \langle x_{n,1}, x_{n,2}, \dots, x_{n,N} \rangle, \tag{8}$$

where $x_{n,i}$ are residues computed by

$$x_{n+1,i} = (g_i \cdot x_{n,i}^{-1} + b_i) \pmod{p_i}, \ 1 \leq i \leq N \tag{9}$$

and g_i and p_i are primitive root and modulo of each channel and b_i is a constant. It is easy to see that each channel implements an ICG.

Apparently, residue number systems, residue arithmetic, and inversive congruential generators are able to generate sequences of pseudorandom numbers with interesting properties. That makes them a viable source of randomness for stochastic optimization algorithms. In the rest of this work, we evaluate the ability of differential evolution together with selected congruential PRNGs to solve problems from a well–know battery of real–parameter optimization functions [1].

5 Differential Evolution

The DE is a popular stochastic evolutionary optimization algorithm that evolves a population of real encoded vectors representing the solutions to given problem [23]. The algorithm starts with an initial population of M real-valued vectors. The vectors are initialized with real values either randomly or so, that they are evenly spread over the problem space. The latter initialization leads to better results of the optimization [23]. The algorithm perturbs selected base vectors with the scaled difference of two (or more) other population vectors in order to produce the trial vectors. The trial vectors compete with members of the current population with the same index called the target vectors. If a trial vector represents a better solution than the corresponding target vector, it takes its place in the population [23].

The two most significant parameters of the DE are scaling factor and mutation probability [23]. The scaling factor, $F \in [0, \infty)$, controls the rate at which the population evolves and the crossover probability, $C \in [0, 1]$, determines the ratio of elements that are transferred to the trial vector from its opponent. The size of the population and the choice of operators are important parameters of the optimization process.

The DE is a successful evolutionary algorithm designed for continuous parameter optimization and driven by the idea of scaled vector differentials. Similarly to other nature–inspired algorithms, it is a stochastic procedure and its implementation has to rely on a particular source of stochasticity. It is well–known that its performance is affected by the choice of PRNG or other source of randomness but this relationship is complex and no clear correlations between PRNG properties and algorithm performance have been found [12,17–19,24]. That motivates further research and empirical evaluation of alternative sources of stochasticity for the use with differential evolution.

6 Experiments

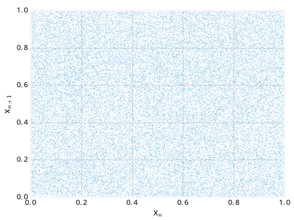
In this work, we compare the performance of a simple differential evolution algorithm, *DE/rand/1*, using four PRNGs based on the congruence relation and residue number systems. Although two of the generators have potential to accelerate the computation in parallel environments, they are in this study investigated from the *solution quality* point of view to address the important question how do they affect the ability of the underlying algorithm to solve complex optimization problems. The algorithm was chosen due to its simple structure, high performance, good results, and natural ability to operate in parallel [22]. It was used with the following fixed parameters: population size 70, scaling factor $F = 0.5$, mutation probability $C = 0.9$, and maximum number of fitness function evaluations 100,000. The parameters were selected on the basis of best practices, previous experience with the algorithm, and intensive trial–and–error runs.

The compared generators and their properties are summarized in Table 1. The parameters were selected so that the generators produce sequences of pseudorandom numbers with similar word length and similar period. The dependence of

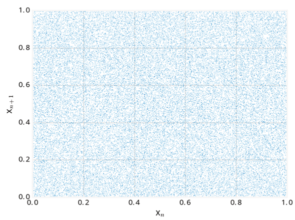
Table 1. Compared PRNGs.

Generator	Description
ICG	ICG (2^{31}, 1, 1)
EICG	EICG (2^{31}, 1, 1)
CICG	Compound ICG [6,11] composed of ICG (1033, 413, 1), ICG(1039, 173, 1), and ICG (2027, 579, 1)
CEICG	Compound EICG [6,11] composed of EICG (1033, 413, 1), EICG (1039, 173, 1), and EICG (2027, 579, 1). An explicit version of $\mathcal{CICG}$
RNS-ICG	The simplest version of RNS–ICG according to [9] with $m = \{1033, 10939, 2027\}$, $g = \{413, 173, 579\}$, and fixed $b = \{1, 1, 1\}$. A RNS implementation of $\mathcal{CICG}$

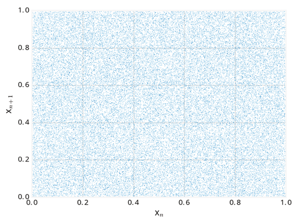
x_{n+1} on x_n is for each investigated generator for a sequence of 100,000 generated pseudorandom numbers from the range $[0, 1]$ shown in Fig. 1. The figures show that there are no visually detectable structures or regularities in the generated sequences.



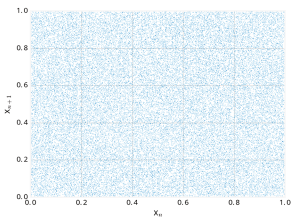
(a) ICG



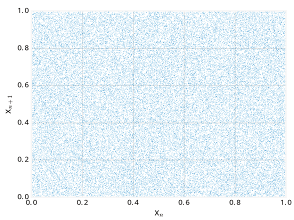
(b) EICG



(c) CICG



(d) CEICG



(e) RNS-ICG

Fig. 1. Dependency of the $(n + 1)$-th generated number on the n-th number.

A well–known suite of test functions, developed for the special session on real–parameter single objective optimization at the IEEE Congress on Evolutionary Computation [1], was used to assess the ability of the algorithm with different generators to solve continuous optimization problems under different circumstances.

It consists of 30 test functions that represent a wide variety of problems with different properties. They include unimodal, multimodal, hybrid, and composite functions with various shapes, ruggedness, separability, symmetry (asymmetry), number of optima and so on [1]. All experiments were executed with problem dimension set to 10 and repeated for each configuration 51 times independently. The results

Table 2. Error of results obtained by DE with ICG and EICG.

F	ICG			EICG		
	Min	Mean	Max	Min	Mean	Max
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0.56	1.943	0	0.722	2.322
5	0.995	12.842	27.339	0.997	12.894	27.17
6	0	0	0	0	0	0
7	12.726	29.82	38.936	13.684	30.569	41.399
8	1.99	16.817	29.669	2.985	17.258	28.763
9	0	0	0	0	0	0
10	7.08	330.575	1;126.33	6.955	323.496	1,059.9
11	0	0.27	1.99	0	0.232	0.995
12	0	9.317	158.746	0	21.089	158.746
13	0	4.329	6.141	0	3.744	5.761
14	0	0.344	1.99	0	0.172	1.99
15	1.010e–6	0.14	0.997	5.300e–7	0.138	1.002
16	2.012e–2	2.698	11.541	3.156e–4	5.278	38.098
17	0	2.862	20.493	0	3.262	17.752
18	5.460e–6	0.468	20.007	2.227e–5	0.467	20.111
19	0	4.116e–3	1.973e–2	0	7.123e–3	3.916e–2
20	0	0.163	1.619	0	0.205	0.995
21	100	175.027	231.975	100	182.064	227.932
22	11.563	95.138	100.985	0	94.431	100.998
23	300	306.517	316.383	302.576	306.389	312.782
24	100	317.714	352.765	100	330.544	353.201
25	397.743	404.274	445.824	397.743	412.209	445.805
26	300	300.907	347.149	300	302.601	388.081
27	389.006	389.457	390.752	388.283	389.491	391.628
28	300	349.101	611.822	300	362.595	611.822
29	226.579	229.721	235.158	225.573	229.151	237.442
30	394.501	24,428.983	1.249e+6	394.501	31,909.564	8.176e+5

of computational experiments are summarized in Tables 2 and 3, respectively. The tables list for each test function minimum, mean, and maximum error of final solutions obtained by the DE with corresponding PRNG across the 51 independent optimization runs. A ranking, comparing all investigated algorithms for each test function with respect to minimum and mean fitness, is shown in Table 4. The table lists the rank (order) each algorithm variant achieved for every test function. If the use of more PRNGs lead to the same fitness, the sign '−' is used in the table to show that no ranking was possible. Table 4 clearly shows that while ICG, EICG, CICG, and CEICG compete for the better ranks, the DE with RNS-ICG as with respect to the error of the minimum (best) obtained result ranked last in almost every case. On the other hand, it was able to place on top of the ranking to mean

Table 3. Error of results obtained by DE with CICG, CEICG, and RNS-ICG.

F	CICG			CEICG			RNS-ICG		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
1	0	0	0	0	0	0	0.703	9.126	30.988
2	0	0	0	0	0	0	$7.455e{-5}$	0.612	14.54
3	0	0	0	0	0	0	3.828	30.673	128.828
4	0	0.564	1.884	0	0.562	2.243	3.221	4.409	5.652
5	0.995	15.911	28.082	1.99	12.635	31.435	6.694	18.695	26.06
6	0	0	0	0	0	0	$2.000e{-8}$	$1.500e{-7}$	$4.100e{-7}$
7	10.808	30.654	39.128	16.963	31.606	42.456	21.443	30.83	40.655
8	1.995	15.883	30.358	2.985	16.207	28.483	8.424	16.07	23.066
9	0	0	0	0	0	0	0	0	0
10	10.307	342.68	1,142.28	0.437	348.968	1,128.85	16.716	295.326	500.78
11	0	0.134	0.995	0	0.216	1.131	0.344	3.143	7.64
12	0	21.883	238.176	0	9.773	130.911	3,949.03	14,396.994	29,441.9
13	0	3.863	6.703	0	4.186	5.761	8.888	13.784	19.425
14	0	0.249	1.99	0	0.23	0.995	0.301	9.56	18.6
15	$2.650e{-6}$	0.158	0.5	$8.000e{-8}$	$7.780e{-2}$	0.999	0.472	1.613	3.064
16	$5.014e{-3}$	2.429	11.688	$1.817e{-5}$	2.615	28.777	0.371	4.821	38.629
17	0	1.361	16.757	0	2.529	17.752	$6.478e{-2}$	2.514	18.656
18	$3.550e{-6}$	1.66	20.112	$3.070e{-6}$	0.487	20.018	5.032	11.54	20.972
19	0	$3.408e{-2}$	1.5	0	$5.611e{-3}$	$1.973e{-2}$	$6.060e{-2}$	0.551	1.319
20	0	0.114	0.624	0	0.205	0.995	0	$3.620e{-2}$	0.312
21	100	181.203	231.679	100	185.887	228.675	102.293	142.979	234.117
22	0	98.279	100.882	0	92.688	100.65	$4.426e{-5}$	99.816	104.519
23	302.576	306.087	313.851	302.576	306.292	319.134	309.233	317.046	326.589
24	100	299.758	343.417	327.09	335.566	354.761	102.107	347.765	370.537
25	397.743	408.562	445.793	397.743	410.533	445.801	397.743	425.672	446.078
26	300	300	300	300	301.813	347.149	300	305.44	347.149
27	388.277	389.708	392.796	389.006	389.522	390.752	389.006	389.661	390.752
28	300	336.525	611.822	300	337.077	611.822	304.777	456.858	604.414
29	225.575	229.551	239.699	226.092	229.424	234.774	249.784	269.106	293.096
30	394.501	55,849.561	1.249e+6	394.501	1,012.626	31,412.5	537.131	7,085.833	38,389.8

Table 4. Ranking (order) of DEs with different PRNGs.

F	According to min. fitness					According to mean fitness				
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
1	-	-	-	-	RNS-ICG	-	-	-	-	RNS-ICG
2	-	-	-	-	RNS-ICG	-	-	-	-	RNS-ICG
3	-	-	-	-	RNS-ICG	-	-	-	-	RNS-ICG
4	-	-	-	-	RNS-ICG	ICG	CEICG	CICG	EICG	RNS-ICG
5	-	-	EICG	CEICG	RNS-ICG	CEICG	ICG	EICG	CICG	RNS-ICG
6	-	-	-	-	RNS-ICG	-	-	-	-	RNS-ICG
7	CICG	ICG	EICG	CEICG	RNS-ICG	ICG	EICG	CICG	RNS-ICG	CEICG
8	ICG	CICG	-	-	RNS-ICG	CICG	RNS-ICG	CEICG	ICG	EICG
9	-	-	-	-	-	-	-	-	-	-
10	CEICG	EICG	ICG	CICG	RNS-ICG	RNS-ICG	EICG	ICG	CICG	CEICG
11	-	-	-	-	RNS-ICG	CICG	CEICG	EICG	ICG	RNS-ICG
12	-	-	-	-	RNS-ICG	ICG	CEICG	EICG	CICG	RNS-ICG
13	-	-	-	-	RNS-ICG	EICG	CICG	CEICG	ICG	RNS-ICG
14	-	-	-	-	RNS-ICG	EICG	CEICG	CICG	ICG	RNS-ICG
15	CEICG	EICG	ICG	CICG	RNS-ICG	EICG	CICG	CEICG	ICG	RNS-ICG
16	CEICG	EICG	CICG	ICG	RNS-ICG	CICG	CEICG	ICG	RNS-ICG	EICG
17	-	-	-	-	RNS-ICG	CICG	RNS-ICG	CEICG	ICG	EICG
18	CEICG	CICG	ICG	EICG	RNS-ICG	EICG	ICG	CEICG	CICG	RNS-ICG
19	-	-	-	-	RNS-ICG	ICG	CEICG	EICG	CICG	RNS-ICG
20	-	-	-	-	-	RNS-ICG	CICG	ICG	CEICG	EICG
21	-	-	-	-	RNS-ICG	RNS-ICG	ICG	CICG	EICG	CEICG
22	-	-	-	ICG	RNS-ICG	CEICG	EICG	ICG	CICG	RNS-ICG
23	ICG	-	-	-	RNS-ICG	CICG	CEICG	EICG	ICG	RNS-ICG
24	-	-	-	RNS-ICG	CEICG	CICG	ICG	EICG	CEICG	RNS-ICG
25	-	-	-	-	-	ICG	CICG	CEICG	EICG	RNS-ICG
26	-	-	-	-	-	CICG	ICG	CEICG	EICG	RNS-ICG
27	CICG	EICG	-	-	-	ICG	EICG	CEICG	RNS-ICG	CICG
28	-	-	-	-	RNS-ICG	CICG	CEICG	ICG	EICG	RNS-ICG
29	EICG	CICG	CEICG	ICG	RNS-ICG	EICG	CEICG	CICG	ICG	RNS-ICG
30	-	-	-	-	RNS-ICG	CEICG	RNS-ICG	ICG	EICG	CICG

fitness in the case of 3 test functions. However, it still performed poorly in case of the remaining test functions.

7 Conclusions

An empirical comparison of the Differential Evolution powered by five different types of PRNGs based on residue number systems and the congruence relation was performed in this work. A well known suite of benchmarking functions was used to assess the ability of the algorithms to find global optima of 30 10–dimensional real–parameter functions with different properties. The comparison confirmed that the ability of the algorithm to converge to optimal solutions is indeed affected by the used PRNG and that different residue number system based PRNGs lead the algorithm to solutions of different quality. The RNS-ICGgenerator, proposed in [9] did not allow the algorithm to reach as good solutions as inversive congruential generators and their combined variants. The CEICG generator yielded the most successful PRNG strategy with respect to the error of the best found solution. That is an encouraging result that suggests a more intensive evaluation of residue number system based sources of stochasticity in evolutionary computation.

Acknowledgement. This work was supported by the Czech Science Foundation under the grant no. GJ16-25694Y and by the projects SP2017/100 "Parallel processing of Big Data IV", of the Student Grant System, VŠB-Technical University of Ostrava.

References

1. Awad, N.H., Ali, M.Z., Liang, J.J., Qu, B.Y., Suganthan, P.N.: Problem definitions and evaluation criteria for the CEC 2017 special session and competition on single objective real-parameter numerical optimization, Technical report, October 2016
2. Bland, I.M., Megson, G.: Systolic random number generation for genetic algorithms. Electron. Lett. **32**(12), 1069–1070 (1996)
3. Cantú-Paz, E.: On random numbers and the performance of genetic algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2002, pp. 311–318. Morgan Kaufmann Publishers Inc., San Francisco (2002)
4. Cárdenas-Montes, M., Vega-Rodríguez, M.A., Gómez-Iglesias, A.: Sensitiveness of evolutionary algorithms to the random number generator. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2011. LNCS, vol. 6593, pp. 371–380. Springer, Heidelberg (2011). doi:10.1007/978-3-642-20282-7_38
5. Chi, H., Cao, Y.: Generating parallel random sequences via parameterizing EICGs for heterogeneous computing environments. In: Taniar, D., Gervasi, O., Murgante, B., Pardede, E., Apduhan, B.O. (eds.) ICCSA 2010. LNCS, vol. 6019, pp. 409–417. Springer, Heidelberg (2010). doi:10.1007/978-3-642-12189-0_36
6. Eichenauer-Herrmann, J.: Explicit inversive congruential pseudorandom numbers: the compound approach. Computing **51**(2), 175–182 (1993)
7. Eichenauer-Herrmann, J.: Construction of inversive congruential pseudorandom number generators with maximal period length. J. Comput. Appl. Math. **40**(3), 345–349 (1992)
8. Eichenauer-Herrmann, J.: Statistical independence of a new class of inversive congruential pseudorandom numbers. Math. Comput. **60**(201), 375–384 (1993)
9. Gayoso, C.A., González, C., Arnone, L., Rabini, M., Moreira, J.C.: Pseudorandom number generator based on the residue number system and its FPGA implementation. In: 2013 7th Argentine School of Micro-nanoelectronics, Technology and Applications, pp. 9–14, August 2013
10. Hellekalek, P.: Don't trust parallel monte carlo! In: Twelfth Workshop on Parallel and Distributed Simulation, 1998, PADS 1998. Proceedings, pp. 82–89, May 1998
11. Hellekalek, P.: Inversive pseudorandom number generators: concepts, results and links. In: Proceedings of the 27th conference on Winter simulation, WSC 1995, Arlington, VA, USA, 3–6 December 1995, pp. 255–262 (1995)
12. Krömer, P., Snásel, V., Zelinka, I.: Randomness and chaos in genetic algorithms and differential evolution. Proc. INCoS **2013**, 196–201 (2013)
13. Krömer, P., Zelinka, I., Snásel, V.: Can deterministic chaos improve differential evolution for the linear ordering problem? In: Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2014, pp. 1443–1448 (2014)

14. Lee, C.Y., Yao, X.: Evolutionary programming using mutations based on the Levy probability distribution. IEEE Trans. Evol. Comput. **8**(1), 1–13 (2004)
15. Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. **8**(1), 3–30 (1998)
16. Maucher, M., Schöning, U., Kestler, H.A.: Search heuristics and the influence of non-perfect randomness: examining genetic algorithms and simulated annealing. Comput. Stat. **26**(2), 303–319 (2011)
17. Meysenburg, M.M., Foster, J.A.: The quality of pseudo-random number generations and simple genetic algorithm performance. In: ICGA, pp. 276–282 (1997)
18. Meysenburg, M.M., Foster, J.A.: Randomness and GA performance, revisited. In: Banzhaf, W., Daida, J., Eiben, A.E., Garzon, M.H., Honavar, V., Jakiela, M., Smith, R.E. (eds.) Proceedings of the Genetic and Evolutionary Computation Conference, vol. 1, pp. 425–432. Morgan Kaufmann, Orlando, 13–17 July 1999
19. Meysenburg, M.M., Hoelting, D., McElvain, D., Foster, J.A.: How random generator quality impacts GA performance. In: GECCO, pp. 480–487 (2002)
20. Ananda Mohan, P.V.: Residue Number Systems: Theory and Applications. Springer, Cham (2016)
21. Omondi, A., Premkumar, B.: Residue Number Systems: Theory and Implementation. Advances in Computer Science and Engineering: Texts. Imperial College Press, London (2007)
22. Penas, D.R., Banga, J.R., González, P., Doallo, R.: A parallel differential evolution algorithm for parameter estimation in dynamic models of biological systems. In: Saez-Rodriguez, J., Rocha, M., Fdez-Riverola, F., De Paz Santana, J. (eds.) 8th International Conference on Practical Applications of Computational Biology and Bioinformatics, PACBB 2014. Advances in Intelligent Systems and Computing, pp. 173–181, vol. 294. Springer, Cham (2014). doi:10.1007/978-3-319-07581-5_21
23. Price, K.V., Storn, R.M., Lampinen, J.A.: Differential Evolution A Practical Approach to Global Optimization. Natural Computing Series. Springer, Berlin (2005)
24. Rajashekharan, L., Shunmuga Velayutham, C.: Is differential evolution sensitive to pseudo random number generator quality? – an investigation. In: Berretti, S., Thampi, S.M., Srivastava, P.R. (eds.) Intelligent Systems Technologies and Applications. AISC, vol. 384, pp. 305–313. Springer, Cham (2016). doi:10.1007/978-3-319-23036-8_26
25. Senkerik, R., Pluhacek, M., Zelinka, I., Jasek, R.: On the adaptivity and complexity embedded into differential evolution. AIP Conf. Proc. **1738**(1), 120030 (2016)
26. Tirronen, V., Äyrämö, S., Weber, M.: Study on the effects of pseudorandom generation quality on the performance of differential evolution. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2011. LNCS, vol. 6593, pp. 361–370. Springer, Heidelberg (2011). doi:10.1007/978-3-642-20282-7_37

Research and Realization of News Gathering and Editing System Based on Distributed Search Engine

Yamin Han[1,2], Kun Liu[1,2](✉), and Kun Ma[1,2]

[1] Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, Shandong, China
ise_liuk@ujn.edu.cn

[2] School of Information Science and Engineering, University of Jinan, Jinan, People's Republic of China

Abstract. According to the investigation and survey on the demands of e23.cn, the depth analysis was performed to explore the new changes in the news material gathering and material search methods in the system in the context of big data. Meanwhile, the related technologies of web crawler, replicated documents elimination and distributed search engine are integrated in the news gathering and editing system. The experimental results showed that the system had the functions of efficiently gathering Web materials, eliminating the replicated documents and fast retrieving information, which can meet the actual needs required by the industry of news.

Keywords: Web crawler · Replicated documents elimination · ElasticSearch · Distributed search

1 Introduction

Presently, with the integrated development of traditional and emerging media, the news industry has been rapidly developed with the application of internet technology in China. On one hand, the news sharing has been enhanced, where the news materials have been referred and news have been reproduced mutually among news media. The news collectors may spend a lot of time to search and browse the news websites when gathering Web news and different editors may collect the same Web news materials, which result in the information redundancy. On the other hand, due to the great increase in the news data, the retrieval efficiency has been significantly decreased. It thus makes the request of new generation for the news industry, namely a set of efficient news gathering and editing system to gather the news materials and fast retrieving the information.

2 System Design

2.1 Design of Gathering Structure

The objective of news material gathering is to grab the unstructured data on HTML pages [1] and these data are saved in the local database in the structured form and the index

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8, DOI 10.1007/978-3-319-65636-6_31

established using the search engine. The library of news material is then established for the search and reference by news staff.

During the news material gathering, the technology of web crawler can be used to gather Web news from the specified Web sites or other electronic readings. The collected contents are then saved in the material library after the treatment of analysis, duplication elimination and classification, as well as being stored in the index established by the search engine. On one hand, the web crawler is capable to automatically, efficiently and timely acquire the extensive news materials to meet the demands of material gathering by the media. On the other hand, the index is established using the search engine, which can increase the search speed and improve the user experience. The main structure of Web news gathering is shown in Fig. 1.

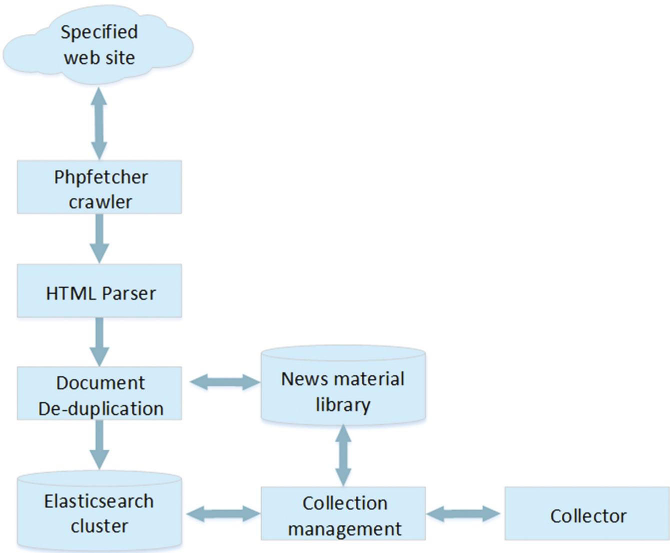


Fig. 1. Structural diagram of web news gathering

The crawler used in this system is established using the crawling framework of Phpfetcher. Such crawling framework consists of four objects, namely Dom, Page, Crawler and Manager.

(1) Dom is used to parse HTML and Dom of HTML can be accessed.
(2) Page corresponds to a specific HTML page and all contents of such page can be acquired.
(3) Crawler refers to the crawling object, setting the rules to crawl the pages.
(4) Manager is responsible for the management of Crawler.

The design for most of news sites and comprehensive sites that contain the news information is generally standard. The classification is relatively clear and the news pages of the same type are saved in the same subdirectory. The news pages have the strong timeliness. Only the specified area is gathered during the material gathering.

The main work of HTML parsing is to parse the gathered Web source files and extract the text and URL of Web pages [2]. The information of HTML page corresponds to the text label. Before extracting the content, HTML parser is used to parse the corresponding HTML page into DOM tree. According to the setting of extracting rule, the corresponding head and foot of label are found to extract the content. The extraction process based on DOM pages is shown in Fig. 2.

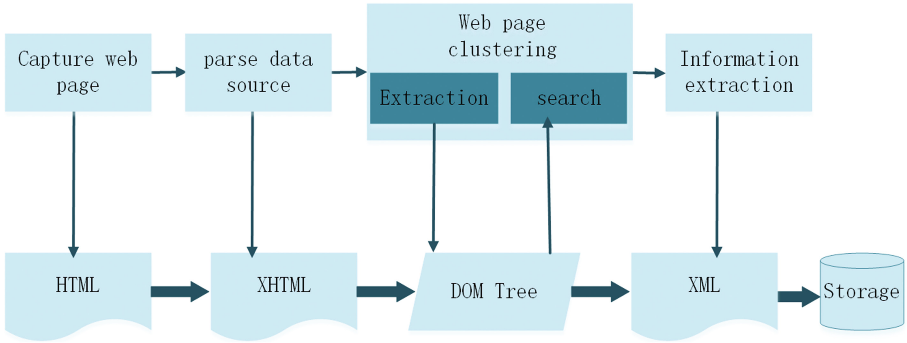


Fig. 2. Extraction process of web information

The Web extraction based on DOM [3] mainly includes four steps: firstly, the Web pages are gotten and the data sources are arranged, where the data sources are mapped to XHTML. The unpaired labels are added with the terminator "/"; the attributes in the label are added with the double quotes; "\" in URL is converted into "/". Secondly, the data source is parsed and XHTML file is constructed as DOM tree. The main process is that all start tags of Web page are recorded in the tag list at first. Then all tags of Web page are found and the end tags and comment tags without the start tags are removed. Repeat the above steps till the treatment of all tags on Web page. In this way, a table with tags and contents is formed, where the whole tree is divided into n sub-trees in the list. Thirdly, DOM is adopted to determine the similarity between the gathered Web page and the sample structure. In case of high similarity, it is called the Web page generated by the same template, namely these Web pages share the public DOM tree. The pattern of similar sample is used to extract the information of Web page. Fourthly, nodes in DOM are converted into XML file and the extraction results are saved in the form of structured storage.

The algorithm for the replicated documents elimination is based on the analysis of documents [4]. The design principle of replicated documents eliminator is that, before and after the punctuation mark in the document to be checked, two Chinese characters or characters are selected as the features. These features are then merged by the order of position in the document as the token. Such token is compared with the document in the library. If the token of some document has the certain proportion of continuously same characters from the token of such document, it can be judged that such document is duplicated in the library and the further treatment of such document is stopped. Due to the timeliness, a piece of news may be posted or reproduced for three days at most.

There are few pieces of news that can last for over 7 days. Despite the certain adjustment on the title, first and last paragraphs after being reproduced, there will be no impact on the treatment of duplicated documents elimination. Thus a document within one week in the library is selected for the comparison during the duplicated documents elimination.

2.2 Index Creation and Material Search

In such system, the index is created using ElasticSearch, where an index is divided into 6 index shardings. All these shardings are uniformly distributed in the valid nodes of cluster to form the distributed structure, in order to reduce the burden of each node [5]. Besides, the copy is set up for each index sharding. It is 1 for the copy in such system, namely the default quantity. When some index sharding is invalid, the copy can be used to recover the corresponding data of this sharding. ElasticSearch is capable to find the new nodes in time and make the certain adjustment on the data, as well as detecting the invalid nodes. When finding the invalid nodes, the data will be automatically distributed to the available nodes. To further manage the cluster in ElasticSearch, the third-party plugin Head is employed to operate and manage the cluster through the visual interface.

The configuration of index mainly covers the factors of index type, parser and storage policy, where the parser includes the analyzer and pre-filter.

The index type is of critical importance for the treatment of fields. The information in such system is in Chinese and the type is the string field. During the parsing, each character is treated as one unit to be parsed. The parser is the core of indexing mechanism, which plays the critical role in how to treat the fields when being indexed or searched. It also includes the analyzer and pre-filter. The original character string is split into the smallest basic unit for indexing. The accuracy of analyzing has the direct impact on the accuracy of search. The default method by ElasticSearch is the one with the unit of characters, but the degree of search matching and search speed cannot meet the demands of news gathering and editing system. IK analyzer developed by the third party is adopted in this system [6]. Then the preprocessor is used for the preprocessing of these basic units, such as removing the disable characters. If the field is only used for the search but not the display, its storage is set as false. In this way, such field will not be saved in the system, which can reduce the space of indexing.

The main steps for the search using ElasticSearch are as follows:

(1) IK analyzer is used to parse the key words that the user inputs.
(2) MultiFieldQueryParser is used to generate Query object, which allows the search of multiple key words on multiple index columns.
(3) The example of IndexSearch is created at first and the distributed search is performed on each node of ElasticSearch. Then the results of each node that meet the requirements are combined, sorted and then saved in the result set of SearchResponse.
(4) ObjectMapper is used to transform JSON data into List pattern. It is then returned to the result list and displayed to the user.

3 Results Design of Gathering Structure

3.1 Analysis on Gathering Results

There is the certain replication for the data gathered by web crawler. The effect is compared and analyzed before and after the replicated documents elimination. The comparison mainly includes:

(1) The same piece of news is posted on three web sites. The gathering is performed one time on three web sites respectively, where the number of such news in the library is counted.
(2) The same piece of news is posted on three web sites. The gathering is performed one time every 6 h on three web sites, with 5 times of gathering in total. The number of such news in the library is counted.
(3) The similar piece of news is posted on five web sites. The gathering is performed one time on five web sites, where the number of such news in the library is counted. The following table gives all Comparison of Gathering Effect.

According to Table 1, the first case is analyzed. Because of no replicated documents elimination, the same piece of news gathered from three web sites are saved in the library, which result in 3 times of news in the library. After the replicated documents elimination, it only counts one time in the library. For the second case, the same piece of news is included in three web sites. In the condition of news gathering at 5 different times, before the replicated documents elimination, there are 13 times in the library. The reason may be that some web site deletes such news and thus it is not 15 times in the library (if the news is not deleted, the theoretical number should be 15 times in the library). For the third case, before the replicated documents elimination, these 5 pieces of news are gathered in the library. After the replicated documents elimination, two pieces of similar news are in the library, because though these two pieces of news are similar, there is the great difference in the content description.

Table 1. Comparison of gathering effect

No.	Before replicated documents elimination	After replicated documents elimination
1	3 times	1 time
2	13 times	1 time
3	5 times	2 times

3.2 Analysis on Search Results System Design

The storage of gathered data is the system database and the index constructed by ElasticSearch. To analyze the search effect by the index of ElasticSearch, the comparison is performed on the response time of search with the increase in the data volume. The volume of experimental test data is 20000 pieces, 100000 pieces and 200000 pieces respectively. MySQL and ElasticSearch are used to search the key word of “Ji Nan” twice in the same data and the ratio of search time by ElasticSearch to the search time by MySQL is calculated, with the experimental results shown in Table 2.

Table 2. Comparison of search speed

Data Volume / Times	20000	100000	200000
1	1: 12.1	1: 9.5	1: 22.6
2	1: 29.3	1: 31.4	1: 101.2

There are two important conclusions from the experiment. The first conclusion is that, with the increase in the volume of experimental data, in the case of search in the same order, the ratio of search speed by ElasticSearch to the search speed by the database becomes smaller. It indicates that with the increase in the search speed by ElasticSearch, it becomes faster when being compared with the search speed by the database. When the volume of data is the same, the search speed by ElasticSearch for the second time becomes slower than the one by ElasticSearch for the first time. It is because ElasticSearch increases the technology of searching the cache. Therefore, when searching the large volume of data, it will be more advantageous when searching by ElasticSearch.

4 Conclusion

The specific works and the demands of news gathering, editing and posting on e23.cn are carefully investigated in this paper and then the news gathering and editing system based on the idea of media convergence is designed and realized. By referring to the successful cases of news gathering and editing system at home and abroad, it intends to expand the gathering channel of news material and news transmission, relieve the burden of news staff and increase the efficiency of news production.

References

1. Jiangwen, S., Kong, X., Song, L.: Non structured data unified storage platform. Comput. Modernization **11**, 69–73 (2016)
2. Shi, R.: A new approach to content extraction from web page. Telecom World **19**, 210–211 (2015)
3. Narawade, S.M., Prabhakar, N.M., Maruti, N.S.: A web based data extraction using hierarchical (DOM) tree approach. Int. J. Innovative Res. Sci. Technol. **2**(11), 255–257 (2016)
4. Xiong, Z., Ya, M., Zhang, Y.: Detection and elimination of similar web pages based on text structure and string of feature code. J. Comput. Appl. **33**(2), 554–557 (2013)
5. Gormley, C., Tong, Z.: Elasticsearch: The Definitive Guide. O'Reilly Media, Sebastopol (2015)
6. Meng, B., Wang, Z.: Two chinese word segmentation algorithms in the realization of the cloud computing platform. Netw. Secur. Technol. Appl. **12**, 67–71 (2014)

Evacuation Guidance Method by Using Augmented Reality and Digital Signage

Kazuyoshi Hikita(✉) and Hiroyoshi Miwa

Graduate School of Science and Technology, Kwansei Gakuin University,
2–1 Gakuen, Sanda-shi, Hyogo, Japan
{hikita,miwa}@kwansei.ac.jp

Abstract. Rapid and appropriate evacuation is important to decrease victims caused by disasters when a natural disaster strikes. The previous evacuation guidance determines evacuation routes beforehand. However, when a large-scale disaster occurs, streets may be blocked for passing through because of blaze or ruin of building and tidal wave, and congestion may be caused because of obstruction. Furthermore, communication infrastructure may broken down at the time of the particularly large-scale disaster, and victims are isolated from communication. An evacuation guidance indicating fixed evacuation routes is not useful in such cases, because an evacuation route is not always passable. Therefore, an evacuation guidance determining appropriate direction of movement rapidly according to road condition, position of evacuation center and the number of evacuees is needed. In this paper, we propose a method for evacuation guidance indicating the appropriate direction of movement by using smart phone with the function of augmented reality and digital signage in a poor communication environment such as in the event of a disaster, and we evaluate the performance of the method by simulation.

1 Introduction

Many disasters resulting in serious damage, such as the Great East Japan Earthquake in 2011, have occurred recently. This enhances the need of disaster risk reduction method. Consequently, there are a number of studies for disaster risk reduction and emergency management. They point out the utilization of advanced information and communication technologies are critical (ex. [1]).

It is especially important to navigate evacuees to evacuation centers quickly in order to decrease the number of victims when a disaster strikes. In general, the previous evacuation guidance determines evacuation routes beforehand [2–4]. However, when a large-scale disaster occurs, streets may be blocked for passing through because of blaze or ruin of building and tidal wave, and congestion may be caused because of obstruction. Furthermore, communication infrastructure may broken down at the time of the particularly large-scale disaster, and victims are isolated from communication. An evacuation guidance indicating fixed

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_32

evacuation routes is not useful in such cases, because an evacuation route is not always passable.

In recent years, research and development of DTN (Delay - and Disruption - Tolerant Networking) has advanced globally as a mean of communication in poor communication environments such as in the event of a disaster or communication congestion. Furthermore, evacuation guidance methods using the store-and-carry-forward routing, one of the DTN technology, have been proposed [5–7]. There is a method, which is similar to the compass routing method, for evacuation guidance in the event of a disaster [8,9]. This is to prompt an evacuee to move in the directions of the evacuation centers by showing the directions and distance information superimposed on the surrounding scenery taken by the camera of the mobile terminal of the evacuee. The information is displayed on the mobile terminal as the augmented reality (AR) screen overlaid with the positions and directions of the evacuation centers. Although this method can expect effectiveness at the time of a disaster, since it shows only the directions and distances of the evacuation centers, it is not possible to indicate the information about impassable points and congested points.

In this paper, we propose a method for evacuation guidance indicating the appropriate directions of movements by using smart phone with the function of AR and digital signage in a poor communication environment such as in the event of a disaster, especially for situations with high population density of evacuees like daytime urban areas. The evacuees in such a situation do not know the evacuation route and a number of evacuees causes congestion; therefore, such a situation needs an effective evacuation guidance method. We evaluate the performance of the proposed method by simulation and show that the method works better than the previous method in the various conditions.

2 Evacuation Guidance Method by Using Augmented Reality and Digital Signage

In this section, we propose a method for evacuation guidance.

It is assumed that the mobile terminal such as a smart phone periodically collects surrounding map information and caches it in normal times. In the mobile terminal used by an evacuee, the current position information of the evacuee is acquired by GPS, and the direction of the evacuation center as the destination is displayed on the screen based on the position information of the evacuation centers acquired in advance. In addition, beacons installed in digital signage and traffic lights are also used. The beacon has a communication function and various sensors. It senses the surrounding and gathers congestion status and impassable points at the time of a disaster. The information is shared among beacons via a communication line such as satellite communication or ad-hoc network. Since traffic lights are installed at intersections on the road, the information collection by the sensors is relatively easy. Similarly, since digital signage is installed at a place with good visibility for that purpose, it is considered that it is effective as an information collection and communication base. It is also possible to install an emergency battery to activate the necessary guidance function even when a power failure occurs. The sensor information collected by the beacon is sent to the center by the communication function. When the communication line is disconnected, it is also possible to add functions to continue communication by DTN technology such as ad-hoc networks. The collected information is sent to all beacons, and the beacon broadcasts to the mobile terminals of the surrounding evacuees by a wireless local area network; in addition, the information is visually indicated by the display function of the digital signage (including traffic lights hereafter) itself. We show the above basic idea in Fig. 1.

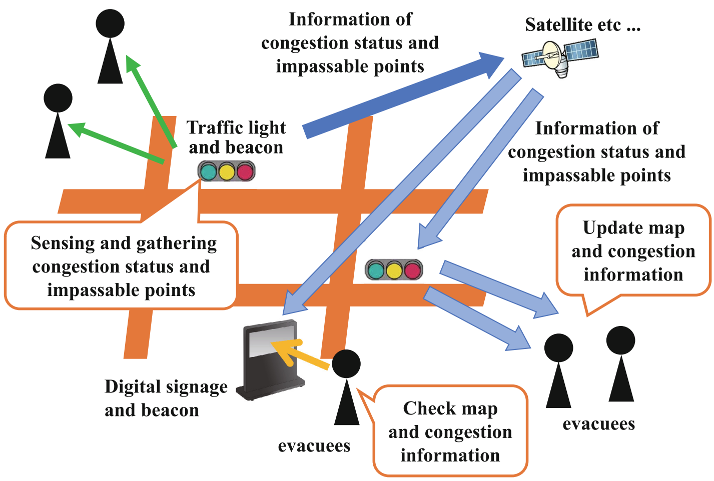


Fig. 1. Collection and distribution of congestion and road condition information based on digital signage and traffic lights.

The mobile terminal reflects the received congestion information and the impassable points on the cached map information. The information is displayed on the mobile terminal as the augmented reality (AR) screen overlaid with the position and direction of the evacuation center (Fig. 2).

Next, we describe movements of evacuees assumed by the above evacuation guidance method. When an evacuee has a mobile terminal, the evacuee gets the information displayed on the mobile terminal (the augmented reality screen overlaid with the position and direction of the evacuation center). In the example of Fig. 3, the evacuee selects the direction indicated by the right arrow, which is the previous method in the reference [9], because the direction that the angle between road and the direction to the evacuation center is smallest is likely to be selected psychologically. However, if the congestion situation of the adjacent intersection is also given to the evacuee as information by a beacon around the evacuee, the evacuee avoids the congestion and select the road indicated by the downward arrow (Fig. 4).

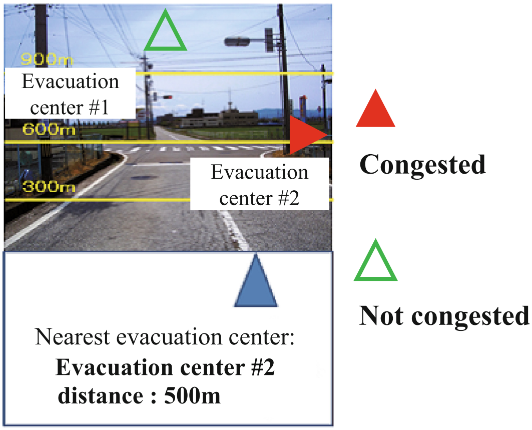


Fig. 2. Example of the display on a mobile terminal.

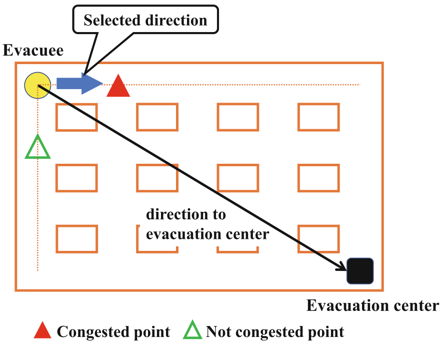


Fig. 3. Evacuation route of evacuee by the previous method.

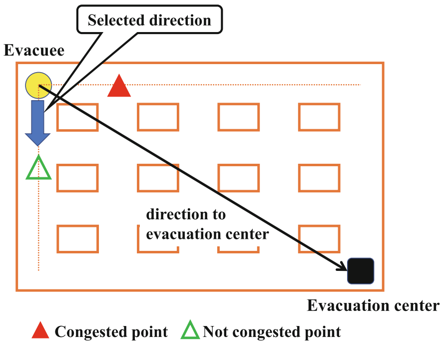


Fig. 4. Evacuation route of evacuee by the proposed method.

When there are multiple evacuation centers, the congestion situation is also indicated and evacuees select based on the distances to the evacuation centers and the congestion situation.

When an evacuee does not have a mobile terminal, the evacuee gets the information displayed on a digital signage around the evacuee and moves the same as the case that an evacuee has a mobile terminal. When an evacuee does not have a mobile terminal and there is no digital signage around the evacuee, the evacuee follows the movement of other evacuees having mobile terminals.

3 Performance Evaluation

3.1 Human Mobility Model

We describe the human mobility model in the following performance evaluation. When an evacuee has a mobile terminal and there is a beacon around the evacuee, the evacuee determines the evacuation direction probabilistically based on the congestion information, and proceed. When an evacuee has a mobile terminal and there is no beacon around the evacuee, the evacuee determines the evacuation direction probabilistically based on only the information of the position and direction of the evacuation center. When an evacuee does not have a mobile terminal and there is a digital signage around the evacuee, the evacuee moves the same as the case that an evacuee has a mobile terminal. When an evacuee does not have a mobile terminal and there is no digital signage around the evacuee, the evacuee follows the movement of other evacuees having mobile terminals probabilistically. When an evacuee does not have a mobile terminal and there is no digital signage around the evacuee, if the evacuee does not follow the movement or there is no evacuee having a mobile terminal around the evacuee, the evacuee randomly determines the direction. If an evacuee reaches a impasse, the evacuee returns the way till other routes are found.

3.2 Simulation

We use a grid graph as the topology of a road network. A vertex corresponds to the center of a section defined by segmenting a road at regular intervals, a branch, and an intersection. An edge interconnects the centers of the sections. The vertex capacity is the maximum number of persons that can be accommodated in a section defined at regular intervals. The edge capacity is the maximum number of persons who can pass through the edge during a unit time. The initial supply of a vertex is the number of persons at the vertex at time 0. The additional supply of a vertex is the number of persons who appear at the vertex from adjacent building, etc. during a unit time. The leaving of a vertex is the number of persons who leave the vertex to an adjacent building, evacuation center, etc. during a unit time. It is assumed that a person who has arrived at the evacuation center leaves the network, that is, they enter an evacuation center.

We assume that there are the traffic lights with beacon at all intersections and all evacuees having the mobile terminals can get the information about the congestion status and impassable points.

We evaluate the performance of the proposed method in Sect. 2 and the method without digital signage and beacon. The method without digital signage and beacon is equivalent to the method of the reference [9]. We call it the previous method hereafter.

In all the following experiments, the total of the initial supply is 2000, the total of the additional supply is 10000, and the total of the leaving is 0 and these supplies are randomly allocated to all vertices.

First, we use the complete grid graph whose vertex capacity is 50 and edge capacity is 40 in Fig. 5. A black vertex indicates an evacuation center. We assumed that all evacuees have the mobile terminal in this experiment.

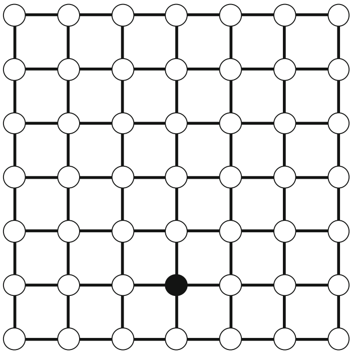


Fig. 5. Road network 1.

We show the results in Figs. 6 and 7. Figure 6 shows the evacuation completion time in each simulation in the road network of Figs. 5 and 7 shows the distribution of evacuation time in the road network of Fig. 5.

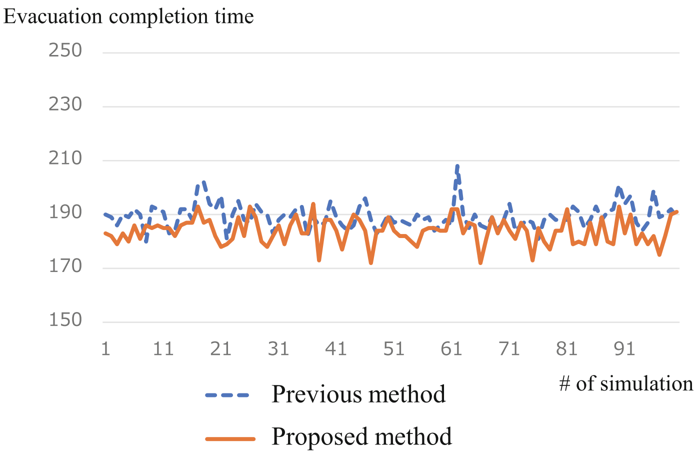


Fig. 6. Evacuation completion time in each simulation in the road network of Fig. 5.

We can see that the evacuation completion time of the proposed method is almost always less than that of the previous method in Fig. 6; the evacuation time of most evacuees is small, but we cannot ignore the number of evacuees whose evacuation time is large in the both methods in Fig. 7.

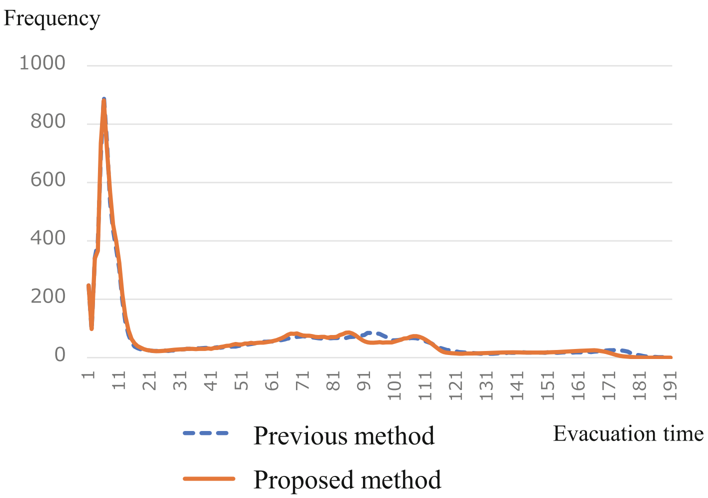


Fig. 7. Distribution of evacuation time in the road network of Fig. 5.

The average evacuation completion time of the proposed method is 183.8 and the average evacuation time of all evacuees is 47.43. On the other hand, the evacuation completion time of the previous method is 189.3 and the average evacuation time of all evacuees is 50.01. The performance of the proposed method is better than the previous method. This is because the map and congestion information distributed from the beacon in the digital signage (including traffic light) is effective to find a better evacuation route.

Next, we observe the influence of impasses. In this experiment, we used the partial grid graph in Fig. 8.

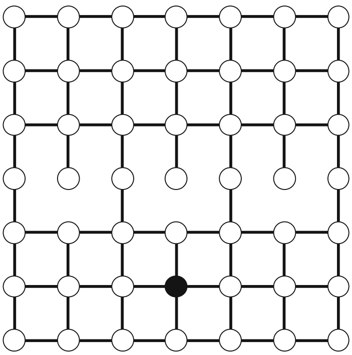


Fig. 8. Road network 2.

We show the results in Figs. 9 and 10.

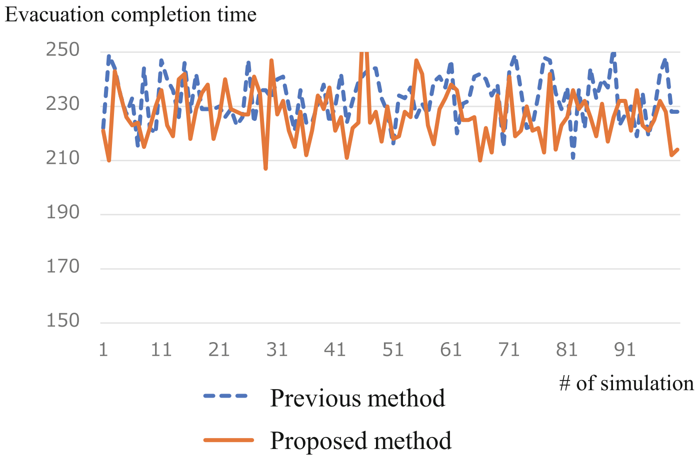


Fig. 9. Evacuation completion time in each simulation in the road network of Fig. 8.

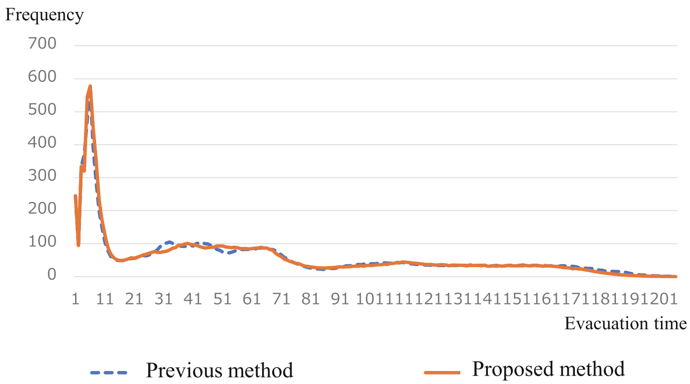


Fig. 10. Distribution of evacuation time in the road network of Fig. 8.

We can see that the evacuation completion time of the proposed method is almost always less than that of the previous method in Fig. 9; the evacuation time of most evacuees is small, but we cannot ignore the number of evacuees whose evacuation time is large in the both methods in Fig. 10.

The average evacuation completion time of the proposed method is 226.8 and the average evacuation time of all evacuees is 56.75. On the other hand, the evacuation completion time of the method without digital signage and beacon is 233.1 and the average evacuation time of all evacuees is 59.77. The performance of the proposed method is better than the previous method. Since the evacuees move so as to avoid congestion in the proposed method, the evacuees move so as to avoid the dead end and the surrounding. However, it causes the concentration to the limited routes and the congestion cannot be avoided eventually. Therefore, it is difficult to shorten the evacuation time largely.

Next, we examine the case that no evacuees have mobile terminals having the function of the proposed method. We can observe whether evacuees can evacuate appropriately based on only the information indicated by digital signage and beacon. We used the road networks of Figs. 11 and 12.

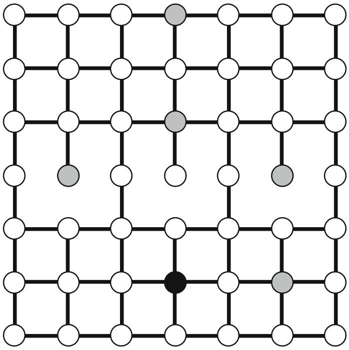


Fig. 11. The position of 5 displays of digital signage.

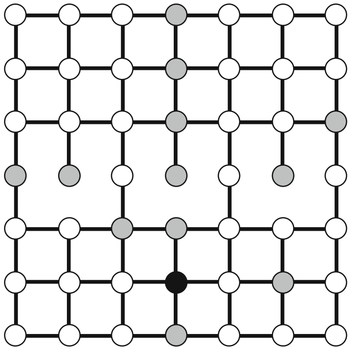


Fig. 12. The position of 10 displays of digital signage.

In this experiment, we compare the cases that the number of the digital signage displays is 10, 5, and 0. The following figures show the position of digital signage with beacon. The gray vertices indicate the position of digital signage and beacon.

Figure 13 shows the evacuation completion time in each simulation where the number of displays of digital signage is 10, 5, 0. Since the evacuation completion time in the case that the number of displays of digital signage is 0 is too large, Fig. 14 focuses on the evacuation completion time where the number of displays of digital signage is 10 and 5. Figure 15 shows the distribution of evacuation time.

When there is no digital signage, the average evacuation completion time is 2063.75 and the average evacuation time of all evacuees is 155.62; when there are 5 digital signage displays, the evacuation completion time is 285.61 and the average evacuation time of all evacuees is 68.35; when there are 10 digital signage displays, the average evacuation completion time is 207.2 and the average evacuation time of all evacuees is 52.20. The evacuation time decreases rapidly according to the increase of digital signage displays. This result means that, even if the mobile terminals having the function based on the proposed method do not become widely used, the appropriate evacuation can be achieved by even only information indicated by digital signage.

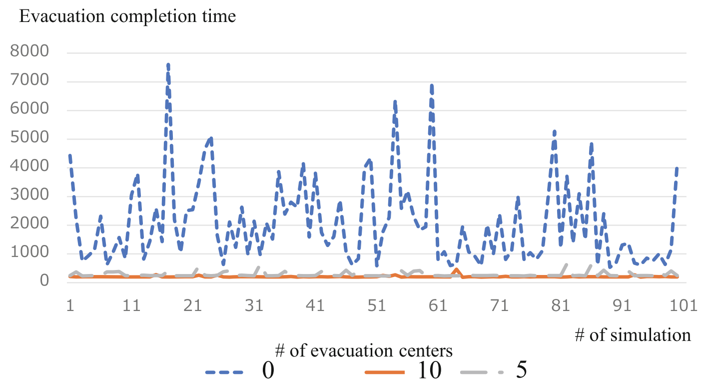


Fig. 13. Evacuation completion time in each simulation where the number of digital signage is 10, 5, 0.

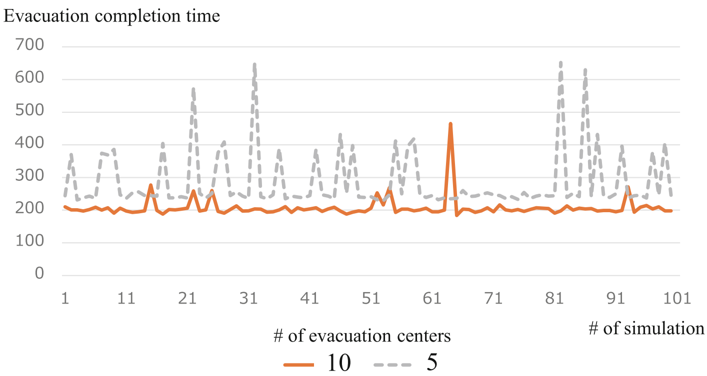


Fig. 14. Evacuation completion time in each simulation where the number of digital signage is 10 and 5.

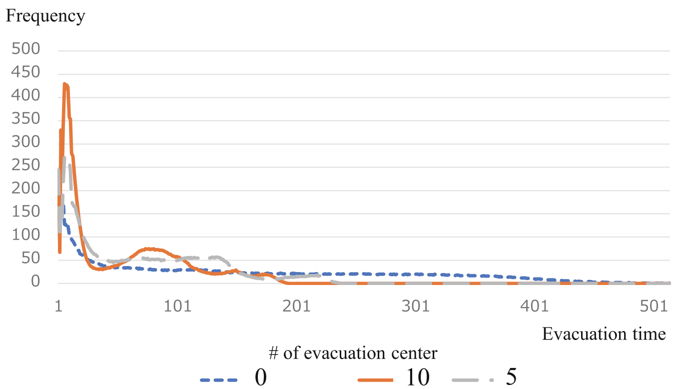


Fig. 15. Distribution of evacuation time where the number of digital signage is 10, 5, 0.

4 Conclusion

In this paper, we proposed the method for evacuation guidance indicating the appropriate direction of movement by using smart phone with the function of augmented reality (AR) and digital signage in a poor communication environment such as in the event of a disaster, especially for situations with high population density of evacuees like daytime urban areas.

We evaluated the performance of the proposed method by simulation. The results indicated that, even if there are dead ends in a road network, the proposed method using the digital signage and beacon is better than the previous method. In addition, we examined the case that no evacuees have mobile terminals having the function of the proposed method and observed whether evacuees can evacuate appropriately based on only the information indicated by digital signage. The results that the evacuation time decreases rapidly according to the increase of digital signage displays show that, even if the mobile terminals having the function based on the proposed method do not become widely used, the appropriate evacuation can be achieved by even only information indicated by digital signage. These results show that the map and congestion information distributed from the beacon in the digital signage is effective to find a better evacuation route and that the information displayed by digital signage is also effective to evacuation. Consequently, the proposed method is more effective than the previous method.

For the future work, it remains to investigate whether the evacuation is more effective by guiding individual people or a group of people, whether the influence of some people do not follow the recommended evacuation route is large, and whether the influence of the accuracy of the guidance is large.

Acknowledgment. This work was partially supported by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (B) (17H01742).

References

1. Bessis, N., Asimakopoulou, E., Norrington, P., Thomas, S., Varagamto, R.: A next generation technology victim location and low level assessment framework for occupational disasters caused by natural hazards. Int. J. Distrib. Syst. Technol. **2**(1), 43–53 (2011)
2. Ford, L.R., Fulkerson, D.R.: Flows in Networks. Princeton University Press, Princeton (1962)
3. Hamachar, H.W., Tjandra, S.A.: Mathmatical modeling of evacuation problems: a state of the art. In: Pedestrian and Evacuation Dynamics, pp. 227–266 (2002)
4. Hadzic, T., Brown, K.N., Sreenan, C.J.: Real-time pedestrian evacuation planning during emergency. In: Proceedings of 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 597–604, 7–9 November 2011
5. Fujihara, A., Miwa, H.: Real-time disaster evacuation guidance using opportunistic communication. In: Proceedings of IEEE/IPSJ-SAINT2012, pp. 326–331 (2012)

6. Fujihara, A., Miwa, H.: Effect of traffic volume in real-time disaster evacuation guidance using opportunistic communications. In: Proceedings of IEEE-INCoS2012, pp. 457–462 (2012)
7. Fujihara, A., Miwa, H.: On the use of congestion information for rerouting in the disaster evacuation guidance using opportunistic communication. In: Proceedings of ADMNET2013, IEEE 37th Annual Computer Software and Applications Conference Workshop, pp. 563–568 (2013)
8. Kranakis, E., Singh, H., Urrutia, J.: Compass routing on geometric networks. In: Proceedings of 11th Canadian Conference on Computational Geometry (1999)
9. http://bousai.or.jp/mitamita/mitamita03/

Evaluation of How Methods for Creating People Flow Data Affect Performance of Epidemic Routing

Yusuke Sakai[1] and Akihiro Fujihara[2(✉)]

[1] Fukui University of Technology, 3-6-1 Gakuen, Fukui-shi, Fukui 910-8505, Japan
lucky39@mx5.fctv.ne.jp
[2] Chiba Institute of Technology, 2-17-1 Tsudanuma,
Narashino, Chiba 275-0016, Japan
akihiro.fujihara@p.chibakoudai.jp

Abstract. Recently various datasets regarding human mobility have been available for research purpose. People Flow Project by Center for Spatial Information Science provides people flow data of Japanese urban areas created from Person Trip Census data. There exist some data in which the number of points of departure and arrival is only one for each representative postal address although real human mobility traces does not have this property. It comes from the reason that Person Trip Census does not include coordinates of longitude and latitude information, but does only representative postal addresses. When we consider performance evaluation of information dissemination by epidemic routing, effects of methods for creating the data on simulation results have not been well considered. In this study, we investigate this effects through creating three types of people flow data using Fukui's Person Trip Census to compare the performance of epidemic routing. The first data is that the points of departure and destination are assigned only one for each representative postal address, and nodes expressing persons or cars move linearly between them. The second data is that the points of departure and destination are the same with the first one, but each node moves along a shortest-path route on the road network. The third data is that the points of departure and destination are assigned to randomly selected points on the road distributed around the addresses and each node moves along a shortest path route. We perform numerical simulations of epidemic routing where a selected node initially have a message to spread to other nodes within a communication distance. We evaluate the impact on temporal epidemic size by changing the communication distance. As a result, we find that when the distance is sufficiently short, the speed of dissemination in the third data becomes much slower than those of the first and second ones. This result indicates we must be careful to use people flow data for evaluating the performance because a illusory pattern that a large number of mobility traces gather into the representative points of departure and arrival cannot be negligible.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_33

1 Introduction

As information communication technologies advances, Data Science (or Data-driven Science) has been taken attention and various datasets can be used for research purpose [1–3]. Although we need to be deeply careful about privacy issues, various datasets regarding human mobility patterns have also been gradually available. People Flow Project [2] by Center for Spatial Information Science (CSIS) in the University of Tokyo has provided people flow data of 21 urban areas in Japan and four cities abroad for researchers who have agreed collaborative reasearch with CSIS. People flow data consist of a vast amount of human mobility traces in a day created from Person Trip (PT) Census Data [4]. PT Ceusus has been done by every Japanese prefecture about once in every ten years to understand how people in Japan move in daily life for applications to urban planning and considering local economics. In the survey, a questionnaire is sent to more than ten thousands of citizens to request their answers to fill when and where they go and stay for how many times in a day. The answers are gathered by prefectural governments and they are aggregated into a dataset. CSIS has been created people flow data of all major urban areas using PT Census. About Fukui prefecture, no people flow data are available even though there exists PT Census data.

Some existing future internet architectures, such as Delay Tolerant Network (DTN) [5,6] and Mobile Opportunistic Network (MON) [7,8], have been proposed and extensively studied in the context of Internet of Things (IoT) [9–11]. Some of the situations use human mobility to carry and forward messages to other nodes, which is called Store-carry-forward routing scheme. Therefore, human mobility datasets are useful to consider performance evaluations of these network systems. Fujihara has investigated the proper communication distance for epidemic routing [12], which is one of famous routing protocols in DTN, using SNS-based people flow data in major Japanese urban areas including Tokyo metropolitan, Chukyo, and Kansai areas [13]. Because network connection between mobile nodes in DTN stochastically changes by their mobility and proximity conditions, information dissemination processes behave like a dynamic percolation along the road network. The result by numerical simulations for evaluating the performance indicates that there exists a scaling relation between the total number of nodes and the proper communication distance to disseminate almost all the nodes in the urban area within one day. This result indicates universality regarding the performance of DTN and MON independent of urban areas. Since universal properties in cities have been discovered in the field of urban studies [14], it is natural to consider the necessary condition to connect distributed peers in future internet architectures beyond IoT generation.

There exist some people flow data in which the number of points of departure and arrival is only one for each representative postal address although real human mobility traces never have such a property. This is because PT Census does not include coordinates of longitude and latitude, but does only postal addresses. When we consider performance evaluation of information dissemination by epidemic routing using such people flow data, effects of methods for creating the data on epidemic routing have not been well considered.

In this study, we investigate the effects of creation methods of people flow data on the performance through creating Fukui's PT Census Data. We create three types of people flow data to compare information dissemination patterns in epidemic routing. The first data is that the points of departure and destination are assigned only one for each representative postal address and nodes expressing persons or cars move linearly between them. The second data is that the points of departure and destination are the same with the first one, but each node moves along a shortest-path route on the road network. The third data is that the points of departure and destination are assigned to a randomly selected points on the road distributed around the representative postal addresses and each node moves along a shortest path route. We perform numerical simulations of epidemic routing where a selected node initially have a message to spread to other nodes within a communication distance. We evaluate how big is the impact on temporal evolution of the epidemic size by changing the communication distance of mobile nodes. As a result, we find that when the communication distance is sufficiently short, the speed of information dissemination by the third data becomes much slower than the first and second ones through nonlinear fitting of logistic curve. This result indicates that we should be careful to use human mobility datasets for evaluating epidemic routing if they have a illusory property that a large number of mobility traces gather into a small number of representative points of departure and arrival.

2 Related Works

There are various existing studies on routing protocols in DTN and MON. Epidemic routing is one of famous routing protocols in DTN [12] which a node forwards possessing message copies to multiple nearby nodes when it comes within the communication distance. Successive forwarding by nodes on the network finally leads to information dissemination to all the nodes that there has a chance to encounter. Therefore, statistical properties on human serendipitous encounter [15,16] resulting from human mobility patterns is important to understand the performance of DTN in general.

Fujihara has investigated the proper communication distance for epidemic routing using SNS-based people flow data in major Japanese urban areas [13]. Fujiwara et al. have also investigated pandemic phenomena and how contagious diseases spread especially in Tokyo metropolitan area [17]. Both the above studies concluded that there exists a non-trivial scaling relation regarding epidemic final size. Universal properties in cities have been discovered in the field of urban studies [14], which is consistent with universal scaling properties on human encounter.

Since DTN is considered to be useful to communicate at disaster situations, studies on epidemic routing in urban areas can be helpful to consider disaster evacuation guidance using opportunistic communication [18–22].

3 Person Trip Census Data

We contacted to Department of City Planning in Fukui Prefectural Government for providing Person-trip Census Data of Fukui. In fukui prefecture, the census was done on 15 November 2015 [4]. The total number of surveyed persons is 43,700 and in total, 110,821 records are gathered. Each record consists of information about person, postal addresses of departure and arrival on the trip, and method of transportation. The original PT Census Data is a sequence of digits as shown in Fig. 1. Each row indicates a single record. For each column in the record, there is one-to-one correspondence between digits and the trip information. For example, postal addresses are expressed by corresponding zone codes and they are summarized in six-digit zone code table as shown in Fig. 2 on the left. To generate coordinates of longitude and latitude, therefore, it is necessary to convert postal addresses to the coordinate information, which is called *geocoding*. Geocoding is provided by web API services. We used Geocoding.jp [23] to do this and the coordinate information can be summarized as shown in Fig. 2 on the right.

```
010010101051115010121015731010121110020100210000001
010010101051115010121015731010121110020100210000001
010010101051115010121015731010121110020100210000001
010010101051115010121015731010121110020100210000001
010010101051115010121015731010121110020100210000001
010010102081115010121025231010121100020100210000001
```

Fig. 1. Original person trip census data.

City name	Zone code	address
Fukui	010111	3, Ote
Fukui	010121	1, Ote
Fukui	010122	2, Ote
Fukui	010131	1, Chuo
Fukui	010132	2, Chuo
⋮	⋮	⋮

Address	Zone code	longitude	latitude
3, Ote, Fukui	010111	136.218927	36.064323
1, Ote, Fukui	010121	136.22684	36.066486
2, Ote, Fukui	010122	136.224185	36.064531
1, Chuo, Fukui	010131	136.220397	36.061245
2, Chuo, Fukui	010132	136.218969	36.058873
⋮	⋮	⋮	⋮

Fig. 2. Table for converting zone code to postal address (left) and geocoding to convert from postal addresses to coordinates of longitude and latitude (right).

4 Methods for Creating People Flow Data

Based on the created dataset explained in the previous section, we interpolated human motility traces by one-minute temporal resolution. We created three types of people flow data: line, road, and random data. We introduce how to create these data and the difference of them in the following subsections. To show the boundary of cities in Fukui prefecture, furthermore, we used the boundary data provided by [24]. We divided the original boundary data into those of every cities in the north part of Fukui prefecture for helping visualization where moving traces are as shown in Fig. 3.

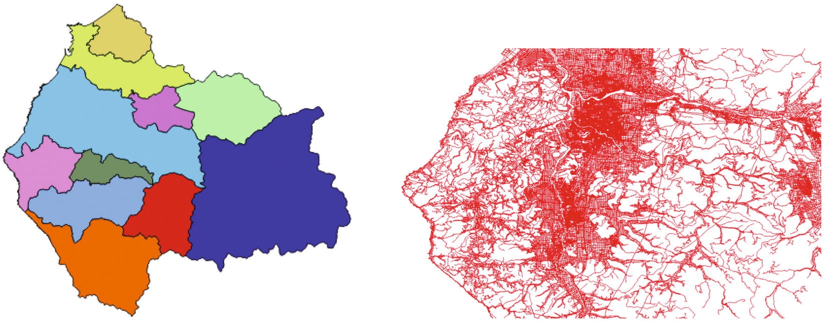


Fig. 3. Visualizations of geographical boundaries (left) and the road network (right) in the north part of Fukui prefecture.

4.1 Line Data

The first data is created to interpolate moving traces by one-minute temporal resolution on the straight line between representative points of departure and arrival. We write a program to automatically calculate moving traces by reading person ID, coordinates of longitude and latitude for the representative points, and datetime to fit the traces on the straight line. The description of line data is shown in Fig. 4.

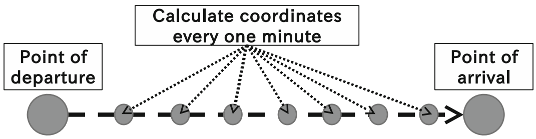


Fig. 4. Description of line data.

4.2 Road Data

The second one is that the points of departure and arrival are the same with the first one, but each node moves along a shortest-path route on a road network. Here, we used Openstreetmap [25] as the data for the road network. The road network in the north part of Fukui prefecture is visualized using QGIS [26] as shown in Fig. 3.

Every representative point of departure and arrival is assign to the nearest point on the road network. We also interpolate moving traces as a shortest-path route using Dijkstra's algorithm [27] and we generate moving points for each trace with one-minute temporal resolution along the network as shown in Fig. 5.

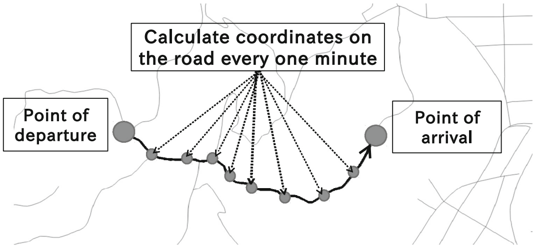


Fig. 5. Description of road data.

4.3 Random Data

The third data is that the points of departure and destination are distributed on a randomly selected points on the road network distributed around the representative postal addresses and each node moves along a shortest path route. In this case, we uniformly distribute these points in the circle with radius 3 km whose center is the original representative points of departure or arrival as illustrated in Fig. 6.

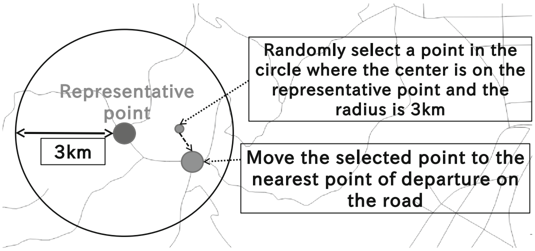


Fig. 6. Description of random data.

As well as the second data, we interpolate moving traces as a shortest-path route and generate moving points for each trace with one-minute temporal resolution along the network.

For these three types of data, we create records whose attributes are use ID (uid), timestamp, and coordinates of longitude and latitude as shown in Table 1. The random data is the most plausible of all because of the method to generate origin-destination point distributions and mobility traces between them by observation. So, we go public with the random data on our website only for research purpose by courtesy of Department of City Planning in Fukui Prefectural Government [28]. For the random data, the total number of records reach 2,047,417 and the number of user IDs is $N = 34{,}255$ in total. PT report says that 70% of all the traffic flow in Fukui is moving by car. This means that the method in which we interpolate mobility traces along the road network works well. Potential missing of this mobility dataset is that the original PT data do not include mobility patterns that come into and go out from Fukui. Also, mobilities by train are not taken into consideration in random data.

Table 1. Sample records and their attributes in people flow data

uid	Timestamp	Latitude	Longitude
1	2005-11-15 09:00:00	36.066486	136.22684
1	2005-11-15 09:30:00	36.054067	136.24709
1	2005-11-15 13:00:00	36.054067	136.24709
1	2005-11-15 13:30:00	36.091209	136.032421
1	2005-11-15 17:00:00	36.091209	136.032421
⋮	⋮	⋮	⋮
1	2005-11-15 20:35:00	36.066486	136.22684
2	2005-11-15 05:00:00	36.041229	136.202362
⋮	⋮	⋮	⋮

5 Simulations of Epidemic Routing

Using human mobility datasets explained in the previous section, we perform numerical simulations of epidemic routing where a selected node initially have a message to spread to other nodes. For each node, there are two states: susceptible and infected. In the susceptible state, the node have no message. In the infected state, the node have the message. We define the communication distance d. If a susceptible node comes close to a infected node within the distance d, the infected node forwards the message to the susceptible node. Then, since the node has the message, the susceptible one becomes infected. We change the value of d and evaluate the speed of information dissemination later. This message forwarding (infection) process is illustrated in Fig. 7.

To perform numerical simulations of epidemic routing, it takes much time (more than one day) to find pairs of susceptible and infected nodes within d

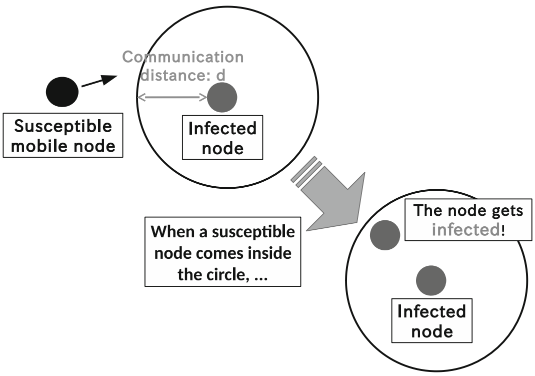


Fig. 7. Message forwarding (infection) process.

because the number of nodes N is too large. In this case, estimated computational time is $O(N^2)$, where N is about 30,000 nodes. Therefore, we consider parallel (multi-core) computing to find the pairs at close range. To do this, we used multiprocessing package of Python language which divides CPU processes into pieces to compute parallely with multiple CPUs. We wrote the program to divide Fukui areas into 100 square areas to compute in parallel for finding user ID of the pairs to forward the message [29]. We use a workstation to compute and visualize simulations of epidemic routing. The workstation specification is as follows: Intel Xeon CPU E5-2643 v3 3.40 GHz $\times$ 24 cores, and DDR4 8 GB $\times$ 16 Memory. We simulated one scenario of epidemic routing using ten CPU cores. As a result, we observed by running our multiprocessing program, the computational time for simulation scenarios successfully reduces to about four hours on average. This means that the computational time becomes one sixth of the time that it originally takes. A snapshot of the numerical simulation of epidemic routing is visualized in Fig. 8.

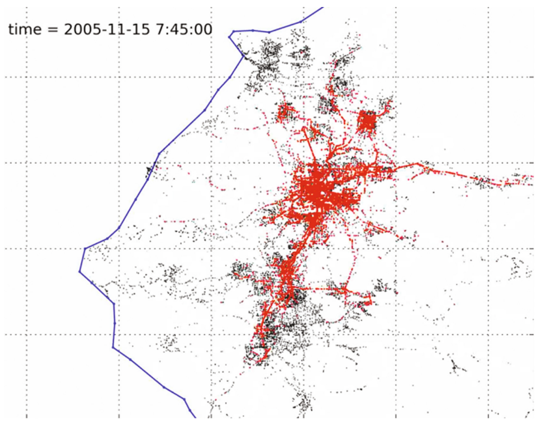


Fig. 8. A snapshot of epidemic routing simulation in Fukui.

Next we consider evaluation measures of the performance for epidemic routing. We consider temporal change of epidemic size (meaning the total number of infected nodes) as the performance measure. We evaluate how big is the impact on temporal epidemic size by changing types of data and the communication distance d. The process of epidemic routing can be interpreted as a epidemic model frequently used in epidemiology called SI model. S means susceptible and I means infected. The temporal evolution of the system without considering spatial factor is described by the following ordinary differential equation with a conservation law.

$$\frac{dS(t)}{dt} = -\alpha SI, \qquad S(t) + I(t) = 1, \tag{1}$$

where t is time, $S(t)$ is the tempral evolution of susceptible state in the system normalized by one, $I(t)$ is that of infected one, and α is a contact rate. By solving the differential equation, it is easy to follow that the time evolution of SI model is described by the following logistic curve having three fitting parameters.

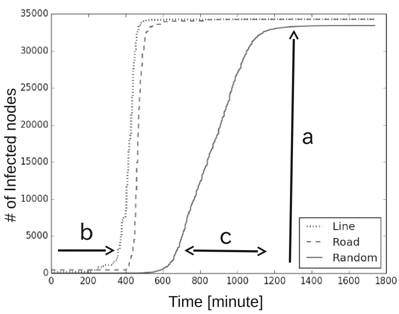
$$S(t) = \frac{a}{1 + b\exp(-ct)}, \tag{2}$$

where a indicates the final epidemic size, b indicates the latency of outbreak, and c is the inverse of relaxation time of epidemic spreading. By applying nonlinear fitting of logistic curve to the time evolution of epidemic size by simulation, we can estimate the parameters a, b, and c for each scenario to compare the speed of information dissemination by epidemic routing. We used nonlinear fitting package in R language to fit these parameters.

6 Results

To perform the simulations, we select one node in Fukui city central randomly as the initial condition. We change the communication distance $d = 5\,\text{m}$ and $d = 1.5\,\text{km}$ and compare three types of people flow data. The time evolutions of epidemic size in three types of data is shown in Fig. 9. As explained the above, the time resolution of people flow data is one minute. The horizontal axis indicates time with minute in a whole day. So, it ranges from zero to $60 \times 24 = 1440\,\text{min}$.

As shown in these figures, both the time evolution curves in the cummulative epidemic size in Fig. 9 on the left looks similar to the logistic curve Eq. (2) for all the three types of data. In Fig. 9, the epidemic size of line and road data drastically increases at commuting rush hours in the morning. Line data tend to show the fastest drastical increase of all. However, the time evolution of random data does not increase in the morning, but gradually increases from noon to another commuting rush hours in the evening. This difference can be evaluated by the result of nonlinear fitting of logistic curve in Table 2. This result



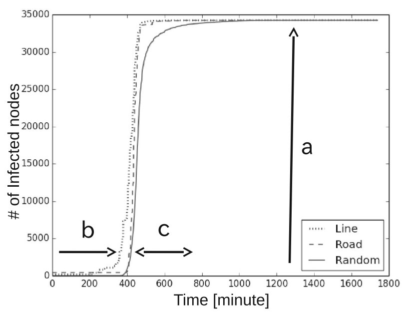


Fig. 9. Time evolution of epidemic size in three types of people flow data with $d = 5\,\text{m}$ (left) and $d = 1.5\,\text{km}$ (right).

indicates that when the communication distance is sufficiently short, the speed of information dissemination indicated by parameter c^{-1} gets much slower in random data compared to the others.

In Fig. 9 on the right, contrary, the epidemic size of all the people flow data drastically increases at commuting rush hours in the morning. But, the latency to the increase seems to be slightly different between them. Line data tend to show the fastest and random data to show the slowest. This difference can be evaluated with parameter b by the result of nonlinear fitting of logistic curve in Table 2.

Table 2. Results of nonlinear fitting of logistic curve to estimate the parameters.

Data type	Line		Road		Random	
Comm. dist. d	5 m	1.5 km	5 m	1.5 km	5 m	1.5 km
a	34302.15	34301.12	34268.97	34281.17	33638.75	33945.13
b	3641539	5426531	49126721	23218283	4070.359	30672989816
c	0.0376094	0.03862014	0.04277773	0.04163123	0.01015323	0.0529438

These results indicate that we should be careful to use human mobility datasets for evaluating epidemic routing if they have a illusory property that a large number of mobility traces gather into a small number of representative points of departure and arrival like line and road data.

7 Discussion

We investigated the effects of the generation methods of people flow data on the performance of epidemic routing simulations through creating Fukui's people flow data from PT Census Data of Fukui prefecture. Because PT Census records only postal addresses as the representative points of departure and arrival for privacy reasons, we should interpolate motility traces in addition to the original data. We considered the methods for creating three types of people flow data. The first data is that the points of departure and destination are placed only one for each representative postal address and nodes move linearly between them. The second data is that the points of departure and destination are the same with the first one, but each node moves along a shortest-path route on the road network from Openstreetmap dataset. The third data is that the points of departure and destination are randomly distributed to a selected point on the road distributed around the representative postal addresses and each node moves along a shortest path route.

We perform numerical simulations of epidemic routing where a selected node initially have a message to spread to other nodes within the communication distance. We evaluate the performance using temporal evolution of epidemic size to consider how the impact of how to create people flow data on temporal epidemic size by changing the communication distance. As a result of nonlinear fitting of logistic curve, we find that when the communication distance is sufficiently short, the speed of information dissemination by the third data becomes much slower than the first and second ones. This result indicates that we should be careful to use human mobility datasets for evaluating epidemic routing if they have a illusory property that a large number of mobility traces gather into a small number of representative points of departure and arrival by the property of PT Census Data, which is the contribution of this work.

As a by-product of this investigation, we have created Fukui's people flow dataset (meaning random data) for the first time. This data is open to the public only for research purpose for free thanks to Department of City Planning in Fukui Prefectural Government where we have been provided Person-trip Census Data of Fukui. Even though the original PT Census was done in 2005, which is more than ten years ago, most mobility patterns in Fukui is by car and the number of roads does not change so much compared to that in ten years ago. So, the authors think the created dataset is useful to estimate human motility patterns in Fukui. If you want to use the data, please go to check the specification of the dataset [30] and contact the authors by email written in this paper.

Simulation results in Fig. 9 and Table 2 show that wherever the initial infected node is selected around the Fukui city center, it finally spreads to almost all the mobile nodes. This is similar in the previous study [13] and it supports the possibility of universal scaling relation between the number of nodes and the proper communication distance. For future work, it is interesting to consider how the method of creating people flow data on the scaling relation.

Since we have no dataset about where residential areas for each cities, we randomly and uniformly distribute the representative points of departure and arrival in the circle area. The radius 3 km is selected for only the reason that the average size of areas indicating the postal address is within 3 km. For future work, it is challenging to consider any evaluation method for how plausible created people flow data are. It is important to try creating more realistic human mobility datasets by combining as many datasets as possible, for example, data of population and residential areas in cities.

We have only simulate epidemic routing, but there are many variants of store-carry-forward routing scheme. It is also interesting to simulate other routing protocols in DTN, for example, Spray and Wait, PROFET, MaxProp, and some other delegation routing methods.

Acknowledgements. This work was partially supported by the Japan Society for the Promotion of Science (JSPS) through KAKENHI (Grants-in-Aid for Scientific Research) Grant Number 17K00141. This research was also supported by the joint research with Center for Spatial Information Science (CSIS), the University of Tokyo (No. 555). About the latter support, the authors thank Drs. Naoya Fujiwara and Hiroshi Kanasugi for useful discussion and comments on the generation method of people flow data. The authors also thank Department of City Planning in Fukui Prefectural Government for providing us Person Trip Census Data of Fukui.

References

1. Blondel, V.D., et al.: A survey of results on mobile phone datasets analysis. EPJ Data Sci. **4**, 10 (2015). http://www.epjdatascience.com/content/4/1/10
2. Peple Flow Project by Center for Spatial Information Science (CSIS). The University of Tokyo. http://pflow.csis.u-tokyo.ac.jp/?page_id=943
3. Nightley and Center for Spatial Information Science (CSIS), SNS-based People Flow Data. http://nightley.jp/archives/1954
4. Summary of The third Person-trip Census Survey of Fukui Prefectural Urban Area and its data specification (Japanese). http://info.pref.fukui.jp/toshi/pt/etsuran2/sentaku/ryuuiten/tebiki.pdf
5. Vasilakos, A.V., Zhang, Y., Spyropoulos, T., Networks, D.T.: Protocols and Applications, Wireless Networks and Mobile Communications Series. CRC Press, Boca Raton (2012)
6. Rodrigues, J. (ed.): Advances in Delay-tolerant Networks Architecture and Enhanced Performance. Woodhead Publishing, Sawston (2014)
7. Denko, M.K.: Mobile Opportunistic Networks: Architectures, Protocols and Applications. Auerbach Publications, Boca Raton (2011)
8. Woungang, I., et al.: Routing in Opportunistic Networks. Springer, New York (2013)
9. Future Smartphones Wonft Need Cell Towers to Connect. http://www.technologyreview.com/news/530996/future-smartphones-wont-need-cell-towers-to-connect/
10. Mascetti, S., Bettini, C., Freni, D., Wang, X.S., Jajodia, S.: Privacy-aware proximity based services. In: Tenth International Conference on Mobile Data Management: Systems, Services, and Middleware, pp. 31–40 (2009)
11. Lin, X., Andrews, J.G., Gosh, A., Ratasuk, R.: An overview of 3GPP device-to-device proximity services. IEEE Commun. Mag. **52**(4), 40–48 (2014)
12. Vahdat, A., Becker, D.: Epidemic routing for partially connected ad hoc networks, Technical report CS-2000-06. Duke University, Department of Computer Science (2000)
13. Fujihara, A.: Estimating proper communication distance for epidemic routing in Japanese urban areas using SNS-based people flow data. In: Proceedings of 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), pp. 1–8. IEEE (2016)
14. Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies. Penguin Press (2017)
15. Fujihara, A., Miwa, H.: Homesick Lévy walk: a mobility model having Ichi-Go Ichi-e and scale-free properties of human encounters. IEEE COMPSAC **2014**, 576–583 (2014)
16. Fujihara, A.: Analyzing scale-free property on human serendipitous encounters using mobile phone data. In: MoMM 2015, pp. 122–125 (2015)
17. Fujiwara, N., et al.: Contribution of hidden modes to nonlinear epidemic dynamics in urban human proximity networks. arXiv:1512.01901 (2015). http://arxiv.org/abs/1512.01901
18. Fujihara, A., Miwa, H.: Real-time disaster evacuation guidance using opportunistic communication. In: IEEE/IPSJ-SAINT2012, pp. 326–331 (2012)
19. Fujihara, A., Miwa, H.: Effect of traffic volume in real-time disaster evacuation guidance using opportunistic communications. In: IEEE-INCoS2012, pp. 457–462 (2012)

20. Fujihara, A., Miwa, H.: On the use of congestion information for rerouting in the disaster evacuation guidance using opportunistic communication. In: IEEE 37th Annual Computer Software and Applications Conference Workshop, ADMNET 2013, pp. 563–568 (2013)
21. Fujihara, A., Miwa, H.: Disaster evacuation guidance using opportunistic communication: the potential for opportunity-based service. In: Big Data and Internet of Things: A Roadmap for Smart Environments, Studies in Computational Intelligence, vol. 546, pp. 425–446 (2014)
22. Fujihara, A., Miwa, H.: Necessary condition for self-organized follow-me evacuation guidance using opportunistic networking. In: Intelligent Networking and Collaborative Systems (INCoS), pp. 213–220 (2014)
23. Geocoding API. http://www.geocoding.jp/api/
24. Japan Basemap information downlaod service (Japanese). http://fgd.gsi.go.jp/download/menu.php
25. OpenStreetMap. https://www.openstreetmap.org/
26. Welxome to the QGIS project! http://qgis.org/en/site/
27. Rossant, C.: IPython Interactive Computing and Visualization Cookbook. O'Reilly Media, Sebastopol (2014)
28. Fukui Prefectural Government. http://www.pref.fukui.lg.jp/english/
29. Python Documentation, 17.2.multiprocessing – Process based parallelism. http://docs.python.jp/3/library/multiprocessing.html#multiprocessing.Process.join
30. Sakai, Y., Fujihara, A.: Fukui Mobility Visualization Project. https://sites.google.com/site/akihirofujihara/fukuimobilityvisualization

IBM Case Analysis by Servitization in IT Industry

Lukas Bickel[1] and Natalia Kryvinska[2(✉)]

[1] Vienna University of Economics and Business (WU Wien), Vienna, Austria
lukas.bickel@s.wu.ac.at
[2] Department of Information Systems, Faculty of Management,
Comenius University in Bratislava, Bratislava, Slovakia
Natalia.Kryvinska@fm.uniba.sk

Abstract. The trend of Servitization has been rising through different industries and has changing them continuously. Due to this, companies undergo major adaptations and transformations. And, an examination of this transition is very important for the accurate enterprise resource planning, as well as business success in general. However, the research on Servitization is mostly conceptual. An empirical investigation is demanded in order analyze the challenges by real companies. For this reason, we perform a case study of the Servitization within an IBM, one of the pioneers in this process in the IT industry. An analysis of the progress in the Servitization process of the company is done by evaluating the transition progress from a product organization to a product-service organization. In the end, short summary and possible future research topics are given.

Keywords: Enterprise · IT industry · Servitization · IBM scenario

1 Introduction

Analysis of the Servitization process development in a company can be done by examining the transition progress from a product organization to a product-service organization. A. Tukker proposed in 2004 a model, in which he provided a product-service-system (PSS) by establishing three basic classifications. These PSS could be product-oriented, use-oriented, or result-oriented [1]. Afterwards A. Neely added integration-oriented- and a solution-oriented- approaches in his research on the financial consequences of Servitization in 2008 [2]. And other scholars in [3] afterwards designed a new illustration with an explicit analysis of the various elements in 2014 (Fig. 1).

In the first phase a company focuses still only on the selling of tangible goods, which are supported by services as add-ons and ownership changes at the point of sale. This is in correlation with the vertical integration of manufacturing companies [1–3].

In the second phase companies are still geared towards selling products but some extra services are added, which are closely related to the product but not a necessity in order to use the product [1–3].

In the third phase services are added with no choice of deciding to use the service or not. This leads to a bundling of product and service during an offer [1–3].

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems,*
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_34

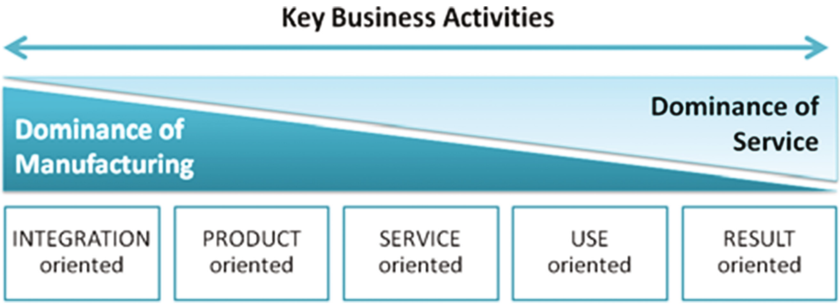


Fig. 1. Servitization transition approaches [3].

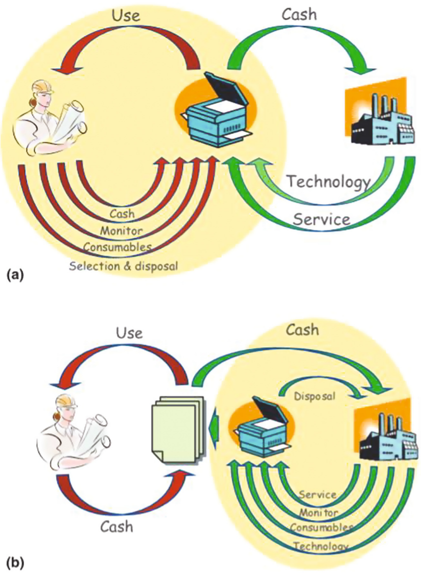


Fig. 2. Switch from ownership (a) - to access (b) by [4, 5].

In the use-oriented phase, the business model switches away from selling products as the product stays in ownership of the provider. This switch implied by Baines et al. [6] is illustrated in Fig. 2.

Tukker divides this phase in three additional types of use-oriented services. The first type is the product lease, where the customer leases the right to use the product. In the second type the product-renting customer rents the right for a certain time to use to product and partly has to share the product with other users. In the last type the product is pooled to a big user base and is mostly used simultaneously [1].

In the last phase the customer and the provider agree on a certain result, which has to be provided. This solution-based approach is the final stage of Servitization. Tukker subdivides this phase also in three types. Firstly, the outsourcing type, which includes outsourcing of whole departments or activities. In the second type a customer agrees to pay according to a certain output of the product and a certain level of usage. The third type bases on the result and the provider is allowed to deliver the result in any given way [1–3].

2 Case Study – IBM Transition Towards Servitization

In this Section we perform an IBM [7] Servitization scenario analysis. We examine a business segment and a revenue distribution within the changes assessment, as well as challenges and positive/negative effects within the transition analysis.

2.1 Changes Assessment

Business Segment

During our observation period (2005–2015), IBM business segments remain the same, but the priority was shifted. In the beginning, there was only a global services segment, which included every service delivered. Over the decade IBM started focusing more and more on services, and split the global services into two business segments. Although IBM had already a strong service oriented PSS (Product Service System), it increased the efforts in certain areas. And, it is getting more towards a result-oriented PSS with its outsourcing [8–18].

In the Fig. 3, we expose results of the change analysis by IBM Servitization transition in the period between 2005 and 2015. Some terms used on the figure are as follows: GBS (Global Business Services), GFS (Global Financial Services), GTS (Global Technology Services), S&H (Systems and Hardware), S&T (Systems and Technology).

In the Figs. 4, 5 and 6, we present the results of our analysis of IBM Servitization transition by three main segments – hardware, software, and services. The analyzed time-period, decade 2005–2015, showed a shift of all segments towards (IT) Products/Services.

As illustrated in Figs. 4, 5 and 6, IBM has decreased the products and services they offer. Most of IBM hardware business was disinvested, as they did not include it in their new business models. IBM software offering has increased over the decade. And now, it includes advanced innovations for further rise towards a result-oriented PSS. Additionally, it is to mention that almost 50% of the software revenue IBM earns is delivered via their cloud service or on demand. This is further evidence of a switch from selling products to selling services. Through the process from offering individual services to an all-round solution package, many of the services offered over the decade have been bundled and technically integrated into each other [8–18].

Revenue Distribution

Further, we perform an assessment of the revenue distribution - using available industrial sources (Figs. 7 and 8). The revenue comparison over the decade displays already the importance of services for IBM in 2005, as IBM was one of the key leaders of the Servitization initiative in the IT sector. By 2015, IBM increased their percentage of services, although to mention, the main services global technology and global business services have been stagnated [8–18].

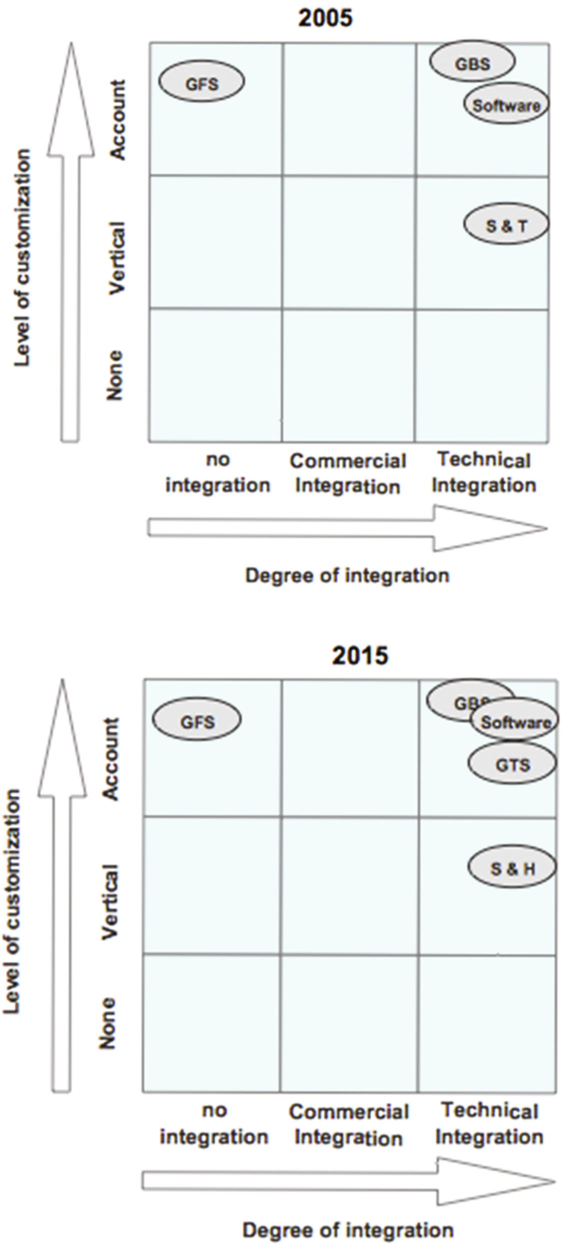


Fig. 3. IBM servitization transition – changes analysis (Own illustration according to [8–18])

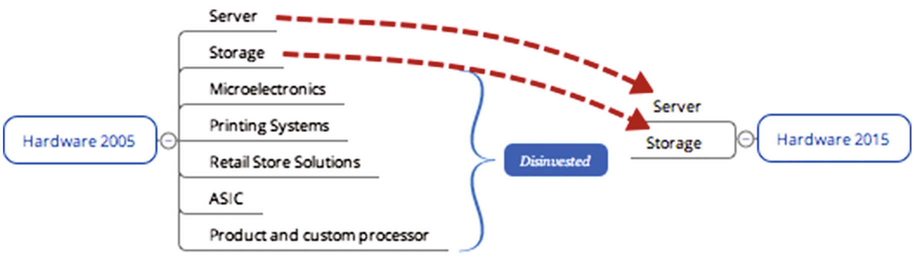


Fig. 4. Hardware offering (own illustration according to the IBM annual reports 2005 and 2015) [8–18].

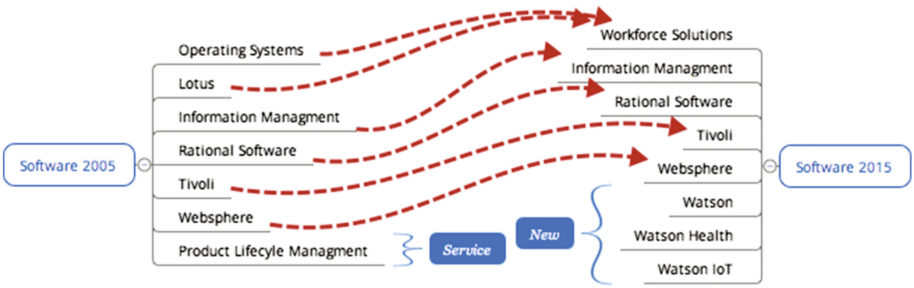


Fig. 5. Software offering (own illustration according to the IBM annual reports 2005 and 2015) [8–18].

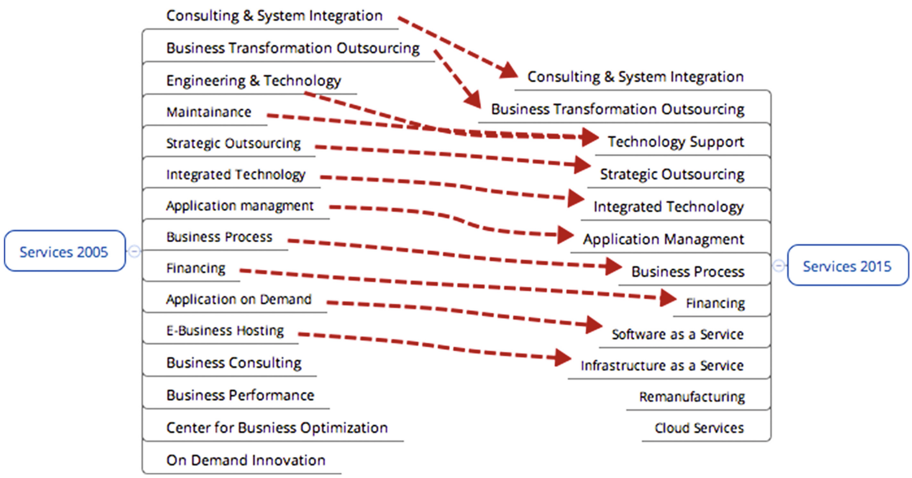


Fig. 6. Service offering (own illustration according to the IBM annual reports 2005 and 2015) [8–18].

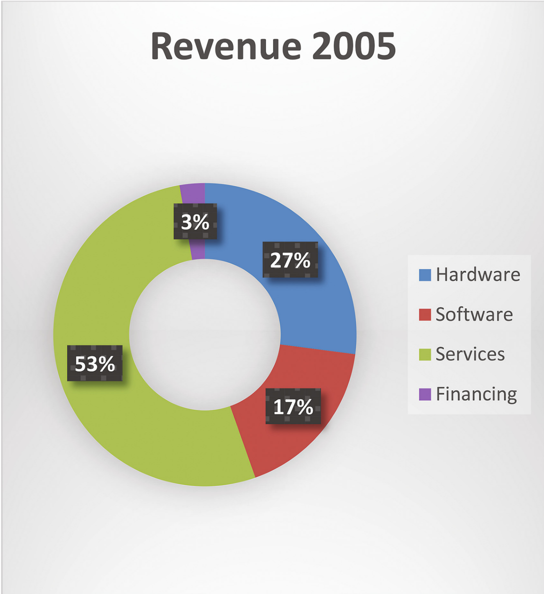


Fig. 7. Own illustration of the IBM revenue 2005 [8].

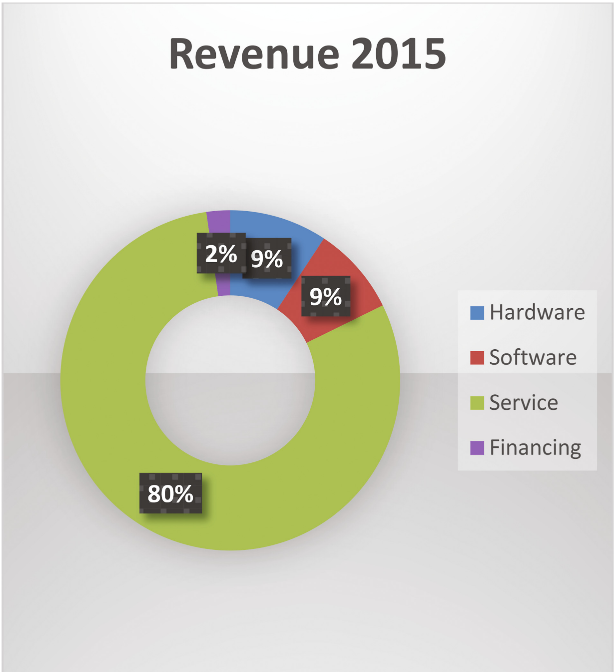


Fig. 8. Own illustration of the IBM revenue 2015 [18].

The reason of so high percentage of services offered in 2015 is an adaptation of the Cloud concept "Software-As-A-Service" in recent years, practically starting pioneering projects last decade.

2.2 Transition Analysis

Challenges

A one of the most challenging topics for the IBM was the switch of their sales model, as they had to change from their usual long-term contracts to a customer wish of short duration contracts also with more flexibility. This has been proven by the duration of contracts that decreased drastically over the 10-year period.

Another challenge to face was - the fast slowing-down of their hardware business, which was a major reason to switch to an alternative Cloud concept "Infrastructure-As-A-Service", as customers were spending less on hardware. This, in turn, has resulted in a huge divesture of their hardware segment, as can be seen in the Fig. 8, and a drop in the revenue of 50%.

IBM also had to rethink their acquisitions and focuses in R&D, which was a huge transition as they have invented important hardware in the past. Now, they are investing overall mostly in software and services. This is represented by the expenditure allocation of R&D and also in the amount of patents in the segment applied.

With an introduction of the Cloud era, IBM also has had to face a challenge of the new way how businesses use software or hardware. IBM has been adapting its vision and shift of its focus again to an even more result-oriented PSS. This happened through a shift to a further service of data analytics [8–18].

Positive and Negative Effects

As with every innovative business model, there are positive but also negative effects. The Servitization has been affecting the IBM, as shown in the Fig. 9, in the way IBM is

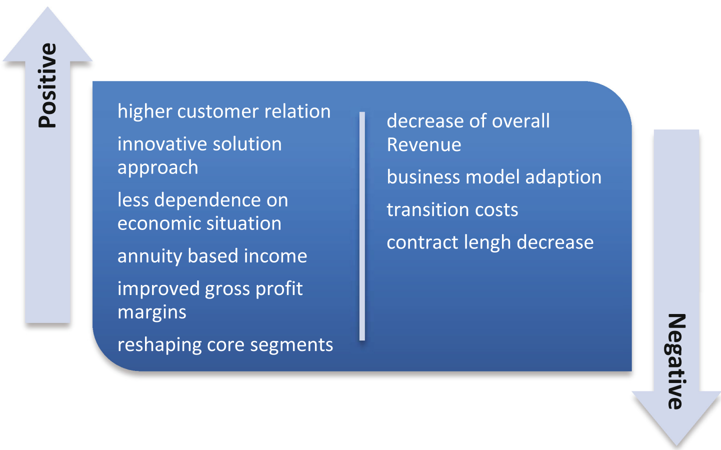


Fig. 9. Positive/Negative effects of servitization [8–18].

doing business, as well as how it reaches its customers [8–18]. The same can be applied to other major IT players.

The servitization transition in general encompasses alterations to the business logic, resources usage, as well as companies' structural design and their business processes performance. Thus, it should be seen as a strategic choice to boost the competitiveness through product differentiation in the service marketplaces.

3 Conclusions

Servitization tend to rise with a transition involvement of various industries and their continuous changes. The consideration of this transition is important for an accurate enterprise resources planning and business success at large. Nevertheless, Servitization studies are mainly conceptual. Pragmatic research is needed to examine the challenges of the real-world companies. Therefore, we conducted in this paper an IBM case study on Servitization. We performed it by means of changes analysis in the business segment and the revenue distribution, as well as through the transition analysis with its challenges and positive/negative effects.

We plan to continue our research on the Servitization in the IT industry. In particular, we intend to perform the case studies of further major players from the branch. Additionally, we intend to perform a comparative analysis of challenges and effects of the Servitization within IT industry, using conducted case studies.

References

1. Tukker, A.: Eight types of product–service system: eight ways to sustainability? Exp. SusProNet. Bus. Strat. Env. **13**, 246–260 (2004)
2. Neely, A.: Exploring the financial consequences of the servitization of manufacturing. Oper. Manag. Res. **1**, 103–118 (2008)
3. Kryvinska, N., Kaczor, S., Strauss, C., Gregus, M.: Servitization strategies and product-service-systems. In: 2014 IEEE World Congress on Services, pp. 254–260 (2014)
4. Baines, T., Lightfoot, H.: Made to Serve: How Manufacturers can Compete Through Servitization and Product Service Systems. Wiley, Chichester (2013)
5. Britchfield, G.: State-of-the-art in product-service systems (2013). https://uldamienobrien.wordpress.com/2013/10/12/state-of-the-art-in-product-service-systems/
6. Baines, T., Lightfoot, H., Benedettini, O., Kay, J.: The servitization of manufacturing: a review of literature and reflection on future challenges. J. Manuf Technol. Manag. **20**, 547–567 (2009)
7. IBM Corporation. http://www.ibm.com/us-en/
8. IBM Corporation: 2005 Annual report (2006). ftp://public.dhe.ibm.com/annualreport/2005/2005_ibm_annual.pdf
9. IBM Corporation: 2006 Annual report (2007). ftp://public.dhe.ibm.com/annualreport/2006/2006_ibm_annual.pdf
10. IBM Corporation: 2007 Annual report (2008). ftp://public.dhe.ibm.com/annualreport/2007/2007_ibm_annual.pdf
11. IBM Corporation: 2008 Annual report (2009). ftp://public.dhe.ibm.com/annualreport/2008/2008_ibm_annual.pdf

12. IBM Corporation: 2009 Annual report (2010). https://www.ibm.com/annualreport/2009/2009_ibm_annual.pdf
13. IBM Corporation: 2010 Annual report (2011). ftp://public.dhe.ibm.com/annualreport/2010/2010_ibm_annual.pdf
14. IBM Corporation: 2011 Annual report (2012). https://www.ibm.com/annualreport/2011/bin/assets/2011_ibm_annual.pdf
15. IBM Corporation: 2012 Annual report (2013). https://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf
16. IBM Corporation: 2013 Annual report (2014). https://www.ibm.com/annualreport/2013/bin/assets/2013_ibm_annual.pdf
17. IBM Corporation: 2014 Annual report (2015). https://www.ibm.com/annualreport/2014/bin/assets/IBM-Annual-Report-2014.pdf
18. IBM Corporation: 2015 Annual report (2016). https://www.ibm.com/annualreport/2015/assets/img/2016/02/IBM-Annual-Report-2015.pdf

Successful Social Network Integration in an Academic Institution: Case Study of Opportunities

Monika Dávideková[1(✉)] and Michal Greguš ml.[2]

[1] Department of Information Systems, Faculty of Management, Comenius University, Bratislava, Slovakia
Monika.Davidekova@fm.uniba.sk

[2] Department of Strategy and Entrepreneurship, Faculty of Management, Comenius University, Bratislava, Slovakia
Michal.Gregusml@fm.uniba.sk

Abstract. The development of information and communication technology (ICT) changed the world irreversibly and created new innovative ways of processing. Nowadays, people interact through social networks that allow creating and maintaining connections with their peers, relatives, colleagues, friends or partners. Social networks alternated the way people communicate with each other or reach out for information. They search for experiences of their connections, their recommendations, assessments or ratings of various objects like commercial offers, jobs and companies, schools, events, etc. People believe in statements and observations of their connections and act based on them. In particular young generation, millennials and later generations like Y, etc. do so to a large extent. Therefore it is important to be active through social networks also for academic institutions to become attractive and reachable to potential future students, employees, companies or partners and to be able to respond to their queries via their preferred communication channel.

There are various types of social networks designed for various purposes. This paper aims to analyze the opportunities of participation and interactivity of selected academic institution in various social networks. The analysis opts to review the importance of particular studied social networks targeting the point of view of people: students, employees, partners, etc. It has the ambition to provide proposals of possible motivating aspects for its integration into the portfolio of public relation management.

The intention of this paper is to clarify the importance of joining various social networks and to exploit their opportunities for an academic institution.

Keywords: Information and Communication Technology (ICT) · Social networks · Synergic effect · Innovative communication channels · Millennials

1 Introduction

With the rise of current digital era, the development of information and communication technology (ICT) new communication channels have emerged allowing real-time

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_35

communication and collaboration across the world [1]. It changed the world and our lives irreversibly beyond the recognition and became an integral part [2] of almost every activity we do.

ICT allows quick finding and distributing of information [3] by replacing physical “space of place” with virtual “space of flows” [4]. In the virtual environment that is accessible for everyone from everywhere due to ubiquity and omnipresence of ICT, the information exchange seems to happen almost instantly or in a relatively very short time - within only a few minutes or seconds [5] – enabling a real-time collaboration in one common virtual space [6, 7]. It removes spatial and timely barriers among collaborating parties [8] by enabling interaction in omnipresent virtual place accessible for everyone.

ICT enables fast communication over distant places for a large number of users with continuously increasing number of communicating users and devices over time [9]. No other former technology has had such an immense impact on our lives as the emergence of ICT [10]. It increased the efficiency of productivity as no previous technology before [11] and fosters further technological innovation, standardization and globalization [12] as well as the inter- and intra-enterprise collaboration [13] in today’s world characterized by dramatic turbulent changes in uncertain environment shaped by multicultural influences [14].

ICT changed almost revolutionary the way people interact [15, 16] and communicate transacting those activities in virtual environment that denotes a very important dimension of working nowadays [17]. It maintains the sense of presence and awareness of remote locations [18]. It provides the basis for computer applications to execute business processes [19] where the virtual communication becomes a central segment of everyday operation [20], not being a mere enabler anymore [21] and representing the dominant driver of output per worker it increases the capital per worker through which it contributes with relatively significant power to the support of the long-run economic growth [22].

The early signs of virtual collaboration can be tracked back into the late 80’s, to the development of a tool for sharing ideas at a distance represented by groupware [23, 24]. This system referred to “*as a piece of software with shared access to its data*” in the early stages evolved into a computer-based system with social group processes [25] denoting an important shift for the system utilization from being operated by one individual towards being practiced commonly by a whole group [1]. Another very important shift in the development of ICT is represented by the emergence of virtual teams where the team members not only share data, but also mutually participate on idea design and development through sharing knowledge, opinions, attitudes, experience and expertise. Along with the development of virtual teams, the groupware evolved into a computer-based system supporting groups engagement in common task execution and processing, now providing a multifaceted interface to a shared virtual environment [26] that gained considerable popularity as an instrument for knowledge sharing and knowledge management [27]. ICT wields a high potential for systemizing, facilitating, exploiting and expediting company-wide knowledge management [28] through virtual teaming. However, for any organization, the strength of social ties is critical to the success, since it determines how deeply individuals interact with each other [29].

ICT revolutionized the way we collaborate into new innovative ways of processing and human interaction. Nowadays, many people create and maintain connections through ICT tools via social networks with their peers, relatives, colleagues, friends or even partners. Social networks alternated the way people interact. They search for news, information, recommendations, assessments or ratings of various offers, jobs, companies, schools, events, etc. They express their political and religious opinions [30]. Adolescents seek social network games as popular online venues [31].

Like computer network that represents a set of machines connected by a set of cables, social network denotes a set of people or other social entities connected by a set of socially-meaningful relationships [32]. Boyd and Ellison [33] define social network sites as "*web-based services that allow individuals to: (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.*" The emergence of social networks can be tracked back to the first launched social network in 1997 and with the time the number of social network sites increased to hundreds of them [34].

People believe in experiences of their connections and act based on them. Today, an individual is not able to learn and experience everything and therefore, his/her connections, other people's experiences become surrogate for knowledge [35], experience and information. Especially the young generation, so called millennials and later generations like Y, teenagers etc. who became nearly inseparable from their cell phones [36] tend to act alike their peers.

According to a poll in 2010 in America, 22% of adolescents log on to their favorite social media site more than 10 times a day and more than half of teenagers do it more than once a day [37]. Thus, a large part of young generation's development is occurring whilst on the internet [38]. According to a European study conducted in all 28 countries of European Union in 2015, 78% of all individuals aged between 16 and 74 are Internet users [39].

Young people respond and react to posts in social network more frequently than to visit official pages. Therefore it is important to be active through social networks to become attractive to this young generation. This is true also for academic institutions that target potential future students and want to able to respond to their queries via the preferred communication channel of adolescents as those preferred social networking over face-to-face communication [40].

People's preferences change over time [41] and so also the inclination toward social network types. There are various types of social networks designed for various purposes. Another type is a professional network that allows creation and maintenance of collegial professional connections and professional relationships across organization [42] also after a switch to new company or department. Well connected people of a company within a social network are able to foster and maintain knowledge flows and their interdependence results in effective knowledge exploitation [43].

It also allows referrals to become more powerful with more detailed information that a brief curriculum vitae may provide. In a professional social network, there is the possibility to create alumni network that could bring graduates together across companies and provide information about the achievements and application of graduates of given institution on that way [44].

Another aspect for an academic institution is the research and scientific endeavor. There are already scientific or researcher's social networks that connect researchers devoted to the same scientific topic across the globe and fosters so the creation of new connections and collaborations. A scientific social network also offers the possibility of connecting companies and corporate researchers to academics [45].

This paper intends to analyze the importance of joining social networks for an academic institution. It aims to include also the point of view of students and graduates as well as employees and potential future attendants as the authors consider all to gain added value from it and therefore to represent stakeholders. This contribution targets to briefly outline opportunities offered by current existing social networks that may not be utilized or are not exploited in excessive manner at the moment in described elaborated case study tailored for chosen academic institution. However, many of provided implications are valid also for other institutions, yet the particular examples of analyzed social networks may change in name and in functionality slightly. The idea remains an analogue one.

This paper is organized as follows: Sect. 2 exemplifies various types of most spread social networks where each type is separately discussed in individual subsection. This section also outlines basic aspects and motivation why the institution shall consider its participation in particular social network with respects to provided gains and difficulties. The conclusion provides a brief summary of propositions provided in this contribution in corresponding parts developed in conducted discussion. Additionally it suggests further research topics for future analyses and explorations in given scientific area.

2 Exemplifying Social Network Types

In this section, various kinds of commonly spread social networks are briefly described. Services of social networks have been distributed and utilized worldwide [46]. Each one is analyzed in its subsection for the convenience of the reader by providing opportunities that a possible enrollment can provide as well as possible adverse effects of such integration are depicted. The provided names or examples of various types of social networks shall represent only concepts that are integrated in present social networks. Those can change at any time as online social networking technologies are being rapidly developed to meet recent trends [47] corresponding to user needs.

2.1 Private Social Network

Examples for private social networks are but not limited to Facebook [48], Path [49], My Space [50], Tagged [51], DiaSpora [52], WeChat [53] etc. The list reflects the currently used private social networks; however, it is subject to change at any time as the development is evolving those systems with time.

This type of social network targets private social connections among friends, relatives, people with similar hobbies, in other words connections with people we share our spare time with. Groups in social network sites denote an appropriate venue to discuss emotions and experiences [54]. People post their opinions, what they do or did, eat, wear, think, want, desire, etc. Social networking sites allow users to communicate with friends, produce and share information [55] as we are now online anywhere anytime [56] and can respond to each other.

This type of social network represents one of the first digital social networks that people enter, join and participate with. Therefore it is also a network where young generation socializes and maintains connections. Individuals read opinions and experiences of their peers and reach out through this communication channel to collect information. According to the results of European study conducted in 2015 [39], 80% of young people (thus four of every five) accesses the Internet on a daily base whereas the remaining population accesses the Internet only by 63%. Furthermore, young generation in the age from 16 to 19 denotes more than 90% of daily Internet users [57].

Targeted/Focused Group

This type of social network represents the communication channel towards young generation that represents potential future attendants of a university or other academic institution. Therefore it is important to create a tailored profile targeting young generation promoting activities that are attractive for millennials, generation Y, teens, pupils etc.

Type of Information to Include

Interesting and attractive information for such focusing group of readers are e.g.:

- Information about international exchange programs like Erasmus, student mobility, cooperation with foreign partner institutions or/and companies etc.
- Possibility of finding an internship or student part time job already during the studies.
- The successfulness of finding a position after graduation and a potential salary in given area.
- An overview or outline of a study program.

This targeted end group builds up their opinion based on experiences of their peers of the same age or around their age more than on numbers and official website of the university. Another considerable aspect is that the targeted group communicates more through pictures and videos more by utilizing text. Therefore, it would be recommendable to include stories told by the students themselves, videos that recap happenings at the university as well as the information about application procedure step by step and other information connected with it in video formats or in form of visual manuals.

This social network is also the communication channel for students during their university studies when the classes are over. It provides the possibility to build groups and to share information, books, reminders, experiences, etc. to any student with an account in the social network. Students will make appointments and learn together or share their opinions, views, works, projects and correct results to written exams and help also others to learn in this way. Such a social network opens the possibility to organize, divide and distribute labor, tasks etc. and to specialize. The group will therefore more efficiently use its own time to learn as fast as possible.

Students share almost everything, thus also the tests, questions they have gotten in a test, the correct answers in a test etc. Therefore it is becoming more and more difficult to create new tests and examination to analyze the knowledge of each one student. With the digital technology increasing speed of information and storage, it is no problem anymore to collect all previous tests and projects or works for all past years. And to learn solely the corresponding answers to conducted exams. This all increases the effort of every teacher or lecturer to objectively assess the work of a student and analyze whether he managed it all by himself/herself or copied amount of data from someone else's work. This even becomes more difficult when the individual just rewrites the context with synonyms to prevent automated check for plagiarism.

Even the evaluation of the knowledge of a given student becomes more difficult as the examiner has to assess how broad and deep is his/her knowledge within a limited time. The probability of having learnt solely some answers to common questions and having no idea of what the student is telling will be still considerably high.

All these and other aspects represent increasing threats of adverse effects of social networks on students.

2.2 Professional Social Network

Examples for professional social network can be, but are not limited to LinkedIn [58], Xing [59], Rally Point [60], Doximity [61], etc. Professional network is designated to connect colleagues within a company across departments or positions. People share their professional curriculums with professional experiences and accomplished educational backgrounds. As already mentioned before, the provided list is subject to change.

Participating in online interactivities provides benefits we cannot deny [62]. ICT and social networks have high potential for systemizing, facilitating, exploiting, expediting and managing company-wide knowledge [28] through virtual connections and teaming. For any organization, the strength of social ties denotes a critical factor to company success by determining how deeply individuals interact with each other [29]. A professional network targets connections and relationships among colleagues of one company or in given field. A newcomer who has just started in a new job can read something about his/her colleagues a see something about their history together with their photos attached to their names and not having to overcome the problems of not remembering everyone's name and not knowing how to address whom.

Targeted/Focused Group

Professional social network targets to connect professionals and experts across a company in given area. It provides profiles with pictures of colleagues for socializing. Professionals in terms of a university or other academic institution are represented by its employees and also by graduates. Therefore, the tailored profile in a professional social network will differ from a private social network in many aspects.

Type of Information to Include

Information that is attractive to employees and graduates:

- Current fields of research, present projects, collaborations and partnerships.
- Possibilities to collaborate with an academic institution for entrepreneurs, companies, corporations, organizations and other institutes.
- Achievements in the past, recognitions of the institutions and its students.
- An overview or outline of future development, strategic goals, mission, vision.
- Arrangements for graduates like alumni meetings, sponsoring, and possibilities of social engagement.
- Possibilities of further development like workshops, trainings, third age university or other specialization.

The targeted group disposes of two different segments: employees and graduates. For graduates, the reference of their alma mater may be a referral aspect too. Graduates are interested in alumni meetings information and well as about possibilities for further development. People visit the web pages of their absolved academic institution only from time to time with a very low frequency and therefore posts about possible alumni meetings in a professional social network where they are connected to the institution will have a far further extend as any announcement on the official web page.

Employees may connect and foster relationships across line boundaries and remain connected even after a switch to another institution. A professional social network may also serve as team web page for a virtual team. Individual team members have the possibility to gain some information about their co-workers even if they are miles away.

This targeted end group is used to work with official website without any hesitation and therefore a short trailer or notification in professional social network with a link to official page may be enough.

2.3 Scientific Social Network

Examples for scientific social network and reference sharing sites can be, but are not limited to ResearchGate [63], Academia.edu [64], Semantic Scholar [65], Mendeley [66] empowered by Elsevier BibSonomy [67], Zotero [68], Cite U Like [69] etc. The list of mentioned references denotes a currently spread social network sited that can change at any time.

Scientific social network sites give scholars the ability to publicize their research outputs and connect with each other [70]. Such sites have actually harnessed the power of the audience they are serving as they provide a space for greater collaboration and facilitate the creation and exchange of user-generated content [71]. It is also a way of fostering and supporting further development in the research field of interest for given academic and research institution. Therefore the participation of researchers in it should be encouraged by the institution as far as possible as people are more motivated to collaborate and maintain their relationships, connections, posts and references when their peers also use it [62].

Scientific social networks enable searching for researchers across the planet independent on their institutional affiliation in given field. This may foster the possibility of connecting business researchers to academic researchers [45] and to interconnect the academic world with the business world. A company interested in further scientific development in a specific field, may search for it in a scientific social network, get insight on achievement results of an institution based on published research articles available in the social network and contact the institution for future collaboration and cooperation.

Targeted/Focused Group
Scientific social network targets to connect researchers, scientific professionals and experts in given area across the world independently from their institutional affiliation. It provides information about conducted and carried out research as well as work under progress.

Type of Information to Include
Information tailored to research fields:

- Past and current fields of research projects, grants, collaborations and partnerships.
- Articles, books, conference papers and other scientific publications, manuals, references.
- Conferences and meetings to be organized or sponsored by the institution or other scientific events.
- An overview or outline of future planned research topics.

A greater publicizing of conducted research may also support the awareness of researchers of given scientific institution about the work and fields of devotion of colleagues who work for the same institution.

The possibility of search across the Earth based on the field enables the possibility of creating mutual partnerships and collaborations among individual institutions that are far away from each other and would have no other possibility to learn about each other. It also creates the room for sponsoring of further research by business corporations that may be interested in the development in their specific field. Such a researcher's network allows further specialization as all research institutions may join their forces in a collaborative research in one scientific field and divide tasks among themselves that may lead to cost reduction as research induces high costs.

3 Conclusion

The emergence of ICT enabled worldwide cooperation and collaboration in real-time by removing timely and spatial barriers. It interconnects individuals and institutions based on their field of interest independent from their origin, funding, affiliation, etc. The evolvement of social networks created a venue for socializing through providing the necessary environment for initiating and maintaining connections, partnerships and relationships and changed the way we behave. The linkages on social networks are gaining on importance and provide unexploited possibilities for synergy effects in various fields. An engagement of an academic institution in social networks opens new opportunities that may provide benefits to a broad spectrum of stakeholders including the institution itself. This paper provided a brief outline of motivating aspects for engagement of an academic institution in diverse social networks. Further more detailed analyses of mentioned aspects may provide powerful insights for further evolvement.

Acknowledgments. This work was supported by the Faculty of Management, Comenius University in Bratislava, Slovakia.

References

1. Dávideková, M., Hvorecký, J.: Collaboration tools for virtual teams in terms of the SECI model. In: Auer, M.E., Guralnick, D., Uhomoibhi, J. (eds.) ICL 2016. AISC, vol. 544, pp. 97–111. Springer, Cham (2017). doi:10.1007/978-3-319-50337-0_9
2. Chen, F., Sager, J.: Train to work effectively in virtual environments: a framework of virtual team competency. In: AMCIS 2007, pp. 1–11. AIS, Keystone (2007), Paper. 104
3. Ogbomo, M.O., Ogbomo, E.F.: Importance of Information and Communication Technologies (ICTs) in making a heathy information society: a case study of Ethiope East local government area of Delta State, Nigeria. Libr. Philos. Pract., 1–8 (2008). 219
4. Sarker, S., Sahay, S.: Information systems development by US-Norwegian virtual teams: implications of time and space. In: 35th Annual Hawaii International Conference on System Sciences (HICSS-35í02), 10 p. IEEE Press, New York (2002)
5. Fryer, B., Stewart, T.A., Chambers, J.: The HBR interview: cisco sees the future. Harvard Bus. Rev. **86**(11), 72–79 (2008)
6. Gunasekare, U.L.T.P.: Virtual teams in Sri Lankan business process outsourcing companies. J. Bus. Econ. Policy **2**(3), 155–160 (2015)
7. Ivanochko, I., Urikova, O., Greguš, M.: Modern mobile technologies for collaborative E-Business. In: Proceedings of the 2014 IEEE International Conference on Intelligent Networking and Collaborative Systems (INCoS 2014), pp. 515–520. IEEE Press, New York (2014). doi:10.1109/INCoS.2014.20
8. Hendriks, P.H.J.: Many rivers to cross: from ICT to knowledge management systems. J. Inf. Technol. **16**(2), 57–72 (2001). doi:10.1080/02683960110054799
9. Dávideková, M., Greguš ml., M.: Correlation properties of QOCCC based on 1D-CCC with parameters (N, N, 2N) and (N, N, N). In: Younas, M., Awan, I., Kryvinska, N., Strauss, C., van Thanh, D. (eds.) MobiWIS 2016. LNCS, vol. 9847, pp. 232–245. Springer, Cham (2016). doi:10.1007/978-3-319-44215-0_19
10. Al-Rodhan, N.R.F.: Information and Communications Technology (ICT): The Politics of Emerging Strategic Technologies. Palgrave Macmillan UK, London (2011)
11. O'kean Alonso, M.A., Hidalgo Pérez, J.M., Rodríguez López, J.: Labor demand and ICT adoption in Spain. Telecommun. Policy **40**(5), 450–470 (2015). doi:10.1016/j.telpol.2015.07.004
12. Mandorf, S., Greguš, M.: The E-business perspective as a solution for inertia against complexity management in SME. In: Proceedings of the 2014 IEEE International Conference on Intelligent Networking and Collaborative Systems (INCoS 2014), pp. 237–241. IEEE Press, New York (2014). doi:10.1109/INCoS.2014.45
13. Ivanochko, I., Urikova, O., Greguš, M.: Mobile technologies enabling collaborative services management. Int. J. Serv. Econ. Manag. **6**(4), 310–326 (2014). doi:10.1504/IJSEM.2014.068261
14. Dávideková, M., Dávideková, S., Greguš ml., M.: A case study of the application of managerial techniques of a selected team manager in the business practice. Ad Alta: J. Interdisc. Res. **5**(2), 20–26 (2015)

15. Jorgenson, D.W., Wu, K.M.: The ICT revolution, world economic growth, and policy issues. Telecommun. Policy **40**, 383–397 (2016). doi:10.1016/j.telpol.2016.01.002
16. Dávideková, M., Greguš ml., M., Bureš, V.: Yet another classification of ICT in knowledge management initiatives: synchronicity and interaction perspective. In: Advanced Science Letters (2016, in press)
17. Gonçalves, P., Ferreira, L., Gonçalves, J., Putnik, G.D., Cruz-Cunha, M.M.: Direct communication versus virtual communication in virtual teams. Proc. Technol. **16**, 3–10 (2014). doi:10.1016/j.protcy.2014.10.062
18. Riemer, K., Frößler, F., Klein, S.: Real time communication-modes of use in distributed teams. In: European Conference on Information Systems (ECIS), pp. 286–297. AIS, Geneva (2007)
19. Broadbent, M., Weill, P.: Management by maxim: how business and IT managers can create IT infrastructures. Sloan Manag. Rev. **38**, 1–31 (1997)
20. Klitmøller, A., Lauring, J.: When global virtual teams share knowledge: media richness, cultural difference and language commonality. J. World Bus. **48**(3), 398–406 (2013). doi: 10.1016/j.jwb.2012.07.023
21. Cherbakov, L., Galambos, G., Harishankar, R., Kalyana, S., Rackham, G.: Impact of service orientation at the business level. IBM Syst. J. **44**(4), 653–668 (2005). doi:10.1147/sj.444.0653
22. Kumar, R.R., Stauvermann, P.J., Samitas, A.: The effects of ICT on output per worker: a study of the Chinese economy. Telecommun. Policy **40**(2–3), 102–115 (2015). doi:10.1016/j.telpol.2015.06.004
23. Johansen, R.: Groupware: Computer Support for Business Teams. The Free Press, New York (1988)
24. Johnson-Lenz, P., Johnson-Lenz, T.: Groupware for a small planet. In: Lloyd, P. (ed.) Groupware in the 21st Century: Computer Supported Cooperative Working Toward the Millennium, pp. 269–284. Greenwood Publishing Group Inc., Westport (1994)
25. Johnson-Lenz, P., Johnson-Lenz, T.: Groupware: the process and impacts of design choices. In: Kerr, E.B., Hiltz, S.R. (eds.) Computer-Mediated Communication Systems: Status and Evaluation. Academic Press, Orlando (1982)
26. Greer, B.M., Luethge, D.J., Robinson, G.: Utilizing virtual technology as a tool to enhance the workforce diversity learning. In: Scott, C.L., Sims, J.D. (eds.) Developing Workforce Diversity Programs, Curriculum, and Degrees in Higher Education, AHEPD, pp. 258–279. IGI Global, Oklahoma (2016)
27. Hendriks, P.H.J.: Why share knowledge? The influence of ICT on the motivation for knowledge sharing. Knowl. Process Manag. **6**(2), 91–100 (1999). doi:10.1002/(SICI)1099-1441(199906)6:2<91:AID-KPM54>3.0.CO;2-M
28. Alavi, M., Leidner, D.E.: Knowledge management systems: issues, challenges, and benefits. CAIS **1**(2es), 1, 1–37 (1999)
29. Torro, O., Pirkkalainen, H.: Strengthening social ties via ICT in the organization. In: Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS 2017), pp. 5511–5520 (2017, in press). doi:http://hdl.handle.net/10125/41828
30. Skeels, M.M., Grudin, J.: When social networks cross boundaries: a case study of work-place use of Facebook and LinkedIn. In: Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP 2009), pp. 95–104. ACM, New York (2009). doi: 10.1145/1531674.1531689
31. Vanwesenbeeck, I., Ponnet, K., Walrave, M.: Young adolescents' advertising literacy and purchase intention in social network games: influence of perspective taking and need for cognition. J. Consumer Behav. **16**(1), 23–33 (2017). doi:10.1002/cb.1596
32. Wellman, B.: An electronic group is virtually a social network. In: Kiesler, S. (ed.) Culture of the Internet, pp. 179–205. Lawrence Erlbaum, New Jersey (1997)

33. Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput. Mediated Commun. **13**(1), 210–230 (2007). doi:10.1111/j.1083-6101.2007.00393.x
34. Steinfield, C., Ellison, N.B., Lampe, C.: Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. J. Appl. Dev. Psychol. **29**(6), 434–445 (2008). doi: 10.1016/j.appdev.2008.07.002
35. Stephenson, K.: What knowledge tears apart, networks make whole. Internal Commun. Focus **36**, 1–6 (1998)
36. Porath, S.: Text messaging and teenagers: a review of the literature. J. Res. Center Educ. Technol. (RCET) **7**(2), 86–99 (2011)
37. Steyer, J.: Is Technology Networking Changing Childhood? A National Poll by Common Sense Media. Common Sense Media, San Francisco (2009)
38. O'Keeffe, G.S., Clarke-Pearson, K.: Council on communications and media: the impact of social media on children, adolescents, and families. Pediatrics **127**(4), 800–804 (2011). doi: 10.1542/peds.2011-0054
39. Eurostat: Information society statistics-households and individuals (2015a). http://ec.europa.eu/eurostat/statistics-explained/index.php/Information_society_statistics_-_households_and_individuals
40. Van Cleemput, K.: I'll see you on IM, text, or call you: a social network approach of adolescents' use of communication media. Bull. Sci. Technol. Soc. **30**(2), 75–85 (2010). doi: 10.1177/0270467610363143
41. Lacerda, A., Ziviani, N.: Building user profiles to improve user experience in recommender systems. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM), pp. 759–764. ACM, New York (2013). doi:10.1145/2433396.2433492
42. Moolenaar, N.M., Daly, A.J., Cornelissen, F., Liou, Y.H., Caillier, S., Riordan, R., Wil-son, K., Cohen, N.A.: Linked to innovation: shaping an innovative climate through network intentionality and educators' social network position. J. Educ. Change **15**(2), 99–123 (2014). doi:10.1007/s10833-014-9230-4
43. Siemens, G.: Connectivism: a learning theory for the digital age. Int. J. Instr. Technol. Distance Learn. (ITDL) **1**, 1–8 (2005). doi:http://er.dut.ac.za/handle/123456789/69
44. Tantawy, R.Y., Farouk, Z., Mohamed, S., Yousef, A.H.: Using professional social networking as an innovative method for data extraction: the ICT alumni index case study. In: 1st International Conference on Innovation & Entrepreneurship, Cairo, April 2012, pp. 1–4. arXiv preprint (2014). doi:arXiv:1410.1348
45. Van Noorden, R.: Online collaboration: scientists and the social network. Nature **512**(7513), 126–129 (2014). doi:10.1038/512126a
46. Park, E., Baek, S., Ohm, J., Chang, H.J.: Determinants of player acceptance of mobile social network games: an application of extended technology acceptance model. Telematics Inform. **31**(1), 3–15 (2014). doi:10.1016/j.tele.2013.07.001
47. Barker, V.: Older adolescents' motivations for social network site use. CyberPsychol. Behav. **10**(3), 478–481 (2009). doi:10.1089/cpb.2008.0228
48. Facebook. https://www.facebook.com/
49. Path. https://path.com/
50. MySpace. https://myspace.com/
51. Tagged. The social network for meeting new people. http://www.tagged.com/?loc=en_US
52. DiaSpora. https://joindiaspora.com/
53. WeChat. https://wx.qq.com/?lang=en_US
54. Neubaum, G., Rösner, L., Rosenthal-von der Pütten, A.M., Krämer, N.C.: Psychosocial functions of social media usage in a disaster situation: a multi-methodological approach. Comput. Hum. Behav. **34**, 28–38 (2014). doi:10.1016/j.chb.2014.01.021

55. Dávideková, M.: Digitalization of society: smartphone - a threat? In: 8th International Research Conference Management Challenges in the 21st Century, pp. 314–320. Vysoká škola manažmentu v Trenčíne, Bratislava (2016)
56. Agger, B.: iTime: labor and life in a smartphone era. Time Soc. **20**(1), 119–136 (2011). doi: 10.1177/0961463X10380730
57. Eurostat: Being young in Europe today - digital world (2015b). http://ec.europa.eu/eurostat/statistics-explained/index.php/Being_young_in_Europe_today_-_digital_world
58. LinkedIn. https://www.linkedin.com/
59. Xing. https://www.xing.com/
60. RallyPoint: The Professional Military Network. https://www.rallypoint.com/
61. Doximity: The Medical Network: Where care comes together. https://www.doximity.com/
62. Zhang, C.-B., Li, Y.-N., Wu, B., Li, D.-J.: How WeChat can retain users: roles of network externalities, social interaction ties, and perceived values in building continuance intention. Comput. Hum. Behav. **69**, 284–293 (2017). doi:10.1016/j.chb.2016.11.069
63. ResearchGate. https://www.researchgate.net/home
64. Academia. https://www.academia.edu/
65. Semantic Scholar: Cut through the clutter. https://www.semanticscholar.org/
66. Mendeley: Empowering researchers. https://www.mendeley.com/
67. Bibsonomy: The blue social bookmark and publication sharing system. https://www.bibsonomy.org/
68. Zotero: Zotero [zoh-TAIR-oh] is a free, easy-to-use tool to help you collect, organize, cite, and share your research sources. https://www.zotero.org/
69. CiteULike: citeulike is a free service for managing and discovering scholarly references. http://www.citeulike.org/
70. Thelwall, M., Kousha, K.: Academia. edu: social network or academic network? J. Assoc. Inf. Sci. Technol. **65**(4), 721–731 (2014). doi:10.1002/asi.23038
71. Al-Debei, M.M., Al-Lozi, E., Papazafeiropoulou, A.: Why people keep coming back to Facebook: explaining and predicting continuance participation from an extended theory of planned behaviour perspective. Decis. Support Syst. **55**(1), 43–54 (2013). doi:10.1016/j.dss.2012.12.032

The 9th International Workshop on Information Network Design (WIND-2017)

Method for Determining Recovery Order Against Intermittent Link Failures

Tsuyoshi Yamasaki and Hiroyoshi Miwa(✉)

Graduate School of Science and Technology, Kwansei Gakuin University,
2-1 Gakuen, Sanda-shi, Hyogo, Japan
{tsuyo,miwa}@kwansei.ac.jp

Abstract. High reliability is required in networks, and it needs to build robust networks. In content delivery services in particular, service interruptions due to disconnection of communication paths between a server and nodes that receive the service must be avoided. Contents delivery services use a master server that contains original contents and edge-servers that hold copies of the contents. It needs that communication paths exist between the master server and the edge-servers even in the network failure. From this background, there is a lot of research on network design method to keep high reliability. On the other hand, recovery scheduling has not considered intermittent link failures, so that, if link failures occur, it was assumed that the failed links are recovered immediately. However, indeed, there is a possibility that network failures occur by an aftershock. In fact, in the Great East Japan Earthquake in 2011 and Kumamoto earthquake in 2016, road networks and electric networks were damaged by not only the main shock but also the aftershocks. In such a case, material resource and human resource are restricted, so that it is necessary to determine the recovery order of the failed links. In this paper, we aim to determine an appropriate recovery order of failed links against intermittent link failures. We formulate the optimization problem determining the recovery order and design a heuristic algorithm. In addition, we evaluate the performance of the algorithm by applying it to the topology of various actual networks.

1 Introduction

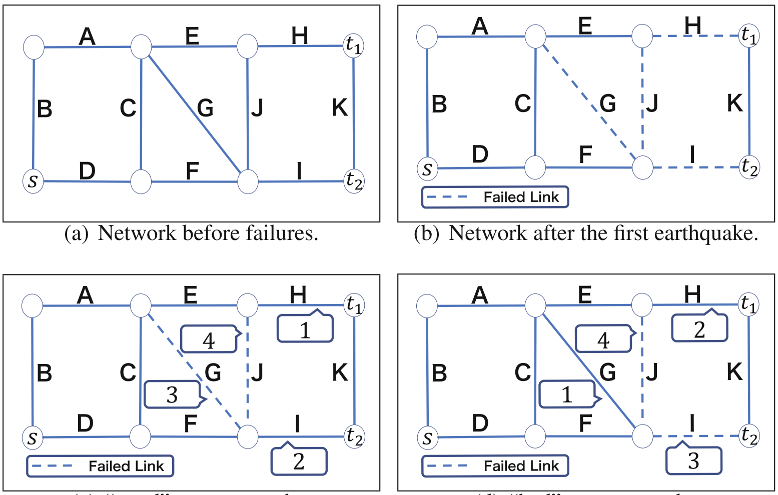
High reliability is required in networks, and it needs to build robust networks. In content delivery services in particular, service interruptions due to disconnection of communication paths between the server and nodes that receive the service must be avoided. Contents delivery services use a master server that contains original contents and edge-servers that hold copies of the contents. It needs that communication paths exist between the master server and the edge-servers even in the network failure. From this background, there is a lot of research on network design method to keep high reliability.

On the other hand, recovery scheduling has not considered intermittent link failures, so that, if link failures occur, it was assumed that the failed links are

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_36

recovered immediately. However, indeed, there is a possibility that network failures occur by an aftershock. In fact, in the Great East Japan Earthquake in 2011 and Kumamoto earthquake in 2016, road networks and electric networks were damaged by not only the main shock but also the aftershocks. In such a case, material resource and human resource are restricted, so that it is necessary to determine the recovery order of the failed links.

Figure 1 shows an example that the difference of recovery order causes difference of connectivity in case of intermittent link failures. In Fig. 1, s is the master server and t_1, t_2 are the edge-servers. In Fig. 1(a), it is assumed that four links fail at first earthquake, and two links are recovered until the next aftershock occurs. Furthermore, it is assumed that one link fails by the next aftershock. The node s and the node t_1, the node s and the node t_2 should be always connected, respectively; however, the difference of the recovery order causes the difference of the connectivity. When the links are recovered by the recovery order in Fig. 1(c), even after the next aftershock occurs, the connectivity is kept, that is, both the node s and the node t_1 and the node s and the node t_2 are connected. However, when the links are recovered by the recovery order in Fig. 1(d), after the next aftershock occurs, the node s is disconnected to the node t_1 and t_2 when the link H fails. The former recovery order keeps the connectivity against the successive failures. Such a "good" recovery order is desirable, when the intermittent failures are predicted.



(a) Network before failures. (b) Network after the first earthquake.

(c) "good" recovery order. (d) "bad" recovery order

Fig. 1. Difference of connectivity caused by difference of recovery order.

As for the research on network reliability, there are some network design methods from the viewpoint of link protection. If a link is protected, when a failure is detected in lower-level layers, at the same time, the path is switched by using an alternative path prepared in advance, such links via IP routes can be regarded as not experiencing failures. It is necessary to achieve high reliability by protecting as few links as possible from a view point of cost. The problems of determining the protected links to keep the connectivity [1], the problems of determining the protected links to keep the diameter [2], and so on, have been studied so far.

The research works on recovery order are extensively studied in the research areas such as road networks. The references [3,4] deal with the problem of determining the recovery segments of a road network by taking the set of actual travel routes of cars into consideration. The road segments on which many cars pass are preferentially recovered. The references [5–7] deal with the recovery order of road segments. In these previous research, it is implicitly assumed that a failure occurs only once, and hence the intermittent failures are not considered. Therefore, when the recovery orders is determined by a previous method, the connectivity may be often lost by intermittent link failures. A method to determine recovery orders against intermittent link failures is needed.

In this paper, we formulate the optimization problem determining an appropriate recovery order of failed links against intermittent link failures and design a heuristic algorithm. In addition, we evaluate the performance of the algorithm by applying it to the topology of an actual network.

2 Problem for Determining Recovery Order Against Intermittent Link Failures

The graph representing a communication network is $G = (V, E)$. Here, V and E are the set of vertices and the set of edges of G, respectively, the number of vertices is $|V| = n$, and the number of edges is $|E| = m$. We associate the weight and capacity with each edge by weight function $w : E \to \mathbb{R}^+$ and capacity function $c : E \to \mathbb{R}^+$. The path length is defined as the sum of the weight of all the edges included in the path. The shortest path between two vertices is defined as the path with the smallest path length among all paths between the vertices. Let $Q = \{(q_1 = \{s_1, t_1\}, d_1), (q_2 = \{s_2, t_2\}, d_2), \ldots, (q_r = \{s_r, t_r\}, d_r)\}$ be the demand set, where, for each pair $q_i = \{s_i, t_i\}$ of vertices $(i = 1, 2, \ldots, r)$, the demand level between s_i and t_i is $d_i \in \mathbb{R}^+$. The demand level between s_i and t_i is routed on the shortest path between s_i and t_i. Let $m(e)$ be the sum of the demands of all the shortest paths included in edge e. For network $N = (G = (V, E), c, w)$ and demand set $Q = \{(q_1, d_1), (q_2, d_2), \ldots, (q_r, d_r)\}$, when $m(e) \leq c(e)$ for all edges $e \in E$, the *capacity constraint* is satisfied. When s_i and t_i is connected for all $q_i = \{s_i, t_i\}$ $(i = 1, 2, \ldots, r)$, the *reachability constraint* is satisfied. If the path length between each vertex pair after an edge is removed is b or less-times of the path length before removing the edge (where b is a positive real number which is an input constant), the *path length constraint* is satisfied. The value b is called the stretch factor. When the capacity constraint, the reachability constraint, and the path length constraint whose stretch factor for all pairs in Q is b, are satisfied, we say that the network is feasible.

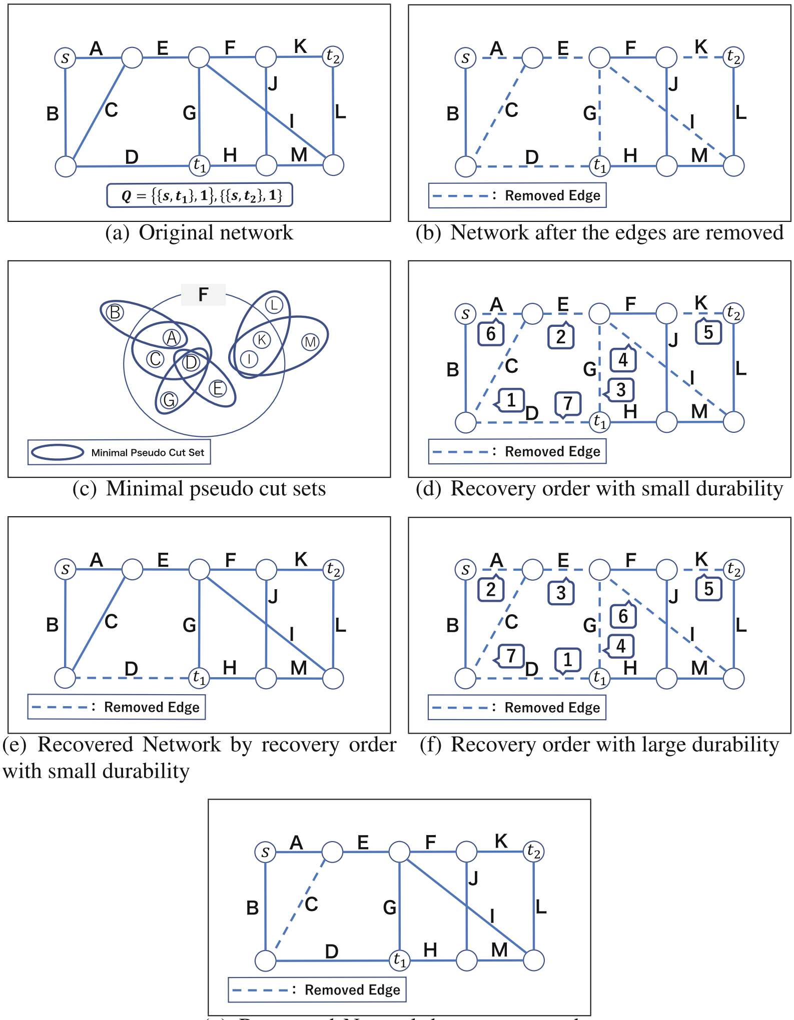
When $Q = \{(\{s, t_1\}, d_1), (\{s, t_2\}, d_2), \ldots, (\{s, t_r\}, d_r\})$, vertex s corresponds to the master server and vertices $\{t_1, t_2, \ldots, t_r\}$ correspond to the edge-servers.

When network $N = (G = (V, E), c, w)$, demand set $Q = \{(q_1, d_1), (q_2, d_2), \cdots, (q_r, d_r)\}$ and the stretch factor b are given, if $E^s(\subseteq E)$ is the minimal subset that the network $((V, E - E^s), c, w)$ is not feasible, E^s is called a minimal pseudo cut set.

For a minimal pseudo cut set E^s and $F(\subseteq E)$ corresponding to the failure links, $|E^s \backslash F|$ is called the remaining number in E^s against F. The minimum of the remaining number in all minimal pseudo cut sets against F is called the remaining number against F and it is referred as $a(F)$. An order of all the edges in F, $\{f_1, f_2, \ldots, f_k\}$, is called a recovery order, when the edges are recovered in the order; the sequence, $\{a(F), a(F - \{f_1\}), a(F - \{f_1, f_2\}), \ldots, a(\emptyset)\}$ is called the remaining number sequence of the recovery order $\{f_1, f_2, \ldots, f_k\}$. Note that the difference of the adjacent values in a remaining number sequence is at most one.

We define the order relation by the lexicographic order relation in all the recovery orders of F. Let the first recovery order be $\{r_1^*, r_2^*, \ldots, r_m^*\}$ in the order relation. For an recovery order $seq = \{r_1, r_2, \ldots, r_m\}$, $\sum_{i=1}^{m}(r_i - r_i^*)$ is called the durability of *seq*. We define the maximal recovery order as the "optimal" recovery order. We describe the basic idea of the optimality of the recovery order as follows: when all the edges in a minimal pseudo cut set are removed, the network becomes infeasible. This implies that, when we assume that the probability that an edge is removed is the same, if the number of the edges included in a minimal pseudo cut set is small, the probability that the network after the next removal of edges corresponding to the next aftershock becomes infeasible is large. Therefore, the removed edges included in a minimal pseudo cut set whose remaining number is small should be recovered before the removed edges included in a minimal pseudo cut whose remaining number is large. In other words, the recovery order that the remaining numbers increase rapidly from the top in a remaining number sequence is desirable. The durability is a measure from this viewpoint.

We show an example in Fig. 2. In the network in Fig. 2(a), we assume that seven edges are simultaneously removed and that six edges are recovered, that is, $Q = \{(\{s, t_1\}, 1), (\{s, t_2\}, 1)\}$, all edges is 1, and the edges in Fig. 2(b) are removed. The minimal pseudo cut set is shown in Fig. 2(c). The remaining number sequence of the recovery order in Fig. 2(d) is $\{0, 0, 0, 1, 1, 1, 1, 2\}$. The network after six edges are recovered is shown in Fig. 2(e). Since the first recovery order is $\{0, 0, 0, 0, 0, 1, 1, 2\}$ (the order of edges K, I, E, G, C, A, D), the durability is two. The edge-connectivity of this network is one, so that, if another edge is removed, the network becomes infeasible. On the other hand, the remaining number sequence of the recovery order in Fig. 2(f) is $\{0, 1, 1, 1, 1, 2, 2, 2\}$. The network after six edges are recovered is shown in Fig. 2(g). The durability of this network is six. The edge-connectivity of this network is two, so that, even if another edge is removed, the network keeps to be feasible. This example indicates that the



(a) Original network
(b) Network after the edges are removed
(c) Minimal pseudo cut sets
(d) Recovery order with small durability
(e) Recovered Network by recovery order with small durability
(f) Recovery order with large durability
(g) Recovered Network by recovery order with large durability

Fig. 2. Example of recovery order, remaining number sequence, and durability.

recovery order with a large durability improves the probability of keeping feasibility against intermittent failures.

We define the problem determining the maximal recovery order as follows:

Problem for Determining Recovery Order (ROP)
INPUT *Network* $N = (G = (V, E), c, w)$ *demand set* $Q = \{(q_1, d_1), (q_2, d_2), \ldots, (q_r, d_r)\}$, *positive real number* b, *and* $F(\subseteq E)$.
OUTPUT recovery order for (N, Q, b, F).
OBJECTIVE: *durability of recovery order (maximize).*

The decision problem, which asks whether there is the recovery order whose durability is more than or equal to d, is also called ROP.

Theorem 1. *ROP is generally NP-hard, even if the weight of all edges is one and the demand level of all demand set is one.*

Proof. We reduce vertex cover problem (VC) known as NP-hard [8] to ROP in polynomial time. A vertex cover of a graph is a set of vertices such that each edge of the graph is incident to at least one vertex of the set. VC is the problem to determine whether there is a vertex cover whose size is a given integer z or less.

Let an instance of VC be $(G = (V, E), z)$ where G is a graph, $|V| = n$, $|E| = m$. Let V be $\{v_1, v_2, \ldots, v_n\}$. We construct the instance of ROP, $(N = (G' = (V', E'), c, w), Q, b = 1, F, d = n - z - 1)$ from $(G = (V, E), z)$ as follows: We make an edge v' for each vertex v in G, vertices s_{xy}, t_{xy} for all edges (x, y) in G, and connect s_{xy} and t_{xy} to both end vertices of edge x' and edge y'. Let the weight of all edges be one and let the capacity of all edges be sufficiently large. Let $Q = \{((s_{xy}, t_{xy}), 1) | (x, y) \in E\}$, $b = 1$, $F = \{v' \in E(G') | v \in V\}$, and $d = n - z - 1$. Thus, we have the instance of ROP, $(N = (G' = (V', E'), c, w), Q, b = 1, F, d = n - z - 1)$. This transformation can be executed in polynomial time.

First, we show that a solution of $(N = (G' = (V', E'), c, w), Q, b = 1, F, d = n - z - 1)$ can be constructed from a solution $\tilde{V}$ ($|\tilde{V}| = \tilde{z}(\leq z)$) of (G, z). The family of the minimal pseudo cut sets of this instance is $\{\{v', w'\} | v', w' \in E(G'), (v, w) \in E(G)\}$ and $F = \{v' \in E(G') | v \in V\}$. The recovery order that the sequence of the edges corresponding to the vertices in $\tilde{V}$ is followed by the sequence of the other edges of $E(G')$ has the remaining number sequence $\{0, 0, \ldots, 0, 1, 1, \ldots, 1, 2\}$ where the number of zero is $\tilde{z}$, the number of one is $n - \tilde{z}$, and the last value is two. Since the first remaining number sequence is $\{0, 0, \ldots, 0, 1, 2\}$ that corresponds to the recovery order that all the edges except a minimal pseudo cut set are recovered and then the edges of the minimal pseudo cut set are recovered, the durability is $(n + 1 - \tilde{z}) - 2 = n - \tilde{z} - 1 \geq n - z - 1 = d$. Therefore, the recovery order is the solution of ROP.

Conversely, we show that a solution of (G, z) can be constructed from a solution of $(N = (G' = (V', E'), c, w), Q, b = 1, F, d = n - z - 1)$. The remaining number sequence whose durability is more than or equal to $n - z - 1$ is $\{0, 0, \ldots, 0, 1, 1, \ldots, 1, 2\}$ where the number of zero is z or less, the value of one is followed, and the last value is two. This is because the values line up in the ascending order and the last value is two. Note that the size of the minimal pseudo cut set is two. The edges in $E(G')$ recovered in the recovery order corresponding to $\{0, 0, \ldots, 0, 1\}$ of this remaining number sequence correspond to a vertex cover of G, because, as at least an edge in each minimal pseudo cut set is recovered and a minimal pseudo cut set corresponds to an edge in G, all edges of G include at least a vertex corresponding to the recovered edge in G'. In addition, as the number of the recovered edges is less than or equal to z, the size of this vertex cover is also less than or equal to z. Thus, this vertex cover is the solution of (G, z).

Consequently, ROP is NP-hard. □

As ROP is NP-hard, we cannot expect a polynomial time algorithm for ROP; therefore, we propose a heuristic algorithm to solve ROP.

We describe the basic idea of the heuristic algorithm as follows: First, we enumerate all minimum pseudo cut sets including at least an edge of F. Let E_h be the set of the edges included in the minimal pseudo cut set whose remaining number against F is the minimum, and let $H(e)$ be the set of the minimum pseudo cut sets including edge e. Let $e_1(\in E_h)$ be the edge that $|H(e_1)|$ is the maximum, and we recover the edge e_1 first. Next, we replace F by $F - \{e_1\}$ and find $e_2(\in E_h)$ similarly. We iterate this procedure for all edges in F, and thus we determine the recovery order. In this procedure, the edge included in the minimum pseudo cut set whose remaining number is small is preferentially recovered, because, if such an edge is removed by the next removal of edges corresponding a successive aftershock, the probability that the network becomes infeasible is high. Especially, since the edge whose recovery increases the remaining numbers of many minimum pseudo cut sets is chosen, the recovery order that the remaining numbers increases rapidly from the top in a remaining number sequence is constructed.

Algorithm 1. Algorithm for Determining Recovery Order

Input: Network $N = (G = (V, E), c, w)$, demand set
$Q = \{\{q_1, d_1\}, \{q_2, d_2\}, \ldots, \{q_r, d_r\}\}$, positive real number b, $F(\subseteq E)$
Output: Recovery order list D

$D \leftarrow \emptyset$
Enumerate all minimal pseudo cut sets in G by the algorithm MDS
while $F \neq \emptyset$ **do**
  Calculate the remaining number in each minimal pseudo cut set against F
  Calculate $|H(e)|$, the number of minimal pseudo cut sets including edge e for each edge e in F
  Determine E_h, the set of the edges included in the minimal pseudo cut set whose remaining number against F is the minimum
  Determine the edge $e(\in E_h)$ that $|H(e)|$ is the maximum and add e to the tail of D
  $F \leftarrow F \setminus \{e\}$
return D

The computational complexity is the exponential order; however, the size of a minimal pseudo cut set is often small in an actual network, so that we can often determine the recovery order of an actual network.

3 Performance Evaluation

We evaluate the performance of the algorithm in Sect. 2 by using the topology of the actual networks in [9]. For that purpose, we compare the proposed algorithm

Algorithm 2. Algorithm determining minimal pseudo cut sets, MDS

Input: Network $N = (G = (V, E), c, w)$ demand set
$Q = \{\{q_1, d_1\}, \{q_2, d_2\}, \ldots, \{q_r, d_r\}\}$, positive real number b
Output: The family of minimal pseudo cut sets, C

```
C ← ∅
for 1 ≤ k ≤ m do
    Determine C_k, the set of the edge subsets whose number of edges k
for 1 ≤ k ≤ m do
    while there is an unscanned edge subset E_k in C_k do
        if (N' = (G' = (V, E − E_k), c, w), b, Q) is infeasible then
            if C has no edge set properly included in E_k then
                C ← E_k
return C
```

(Prop) and the algorithm (Simp) that determines the recovery order by the order of failures of links.

Considering the situation of the intermittent aftershocks, in the following numerical experiments, we assume that an event of the removal of links are succesively continued. When an event occurs at the time t and the next event occurs at the time t', the period $[t, t')$ is referred to as a round. The removed links are tried to be recovered during a round, but all removed links are not always recovered and some links are recovered in the next round or later together with the removed links in the next event. In the following numerical experiments, the removed links are randomly chosen from the non-removed links at an event. Note that a link may be removed several times. At each round, the algorithms Prop or Simp determines the recovery order to all the removed links.

Let $Q = \{(\{s, t_1\}, 1), (\{s, t_2\}, 1), \ldots, (\{s, t_r\}, 1)\}$, where s corresponds to the master server and $\{t_1, t_2, \ldots, t_r\}$ correspond to the edge-servers. We assume that the capacity constraint is always satisfied. When the reachability constraint and the path length constraint for a pair in Q are satisfied, we say that the pair is feasible. The measure is the average of the number of the feasible pairs. We show the network topology used for the evaluation in Fig. 3.

The Table 1, where n is the number of nodes and m is the number of links, and b is the stretch factor, indicates the results. The number of the rounds is ten. The number of the recovered links in a round is randomly chosen in the range from $0.2m$ to $0.3m$. We evaluate the case that the number of the failed links in an event is randomly chosen in the range from $0.2m$ to $0.3m$ and the case that the number of the failed links in an event is randomly chosen in the range from $0.3m$ to $0.4m$.

In Table 1, when the number of the recovered links in a round and the number of the failed links in a round are the same range from $0.2m$ to $0.3m$, the number of the feasible pairs by the algorithm Prop is larger than that by the algorithm Simp. There are many independent paths between the master server and an

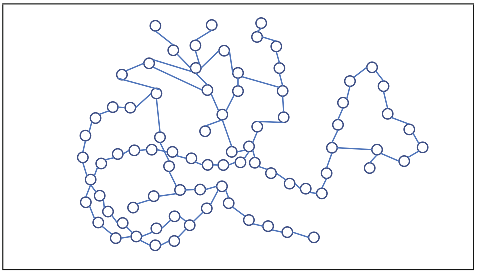


Fig. 3. Network used for evaluation

Table 1. Results on the network in Fig. 3

The number of edge-servers is 6		n = 82, m = 92							
		Stretch Factor							
		b = 1.05				b = 1.5			
		Failure range				Failure range			
		0.2 ~ 0.3		0.3 ~ 0.4		0.2 ~ 0.3		0.3 ~ 0.4	
		Recovery Alg.		Recovery Alg.		Recovery Alg.		Recovery Alg.	
		Prop	Simp	Prop	Simp	Prop	Simp	Prop	Simp
Round	1	4.46	1.76	0.28	0.04	3.85	2.01	0.41	0.01
	2	4.70	2.85	3.76	1.22	4.09	2.96	3.35	1.65
	3	4.72	2.55	3.52	0.89	4.10	2.59	2.91	0.62
	4	4.67	2.27	3.01	0.45	4.01	2.52	2.50	0.40
	5	4.55	2.11	2.44	0.28	3.96	2.52	2.19	0.15
	6	4.50	1.91	1.89	0.13	3.95	2.41	1.69	0.07
	7	4.51	1.90	1.16	0.12	3.91	2.14	1.18	0.06
	8	4.52	1.70	0.56	0.03	3.94	2.24	0.82	0.04
	9	4.35	1.84	0.34	0.03	3.85	1.94	0.56	0.02
	10	4.49	1.66	0.29	0.02	3.78	2.01	0.47	0.00

edge-server in this network, so that the number of the minimal pseudo cut sets including an edge is large; therefore, the recovery of an edge chosen appropriately by the algorithm Prop causes the increase of the feasible pairs. When the number of the failed links in a round is the range from $0.3m$ to $0.4m$, the number of the feasible pairs when $b = 1.5$ is sometimes larger than that when $b = 1.05$. When the stretch factor is large, the number of available independent paths is large; therefore, the large number of pairs becomes feasible by recovering the small number of edges.

4 Conclusion

In this paper, we addressed the problem of determining an appropriate recovery order of failed links against intermittent link failures. Especially, we formulated the problem to keep the connectivity between a master server and some edge-servers in contents delivery services as the optimization problem. First, we proved that the problem is NP-hard. Furthermore, we designed the heuristic algorithm and we evaluated the performance of the algorithm by applying it to the topology of the actual network. The results of the performance evaluation of the algorithm indicated that the difference of the recovery order influences the keep of the connectivity between the master server and the edge-servers and that the proposed algorithm can keep the connectivity against intermittent link failures more than the algorithm that determines the recovery order by the order of failures of links.

Acknowledgment. This work was partially supported by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (B) (17H01742).

References

1. Yamasaki, T., Anan, M., Miwa, H.: Network design method based on link protection taking account of the connectivity and distance between sites. In: Proceedings of International Conference on Intelligent Networking and Collaborative System (INCoS2016), Ostrava, Czech Republic, 7–9 September 2016, pp. 445–450 (2016)
2. Imagawa, K., Miwa, H.: Approximation algorithm for finding protected links to keep small diameter against link failures. In: Proceedings of International Conference on Intelligent Networking and Collaborative System (INCoS2011), 30 November–2 December 2011, pp. 575–580 (2011)
3. Sen, C., Yun-jun, L., Ying-wu, C., Yong-ping, S.: Optimized road network emergency restoration scheduling for relief distribution. In: Proceedings of International Conference on E-Business and E-Government (ICEE), 6–8 May 2011, pp. 1–4 (2011)
4. Itabashi, R., Katsuya, S., Takahiko, K., Yasuo, A.: A location-routing decision model for first-aid restoration of a damaged road network. J. East. Asia Soc. Transp. Stud. **11**(25), 557–565 (2015)
5. Furuta, H., Ishibashi, K., Shun, H.: Optimal restoration scheduling of damaged network under uncertain environment by using improved genetic algorithm. Tsinghua Sci. Technol. **13**(2), 400–405 (2008)
6. Lertworawanich, P.: Highway network restoration after the great flood in Thailands. Nat. Hazards **64**, 873–886 (2012)
7. Zamanifar, M.: Recovery planning model for roadways network after natural hazard. Nat. Hazards, 1–18 (2017)
8. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H Freeman and Company, New York (2003)
9. CAIDA. http://www.caida.org/

Method for Finding Protected Links to Keep Robust Network Against Link Failure Considering Failure Probability

Keyaki Uji and Hiroyoshi Miwa(✉)

School of Science and Technology, Kwansei Gakuin University,
2-1 Gakuen, Sanda-shi, Hyogo, Japan
{keyaki-uji,miwa}@kwansei.ac.jp

Abstract. It is important to design and operate a reliable network which is sufficiently robust against network failures. Particularly, in the content distribution service, it is necessary to minimize the probability of service interruption due to disconnection of the communication path between the node receiving the service and the server. For that purpose, it is desirable to protect all links so that the failure probability of a link is almost zero by sufficiently backup resource and rapid recovery system. We can consider such a protected link does not fail. However, link protection needs much cost; therefore, it is important to find the smallest number of links to be protected so that a network resulting from failures of any non-protected links provides the connectivity. In this paper, we take failure probability of links into consideration. In the previous research, it was assumed that the links are protected so that the probability that the whole network does not satisfy the condition of the connectivity is zero. Practically, however, it is sufficient to reduce failure probability of a network sufficiently small. It implies that we can design a practically reliable network with the small number of protected links. First, we formulate this problem and show that this problem is NP-hard. Furthermore, we propose two heuristic algorithms to solve the problem. In addition, we show some examples of application of the heuristic algorithms to the topology of the various actual networks and evaluate the performance of the algorithms.

1 Introduction

The information networks, which is important as social infrastructure, must be sufficiently robust against network failures. However, it is difficult to avoid failures especially in a large network. Therefore, it is necessary to design a network that keeps communication even during failures.

Particularly, the importance of the content distribution service is increasing recently. Generally, in a system for content distribution, a master server having the original content and some edge servers holding the copy and receiving access from the user are used. To offer the reliable content distribution service, it is

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_37

necessary to decrease the probability of service interruption due to disconnection of the communication path between the master server and the edge servers even in the event of a failure. For that purpose, it is desirable to protect all links so that the failure probability of a link is almost zero by sufficiently backup resource and rapid recovery system. However, link protection needs much cost; therefore, it is important to find the smallest number of links to be protected so that a network resulting from failures of any non-protected links provides the connectivity.

A measure for evaluating the reliability of a network is the number of connected components when the network becomes disconnected due to a failure [1]. The reference [1] deals with an optimization problem for maximizing the number of connected components disconnected due to the failure of nodes. There is also some results on link protection. The reference [2] deals with the problem of determining a set of protected links that keep the reachability to a specific node and proposes an approximation algorithm to this problem. The reference [3] deals with the problem of determining protected links so that each connected component includes at least a specified nodes after a link failure. The reference [4] deals with the problem of determining protected links so that the master server and all edge servers are connected and that the capacity constraint is satisfied and that the increase ratio of the distance in a network to the distance in a failed network does not exceed a threshold.

When each edge has the probability that the edge is removed, the network reliability that the whole network satisfies the condition of the connectivity is defined (ex. [5]).

As described above, there are a lot of the studies of the network design problems based on link protection or failure probability; however, there is no study of the network design problem considering both link protection and failure probability.

In this paper, we take failure probability into consideration. In the previous research, for sake of simplicity, it was assumed that the links are protected so that the probability that the whole network does not satisfy the condition of the connectivity is zero. Practically, however, it is sufficient to reduce failure probability of a network sufficiently small. It implies that we can design a practically reliable network with the small number of protected links.

First, we formulate the problem minimizing the number of the protected links so that, even if any non-protected links fail, the failure probability of a whole network in a resulting network after links fail is below a given threshold. Next, we show that this problem is NP-hard. Furthermore, we propose two heuristic algorithms to solve the problem. In addition, we show some examples of application of the heuristic algorithms to the topology of the various actual networks and evaluate the performance of the algorithms.

2 Link Protection Problem Considering Failure Probability

Let $G = (V, E)$, where V and E are the vertex set and the edge set of G, respectively, be the graph representing a network structure. We associate the failure probability with each edge by the failure probability function $h : E \to \mathbb{R}^+$. Let $Q = \{q_1 = \{s_1, t_1\}, q_2 = \{s_2, t_2\}, \ldots, q_r = \{s_r, t_r\}\}$ be the set of vertex pairs.

A subset of edges, $E_P(\subseteq E)$, corresponds to a set of protected links; a subset of edges, $E_K(\subseteq E)$, corresponds to a set of failed links; and G' corresponds to the network resulting from the link failure. The protection of edge e is equivalent to change the failure probability of e, $h(e)$, to zero.

For the set of the protected edges, $E_P(\subseteq E)$, let the set of $E_K(\subseteq E, |E_K| = i)$ such that at least a vertex pair included in Q is disconnected in $G' = (V, E - E_K)$ where $E_P \cap E_K = \emptyset$, be $K(i, E_P)$. When E_P is the set of the protected edges, the network failure probability $C(E_P)$ be described as follows:

$$\sum_{i=1}^{m} \sum_{E_K \in K(i,E_P)} \prod_{\forall e \in E_K} h(e) \prod_{\forall e' \notin E_K \cup E_P} (1 - h(e'))$$

$C(E_P)$ is the probability that, when each non-protected edge independently and probabilistically fails, there are at least a disconnected vertex pairs in the resulting network. $C(E_P)$ is called the network failure probability when E_P is the set of the protected edges.

Link Protection Problem Considering Failure Probability (PPP)
Input: *Network* $N = (G{=}(V, E), h)$ *Set of vertex pairs* Q, *a positive integer* p.
Output: *Set of protected edges,* $E_P(\subseteq E)$.
Constraint: $|E_P| \leq p$
Objective: $C(E_P)$ *(minimize)*

When $Q = \{\{s, t_1\}, \{s, t_2\}, \ldots, \{s, t_r\}\}$, vertex s corresponds to the master server and vertices $\{t_1, t_2, \ldots, t_r\}$ correspond to the edge servers, and the solution of PPP corresponds to the set of the protected links to minimize the network failure probability such that at least a pair of the master server and an edge server are disconnected.

We show an example of PPP as follows (see Figs. 1, 2 and 3). The failure probabilities of the edges are given in Fig. 1; $Q = \{\{s, t_1\}, \{s, t_2\}, \{s, t_3\}\}$; $p = 1$. The thick edges are protected edges in Figs. 2 and 3. In both figures, all the events that at least a vertex pair is disconnected and their probability are enumerated. The sum of the probabilities means the network failure probability. In Fig. 2, the network failure probability is 0.212; on the other hand, in Fig. 3, the network failure probability is 0.154, which is less than that of Fig. 2. This means that the protected edges in Fig. 3 is better than those in Fig. 2 from the viewpoint of the network failure probability. In fact, the protected edges in Fig. 3 is the optimum solution of this example.

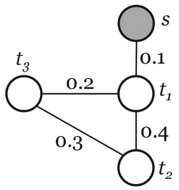


Fig. 1. Network and failure probability of edges.

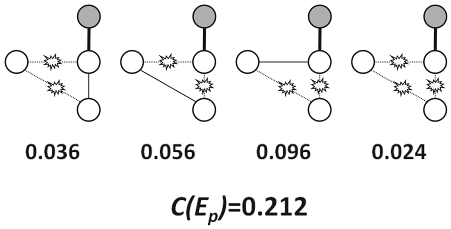


Fig. 2. A protected edge for network in Fig. 1.

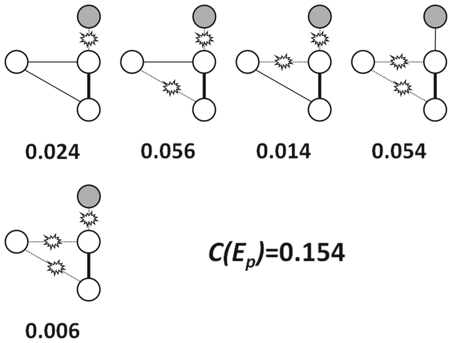


Fig. 3. Another protected edge for network in Fig. 1.

The following theorem holds.

Theorem 1. *PPP is generally NP-hard.*

Proof. We describe the problem PLLA [6] as follow: Let $Q = \{(q_1 = \{s_1, t_1\}, d_1), (q_2 = \{s_2, t_2\}, d_2), \ldots, (q_r = \{s_r, t_r\}, d_r)\}$ be the demand set, where, for each pair $q_i = \{s_i, t_i\}$ of vertices $(i = 1, 2, \ldots, r)$, the demand between s_i and t_i is $d_i \in \mathbb{R}^+$. We associate the weight and capacity with each edge by weight function $w : E \to \mathbb{R}^+$ and capacity function $c : E \to \mathbb{R}_0^+$, and let a network be $N = (G = (V, E), c, w)$. For network $N = (G = (V, E), c, w)$ and demand set $Q = \{(q_1, d_1), (q_2, d_2), \ldots, (q_r, d_r)\}$, when $m(e) \le c(e)$ for all edges $e \in E$, the *capacity constraint* is satisfied. When s_i and t_i is connected for all $q_i = \{s_i, t_i\}$ $(i = 1, 2, \ldots, r)$, the *reachability constraint* is satisfied. When both the capacity and reachability constraints are satisfied, we say that the network and demand set are feasible. When $E_P(\subseteq E)$ satisfies that, for any $E_K(\subseteq E)$ such that $|E_K| \leq k$ for integer $k(\geq 0)$ and such that $E_K \cap E_P = \emptyset$, network $N' = (G' = (V, E - E_K), c, w)$ and demand set $Q = \{(q_1, d_1), (q_2, d_2), \ldots, (q_r, d_r)\}$ satisfies the capacity and reachability constraints, E_P is called k-protected edge set.

PLLA
INSTANCE: *Network* $N = (G = (V, E), c, w)$*, demand set* $Q = \{(q_1, d_1), (q_2, d_2), \ldots, (q_r, d_r)\}$*, integers* p, k (≥ 0)*.*
QUESTION: *Is there the* k*-protected edge set* $E_P(\in E)$ *such that* $|E_P| \leq p$*?*

PLLA is NP-hard, even if the demand of all pairs is one and the capacity of all edges is sufficiently large [6]. The restricted problem of PPP such that the failure probability of all edge is one corresponds to this restricted problem of PLLA. Therefore, PPP is also NP-hard. □

3 Heuristic Algorithms for PPP

Since PPP is NP-hard, we cannot expect a polynomial time algorithm for PPP. In this section, we present two heuristic algorithms to solve PPP in polynomial time.

For network $N = (G = (V, E), h)$ and set of vertex pairs Q, a minimal set $E_s \subseteq E$ such that at least a pair in Q is disconnected in $((V, E \backslash E_s), h)$ is called the *minimal pseudo cut*. The product of the failure probability of all the edges included in a minimal pseudo cut is called the weight of the minimal pseudo cut.

For graph $G = (V, E)$, we consider hyper graph $H = (E, E_H)$ that the vertex set is E and the edge set E_H is the set of all the minimal pseudo cuts of G. If $E'(\subseteq E)$ whose size is less than or equal to p is a vertex cover of H, each minimal pseudo cut has at least an edge of E' for all the minimal pseudo cuts; therefore, the vertices in all the pairs in Q are connected in the graph resulted from the removal of any set of edges except the edges of E'. Even if $E'(\subseteq E)$ whose size is less than or equal to p is not a vertex cover of H, when the edges in E' are included in many minimal pseudo cuts and the sum of the weight of the minimal pseudo cuts not including E' is small, $C(E_P)$ is also small.

For $e \in E$, let the minimal pseudo cuts including e be $\{E_1, E_2, \ldots, E_k\}$. Let $P(e)$ be the sum of the failure probability that all the edges in E_i are removed, $\sum_{i=1}^{k} \prod_{\forall f \in E_i} h(f) \prod_{\forall f' \notin E_i} (1 - h(f'))$. By protecting edge e, all the edges contained in any minimal pseudo cut in $\{E_1, E_2, \ldots, E_k\}$ are not simultaneously removed; therefore, the value of $P(e)$ can be reduced from the network failure probability. It implies that the protection of edges with large value of $P(e)$ leads to the efficient reduction of the network failure probability.

We show the basic idea of the algorithms based on the above consideration. The first algorithm sorts all edges in descending order of $P(e)$ $(e \in E)$ and protects p largest edges. The second algorithm iterates p times the following procedure: the edge e^* of the largest $P(e)$ is chosen and protected. Then, the set of the minimal pseudo cuts is recalculated. Note that, when an edge is protected,

Algorithm 1. Algorithm Simple-PPP-general

Input: Network $N = (G{=}(V, E), h)$, $|V| = n, |E| = m$, Set of vertex pairs Q, a positive integer p.
Output: E_P.
$E_P \leftarrow \emptyset$, $D \leftarrow \emptyset$
for $i = 1$ **to** m **do**
  while *there is an unscanned set of edges with the size of* i, E_s **do**
    if *vertices in at least a pair in* Q *is disconnected, when all the edges in* E_s *are removed* **then**
      if *no subset of* E_s *is included in* D **then**
        $D \leftarrow E_s$
while *there is an unscanned edge* $e \in E$ **do**
  Let $\{E_1, E_2, \ldots, E_k\}(\subseteq D)$ be the set of the minimal psuedo cuts including e.
  $P(e) \leftarrow \sum_{i=1}^{k} \prod_{\forall f \in E_i} h(f) \prod_{\forall f' \notin E_i} (1 - h(f'))$
Sort $P(e)$ $(e \in E)$ in descending order.
Let E_P be the set of p edges in the order of the large $P(e)$ $(e \in E)$.
return E_P

Algorithm 2. Algorithm PPP-general

Input: Network $N = (G{=}(V, E), h)$, $|V| = n, |E| = m$, Set of vertex pairs Q, a positive integer p.
Output: E_P.
$E_P \leftarrow \emptyset$, $D \leftarrow \emptyset$
for $i = 1$ **to** p **do**
  for $i = 1$ **to** m **do**
    while *there is an unscanned set of edges with the size of* i, E_s **do**
      if *vertices in at least a pair in* Q *is disconnected, when all the edges in* E_s *are removed* **then**
        if *no subset of* E_s *is included in* D **then**
          $D \leftarrow E_s$
  while *there is an unscanned edge* $e \in E$ **do**
    Let $\{E_1, E_2, \ldots, E_k\}(\subseteq D)$ be the set of the minimal psedo cuts including e.
    $P(e) \leftarrow \sum_{i=1}^{m} \prod_{\forall f \in E_i} h(f) \prod_{\forall f' \notin E_i} (1 - h(f'))$
  $E_P \leftarrow$ the edge e whose $P(e)$ is the largest.
return E_P

since the failure probability of the edge changes to zero, the minimal pseudo cuts generally changes. Subsequently, $P(e)$ $(e \in E \backslash \{e^*\})$ are also recalculated.

These algorithms cannot always output the optimal solution, because $P(e)$ $(e \in E)$ is not mutually independent. However, since these algorithms protect edges e whose $P(e)$ is comparatively large, it is expected that the network failure probability becomes small.

4 Performance Evaluation

In this section, we evaluate the performance of the proposed two algorithms in Sect. 3 by using the graph structures of actual ISP backbone networks provided by CAIDA (Center for Applied Internet Data Analysis) [7].

In this section, $Q = \{\{s, t_1\}, \{s, t_2\}, \ldots, \{s, t_r\}\}$, vertex s is the master server, and vertices $\{t_1, t_2, \ldots, t_r\}$ are the edge servers. The failure probability is randomly determined in the range of $(0, 0.1]$.

We examined the relationship between the number of the protected edges and the network failure probability of the proposed two algorithms.

First, in Table 1(a), we observed how the network failure probability changes according to the increase of the number of the protected edges.

Table 1. Relation between network failure probability and the number of protected edges.

(a) Decrease of network failure probability according to increase of the number of protected edges

m	n	p	$C(E_p)$	
			Simple-PPP-general	PPP-general
20	22	2	0.4118	0.4118
		6	0.2281	0.2281
		10	0.1835	0.0979
		14	0.0896	0.0127
22	25	2	0.4071	0.4071
		6	0.1986	0.1986
		10	0.1464	0.0527
		14	0.0472	0.0023
16	23	2	0.0255	0.0255
		6	0.0044	0.0031
		10	0.0029	0.0000
		14	0.0008	0.0000
28	30	2	0.5679	0.5679
		6	0.4082	0.4082
		10	0.2393	0.2393
		14	0.1862	0.1134

(b) Increase of the number of protected edges according to decrease of network failure probability.

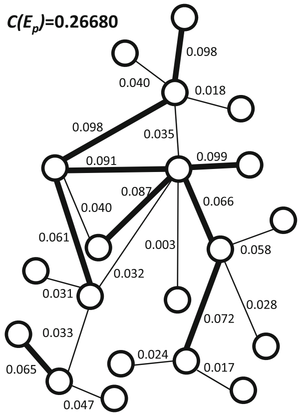
m	n	$C(E_p)$	p	
			Simple-PPP-general	PPP-general
20	22	0.1	14	10
		0.05	16	12
		0.01	21	15
		0.005	21	17
22	25	0.1	12	9
		0.05	14	11
		0.01	18	13
		0.005	18	14
16	23	0.001	14	8
		0.0005	16	9
		0.0001	16	9
		0.00005	16	10
28	30	0.1	22	15
		0.05	24	18
		0.01	29	23
		0.005	30	24

The network failure probability decreases according to the increase of the number of the protected edges in both the algorithm Simple-PPP-general and the algorithm PPP-general. The network failure probability by algorithm PPP-general is less than that of algorithm Simple-PPP-general. In the case that $m =$ 20, $n = 22$, $p = 2$, the algorithm Simple-PPP-general needed 22.870 s to output the result; on the other hand, PPP-general needed 4 min 13.610 s to output the result (HP Z800, CPU Intel Xeon X5687 3.6 GHz, 8 GB).

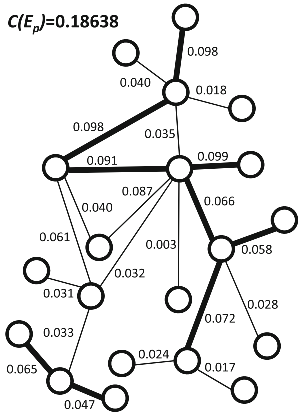
Next, in Table 1(b), we observed how the number of the protected edges changes according to the decrease of the network failure probability.

The number of the protected edges rapidly increases according to the decrease of the required network failure probability in both the algorithm Simple-PPP-general and the algorithm PPP-general. The number of the protected edges by the algorithm PPP-general is less than that of the algorithm Simple-PPP-general.

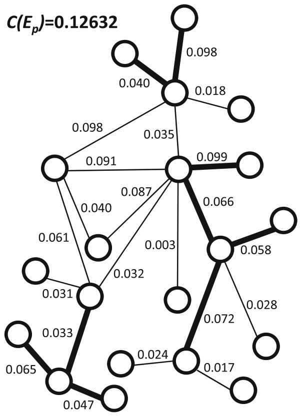
In Fig. 4(a), (b) and (c), we show the topology of the actual networks and the protected edges. The value by each edge is the failure probability of the edge



(a) Descending order of the failure probability of edges



(b) Algorithm Simple-PPP-general



(c) Algorithm PPP-general

Fig. 4. Protected edges.

and the value of $C(E_P)$ is the network failure probability, and the thick line is a protected edge. In these examples, we determined the protected edges whose number of edges is nine or less by the algorithms. Figure 4(a) shows the result by the algorithm that the protected edges are determined by the descending order of the failure probability, Fig. 4(b) shows the result by the algorithm Simple-PPP-general, and Fig. 4(c) shows the result by the algorithm PPP-general.

The network failure probability in Fig. 4(a) is larger than the other examples. It implies that only edges with large failure probability are not always important to improve the reliability, but the graph structure is also important. For example, the protected edges in Fig. 4(b) and (c) tend to be included in a cut with small size; if the size of the cut disconnecting the vertices in a pair is small, since the network tends to be infeasible, the edges included in such a cut tend to be protected. Indeed, in Fig. 4(b) and (c), even if the failure probability of an edge is small, when the edge is important from the viewpoint of the connectivity, it tends to be protected.

The network failure probability in Fig. 4(c) is less than that in Fig. 4(b). It means that the performance of the algorithm PPP-general is better than that of the algorithm Simple-PPP-general. However, the computation time of the algorithm PPP-general is larger than that of the algorithm Simple-PPP-general. The computation time is at most about 5 min, so that we can use the algorithm PPP-general for a network of this size; however, the computational complexity is exponential order, so that it is difficult to use the algorithm for a large-size network. It is necessary to design a more efficient algorithm.

5 Conclusion

In this paper, we focused on the design of the reliable content distribution service to decrease the probability of service interruption due to disconnection of the communication path between the master server and the edge servers even in the event of a failure. We defined and analyzed a new network design problem from the viewpoint of link protection. This problem takes the failure probability into consideration, and it determines the smallest number of the protected links so that, even if any non-protected links fail, the failure probability of a whole network in the resulting network after links fail is below a given threshold. First, we formulated this problem as the optimization problem and proved that this problem is NP-hard. Furthermore, we introduced a new concept of the graph theory, pseudo cut, and we proposed two heuristic algorithms to solve the problem based on the pseudo cut. In addition, we showed some examples of application of the heuristic algorithms to the topology of the various actual networks and evaluated the performance of the algorithms. As a result, it worked for the topology of the actual networks. In addition, we found that the algorithm recalculating pseudo cuts for every detection of a protected link can find smaller number of protected links to achieve the same network failure probability condition than the other algorithm.

Acknowledgment. This work was partially supported by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (B) (17H01742).

References

1. Arulselvan, A., Commander, C.W., Elefteriadou, L., Pardalos, P.M.: Detecting critical nodes in sparse graph. Proc. Comput. Oper. Res. **36**(7), 2193–2200 (2009)
2. Imagawa, K., Miwa, H.: Detecting protected links to keep reachability to server against failures. In: Proceedings of ICOIN 2013, Bangkok, 28–30 January 2013
3. Irie, D., Kurimoto, S., Miwa, H.: Detecting critical protected links to keep connectivity to servers against link failures. In: Proceedings of NTMS 2015, Paris, 27–29 July 2015, pp. 1–5 (2015)
4. Yamasaki, T., Anan, M., Miwa, H.: Network design method based on link protection taking account of the connectivity and distance between sites. In: Proceedings of INCoS 2016, Ostrava, 7–9 September 2016, pp. 339–343 (2016)
5. Koide, T., Shinmori, S., Ishii, H.: The evaluations on lower bounds of all-terminal reliability by arc-packings for general networks. IEICE Trans. Fund. **E82-A**(5), 784–791 (1995)
6. Noguchi, A., Fujimura, T., Miwa, H.: Network design method by link protection for network load alleviation against failures. In: Proceedings of INCoS 2011, Fukuoka, Japan, 30 November–2 December 2011, pp. 581–586 (2011)
7. CAIDA. http://www.caida.org/

Network Design Method by Link Protection to Keep Connectivity and Communication Quality to Servers

Daishi Irie(✉) and Hiroyoshi Miwa

Graduate School of Science and Technology, Kwansei Gakuin University,
2-1 Gakuen, Sanda-shi, Hyogo, Japan
{daishi.irie,miwa}@kwansei.ac.jp

Abstract. Information network must be sufficiently reliable. The most reliable method is to protect all links so that the failure probability of a link is sufficiently small by sufficient backup resource and rapid recovery system. However, as link protection needs much cost, it is necessary to find the smallest number of links to be protected so that a network resulting from failures of any non-protected links can provide sufficient connectivity. Especially, in content delivery services, a request from a user is navigated to one of the mirror servers so that small delay time and balance of loads can be achieved; the network topology must be designed so that the navigation can be efficiently achieved even after failure. In this paper, we focus on the problem of finding the protected links so that the conditions that include the stretch factor and the fragmentation factor are satisfied. First, we formulate this problem and prove that this problem is NP-hard. Next, we present a polynomial-time algorithm to solve the problem that the number of the simultaneous link failures is restricted to one. Furthermore, we present a polynomial-time approximation algorithm with the approximation ratio of a constant that is the number of the simultaneous link failures. In addition, we apply the approximation algorithms to the topology of actual networks and evaluate the approximation ratio.

1 Introduction

Information network must be sufficiently reliable. The users require more network resource to use multimedia applications, so that content delivery service is getting more important; consequently, service continuity must be ensured. Especially, service interruptions by disconnection of communication paths in a network even after failure, must be avoided. Therefore, it is necessary for network service providers to design and operate reliable network which is robust against network failures.

The edge-connectivity and the vertex-connectivity of a graph have been extensively used as the measures for evaluation of the reliability of a network. As

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_38

for a network design method to improve a network that is lack of the connectivity, there is a method in which the connectivity is increased to improve reliability by augmenting the edges of the graph [1,2]. Another measure is the number of connected components in a network when the network becomes disconnected due to a failure [3].

The most reliable method is to protect all links so that the failure probability of a link is sufficiently small by sufficient backup resource and rapid recovery system. Indeed, we can make the failure probability of a link in an IP layer sufficiently small by fast switch function and backup resource reserved in advance in a lower layer, although backup resource must avoid sharing network facility such as fibers and routers in the lower layer so as to give no impact on the Internet services. If such a recovery system is provided in the lower layer, we can consider that a link between two IP routers does not fail, because a failure of the link is rapidly recovered and the failure cannot be detected in the IP layer. Such a link is called a *protected link*. However, link protection needs much cost. As it is not practical to protect all links, only critical links whose failures significantly degrade the performance of a network should be protected. Therefore, it is necessary to find the smallest number of links to be protected so that a network resulting from failures of any non-protected links provides connectivity.

Related to this problem, the problem of determining protected links so that all connected components include specified servers in a resulting network after links fail was investigated [4]. However, it is not sufficient to apply the proposed network design method to an actual network, because the performance such as delay time and load balance should be considered.

In content delivery service, generally, a request from a user is navigated to one of the mirror servers so that small delay time and balance of loads can be achieved; therefore, the network topology must be designed so that the navigation can be efficiently achieved even after a failure. In other words, the stretch factor, the maximal ratio of the distance between nodes in an original network to that in the resulting network after a failure must be small so that the delay will not become large; the fragmentation factor, that is, the ratio of the size of the maximal connected component to the size of the minimal connected component must be small after a failure so that the resulting network after a failure will not be fragmented to unbalanced connected components.

In this paper, we focus on the problem of finding the protected links so that the following conditions are satisfied: even if any non-protected links fail, all connected components in the resulting network include specified servers, and the stretch factor and the fragmentation factor are not more than a threshold respectively.

We show an example. Figure 1 shows a network, where the servers are located in node 1 and node 4. In Fig. 2, when links $(1,2),(1,3)$ simultaneously fail, since the shortest path from node 2 to node 1 that is the nearest servers is disconnected, the nearest server from node 1 changes to node 4; consequently, the distance increases from one to three. If the increase is restricted to twice or less from the viewpoint of QoS, this constraint is not satisfied. In addition, after

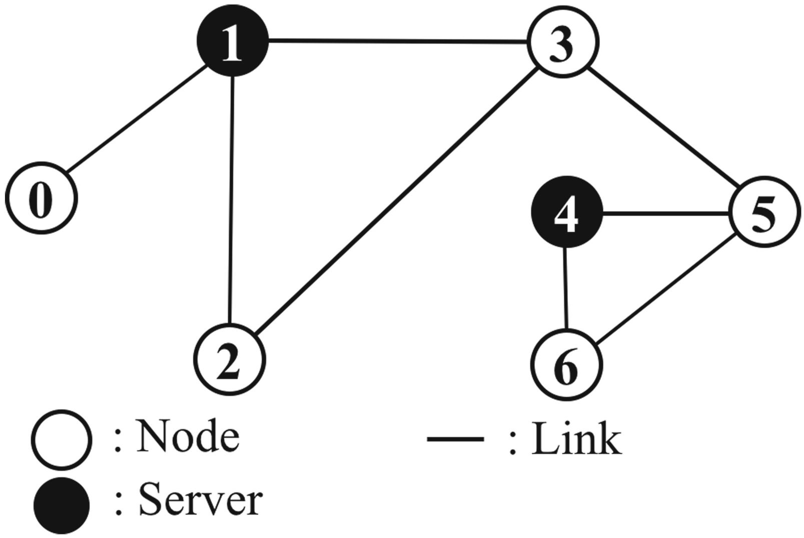


Fig. 1. A network and servers.

the link failure, the server in node 4 must receive accesses from node 2, node 3, node 5, and node 6; on the other hand, the server in node 1 receives accesses from only node 0, and the server load is unbalanced. If the maximal ratio of the size of the connected components including servers is restricted to two or less from the viewpoint of QoS, this constraint is not satisfied. On the other hand, when the links indicated by bold line are protected (Fig. 3), even if any two links except protected links fail, the constraints are always satisfied.

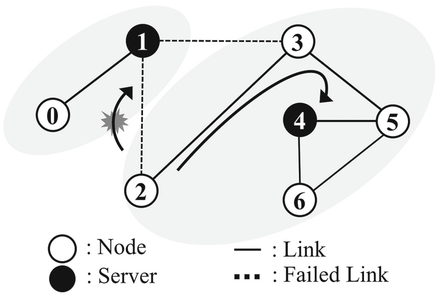


Fig. 2. Increase of path length and unbalanced fragmentation by link failure.

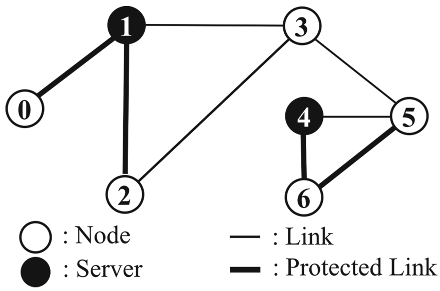


Fig. 3. Protected links.

The reference [5] deal with the problem of determining a set of protected links that keep the reachability to a specific node, and an approximation algorithm is proposed. The problem of determining protected links so that between the master server and all edge servers are connected and that the capacity constraint is satisfied and that the increase ratio of the distance in a network to the distance in a failed network does not exceed a threshold, is investigated in [6].

The reference [4] deals with the problem of determining protected links so that all connected components include specified servers in a resulting network after links fail. This problem is the original version of the problem in this paper. The reference [7] deals with the problem of determining server placement and link protection simultaneously. The problem in this paper considers not only the connectivity but also delay time and load balance, and it is an extended version of the problem in [4].

In this paper, first, we formulate this link protection problem and show that this problem is NP-hard. Next, we present a polynomial-time algorithm to solve the problem that the number of the simultaneous link failures is restricted to one. Furthermore, we present a polynomial-time approximation algorithm with the approximation ratio of a constant, the number of the simultaneous link failures. In addition, we apply the approximation algorithms to the topology of actual networks and evaluate the approximation ratio. The results indicate that the approximation ratio in the actual networks is much smaller than the theoretical upper bound. It follows that our algorithms are efficient when we determine protected links in an actual network.

2 Link Protection Problem to Keep Connectivity to Servers

Let the network be the graph $G = (V, E)$ where V is vertex set of G and E is edge set of G. Let the upper bound of the number of protected edges be p, the number of removed edges be k, and the lower bound of the number of servers included in each connected components be L.

The path length is defined as the number of all the edges included in the path. The shortest path between two vertices is defined as the path with the smallest path length among all paths between the vertices. The distance between two vertices v and w in $G, d_G(v, w)$ is defined as the length of the shortest path between v and w.

A subset of edges $E_P(\subseteq E)$ corresponds to a set of protected links, a subset of edges $E_K(\subseteq E)$ corresponds to a set of failed links, a subset of vertices $W(\subseteq V)$ corresponds to a set of servers, and $G_r = (V, E \backslash E_K)$ corresponds to the network resulting from the link failure.

The stretch factor of $G_r = (V, E \backslash E_K)$, $sf(G_r)$, is defined as $\max_{v,w} d_{G_r}(v, w)/d_G(v, w)$; the fragmentation factor of G_r, $ff(G_r)$ is defined as the ratio of the size of the maximal connected component to the size of the minimal connected component. Let the upper bound of $sf(G_r)$ be b, let the upper bound of $ff(G_r)$ be r.

When all connected components of $G_r = (V, E \backslash E_K)$ includes at least L vertices of W, $sf(G_r) \leq b$, and $ff(G_r) \leq r$, for all E_K such that $E_K \subseteq E, |E_K| \leq k$, $E_P \cap E_K = \emptyset$, E_P is called (k, L, b, r, W)-protected edges. The condition that $E_P \cap$

$E_K = \emptyset$ means that the protected links do not fail. If E_P is (k, L, b, r, W)-protected edges, even if any non-protected k edges simultaneously removed, all connected components in the resulting graph include at least L vertices of W, $sf(G_r) \leq b$, and $ff(G_r) \leq r$.

Link Protection Problem to Keep Connectivity to Servers (LPPCS)
INSTANCE: *A connected graph* $G = (V, E)$, $W \subseteq V$, *positive integers* p, k, L, b, r.
QUESTION: *Is there* (k, L, b, r, W)*-protected edges* E_P *such that* $|E_P| \leq p$*?*

In the network whose protected links are determined by the solution of LPPCS, even if any non-protected k links simultaneously fail, all connected components in the resulting network include at least L servers, the increase of delay is small, and the load to the server in each connected component in the resulting network is balanced.

We show an example in Fig. 4. In this example, $p = 5, k = 3, L = 1, b = 2$, $r = 2$. The vertices of W is vertex 1 and vertex 5. The solution of this instance, $(3, 1, 2, 2, W)$-protected edges, is shown by five bold edges. Even if any three edges are removed except the protected edges, the constraint is satisfied.

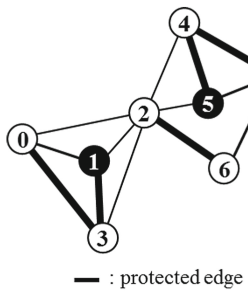


Fig. 4. An instance of LPPCS ($p = 5, k = 3, L = 1, b = 2, r = 2$).

When b is restricted to $|E|$ and r is restricted to $|V|$, LPPCS is equivalent to PLCS [4]; therefore, LPPCS is NP-hard.

Theorem 1. *LPPCS is generally NP-hard.*

3 Polynomial-Time Algorithms

In this section, we address the optimization version of LPPCS whose objective function is to minimize the number of protected edges. In the rest of this paper, we refer to the optimization version as LPPCS.

First, we show that, when $k = 1$, LPPCS can be solved in polynomial time.

Algorithm 1. Algorithm ProtectEdgesSLF

Input: A connected graph $G = (V, E)$, $W \subseteq V$, positive integers k, L, b, r.
Output: E_P.

```
E_P ← ∅
for each edge e ∈ E do
    if there is a connected component in G_e = (V, E\{e}) that does not include
    L or more vertices of W, sf(G_r) > b, or ff(G_r) > r then
        E_P ← E_P ∪ {e}
return E_P
```

E_P output by the algorithm **ProtectEdgesSLF** is $(1, L, b, r, W)$-protected edges, because, if at least an edge of E_P is removed, the constraint is not satisfied. It takes $O(n + m)$ time to enumerate all connected components and check whether they includes L or more vertices of W. It takes $O(n + m) \times O(n^2) = O(n^3 + mn^2)$ to check $sf(G_r) > b$; it takes $O(n+m)$ to check $ff(G_r) > r$. As all edges are checked, this procedure is iterated m times. In total, the computational complexity of the algorithm **ProtectEdgesSLF** is $O(n^3m + m^2n^2)$. Thus, we have the following theorem.

Theorem 2. *Algorithm* **ProtectEdgesSLF** *outputs the smallest number of the protected edges for LPPCS when $k = 1$ in $O(n^3m + m^2n^2)$ time, where n is the number of vertices and m is the number of edges.*

When $k(\geq 2)$ is a constant integer, LPPCS has the polynomial-time approximation algorithm with the approximation ratio of k.

The basic idea is the reduction from LPPCS to the vertex cover problem (VC) of hyper graph. First, the algorithm **ProtectEdges-k-apx** converts an instance of LPPCS to an instance of VC. Then **approxVC**(H) that is based on a maximal matching algorithm for a hyper graph outputs the approximation solution of the instance of VC. The maximal matching algorithm chooses a hyper edge. If it includes no vertex of the vertex cover, then all the vertices in the hyper edge are added to the vertex cover; otherwise, the algorithm chooses another hyper edge. This procedure is iterated until all hyper edges are checked. At each iteration, a hyper edge is chosen in the descending order of the number of the hyper edges sharing vertices with other hyper edges. As many hyper edges are covered in the early stages by this greedy procedure, we can expect that the size of the vertex cover is small. The maximal matching algorithm guarantees the approximation ratio of k, the size of a hyper edge, regardless of the order of choosing hyper edges. Finally, the algorithm **ProtectEdges-k-apx** converts the solution to the solution of the instance of LPPCS.

Algorithm 2. Algorithm ProtectEdges-k-apx

Input: A connected graph $G = (V, E)$, $W \subseteq V$, positive integers k, L, b, r.
Output: E_P.

```
E_P ← ∅
i ← 1
while i ≤ k do
    E_H ← ∅
    for all E'(∈ E\E_P, |E'| = i) do
        if there is a connected component in G_{E'} = (V, (E\E_P) \ E') that does
        not include L or more vertices of W, sf(G_r) > b, or ff(G_r) > r then
            E_H ← E_H ∪ E'
    E_i ← approxVC(H_i = (E, E_H))
    E_P ← E_P ∪ E_i
    i ← i + 1
return E_P
```

Algorithm 3. Procedure approxVC(H)

Input: Hyper graph $H = (V, E)$.
Output: A vertex cover V'.

```
V' ← ∅
Label "uncovered" for all vertices v ∈ V
Sort the hyper edges {e_1, e_2, ..., e_h}(∈ E, |E| = h) in the descending order of
the number of the hyper edges sharing vertices with other hyper edges
for i = 1 to h do
    if all vertices in e_i are "uncovered" then
        Label "covered" to all vertices in e_i
        V' ← V' ∪ {all vertices in e_i}
return V'
```

Theorem 3. *Algorithm* **ProtectEdges-k-apx** *is the approximation algorithm in polynomial time with approximation ratio of k when k is a constant positive integer.*

4 Performance Evaluation

In this section, we evaluate the performance of the approximation algorithm in Sect. 3 for the graph structures of actual ISP backbone networks [8].

First, we evaluated the average of the number of protected edges for 100 randomly chosen W ($|W| = 1, 2, 3$) under the condition that $b = 1000, r = 1000$ and the condition that $b = 2.0, r = 1.5$, respectively. We show the results in Table 1, where n is the number of vertices; m is the number of edges; L is one. The values of $p2$ and $p3$ are the averages of the number of protected edges by **AlgorithmProtectEdges-k-apx** when $k = 2$ and $k = 3$, respectively.

Table 1. ISP backbone networks and protected edges.

$b = 1000, r = 1000$									$b = 2.0, r = 1.5$								
No.	n	m	$\|W\|$	$p1$	$p2$	$p3$	$p2opt$	$p3opt$	No	n	m	$\|W\|$	$p1$	$p2$	$p3$	$p2opt$	$p3opt$
7	35	50	1	2	30.14	39.15	22	27	7	35	50	1	6.48	31.56	41.82	24.72	29.75
			2	1.81	28.68	36.66	20.68	25.55				2	7.1	32.38	40.68	25.14	29.54
			3	1.8	28.02	35.61	20.15	24.87				3	7.08	32.12	39.9	24.86	29.2
9	19	33	1	0	12.25	15.93	7	10	9	19	33	1	1.39	15.05	20.88	9.41	12.93
			2	0	10.96	15.12	6.25	9.09				2	1.44	15.1	20.46	9.21	12.91
			3	0	10.4	14.79	5.94	8.63				3	1.02	14.7	19.83	8.73	12.26
10	37	44	1	3	31	39	27	32	10	37	44	1	10.87	36.51	41.22	31.82	34.94
			2	2.88	30.16	36.81	24.95	30.14				2	11.38	36.5	41.16	31.52	34.33
			3	2.83	29.71	36.09	24.18	29.48				3	11.83	36.57	40.8	30.96	33.69
31	28	56	1	6	27.69	30.39	19	20	31	28	56	1	11.07	25.47	25.86	19.42	20.31
			2	5.54	23.56	25.71	17.58	18.78				2	11.06	25.34	25.2	19.26	20.28
			3	5.37	22.44	24.42	16.89	17.93				3	11.15	25.55	25.47	19.34	20.35
33	22	39	1	3	15.28	18.54	9	12	33	22	39	1	4.61	17.73	20.94	11.4	14.53
			2	2.7	13.86	16.53	8.27	11.01				2	4.59	17.15	21.42	11.09	14.58
			3	2.71	13.15	15.39	7.93	10.45				3	4.12	16.9	21.57	10.77	14.43

The values of $p1$, $p2opt$, and $p3opt$ are the averages of the smallest number of protect edges when $k = 1, 2, 3$, respectively.

The number of the protected edges decreases when $b = 1000, r = 1000$ according to the increase of the size of W; on the other hand, the number of the protected edges does not always decreases when $b = 2.0, r = 1.5$. In general, the increase of the size of W means the relaxation of the constraints, but, when the constraints of b and r are severe, the constraints are not always eventually relaxed.

Next, we compared the proposed algorithm (ProposedAlg) and the algorithm (SimpleAlg) in [4]. In this experiment, $b = 2.0, r = 1.5$. The original algorithm in [4] does not consider b and r, but we modified the algorithm. The essential difference between ProposedAlg and SimpleAlg is the method of choosing hyper edges in the procedure **approxVC**. SimpleAlg chooses a hyper edge randomly; however, ProposedAlg chooses it in the specified order. The average of the approximation ratio of SimpleAlg is 1.28 and 1.44 when $k = 2, 3$, respectively; on the other hand, the average of the approximation ratio of ProposedAlg is 1.26 and 1.29 when $k = 2, 3$, respectively. The proposed algorithm can output the better solution than the previous algorithm. The both algorithms have the theoretical upper bound of the approximation ratio of 2 and 3 when $k = 2, 3$, respectively; however, the average of the approximation ratio is much smaller when the algorithms are applied to the topology of the actual networks.

Next, we evaluated the change of the number of protected edges and the approximation ratio when b increases. In this experiment, we show the average number of protected edges for 100 times randomly chosen W ($|W| = 1, 2, 3$), where $k = 3$, $r = 1000$, the range $[1.25, 5.00]$ of b. We show the results in Figs. 5 and 6. Figure 5 shows the relationship between the value of b and the average of

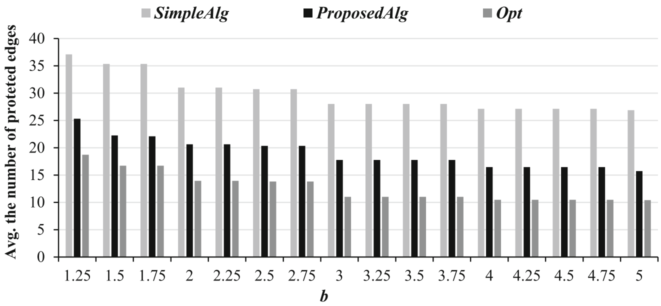


Fig. 5. The relationship between the value of b and the average of the number of protected edges.

the number of protected links for a network in [8]. In Fig. 5, we show also the optimal solution (Opt), that is, the smallest number of protected edges. Figure 6 shows the relationship between b and the average of the approximation ratio.

In Fig. 5, we can observe that the number of protected edges decreases according to the increase of the value of b. The number of the protected edges by the proposed algorithm (ProposedAlg) is always much smaller than that by the previous Algorithm (SimpleAlg). Especially, the number of the protected edges by ProposedAlg in the case that $b = 1.25$ is smaller than that by SimpleAlg in the case that $b = 5.00$. In general, when b is small, as the constraint is severe, the number of the protected edges is large. The above result means that the proposed algorithm can output a better solution than the previous algorithm even if the constraint is severe. In Fig. 6, we can observe that the approximation ratio by the proposed algorithm is much smaller than that by the previous algorithm; regardless of the value of b, the approximation ratio by the proposed algorithm is about 1.5, which is much smaller than the theoretical upper bound.

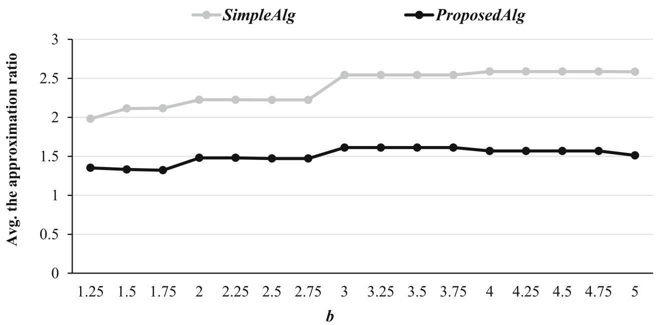


Fig. 6. The relationship between the value of b and the average of the approximation ratio.

5 Conclusion

In content delivery service, generally, a request from a user is navigated to one of the mirror servers so that small delay time and balance of loads can be achieved; therefore, the network topology must be designed so that the navigation can be efficiently achieved even after a failure.

In this paper, we defined and analyzed the problem of finding the protected links so that the following conditions are satisfied: even if any non-protected links fail, all connected components in the resulting network include specified servers, and the stretch factor and the fragmentation factor are not more than a threshold respectively. The small stretch factor corresponds to the small increase ratio of delay; the small fragmentation factor means that the resulting network after failure will not be fragmented to unbalanced connected components. First, we formulated this link protection problem and proved that this problem is NP-hard. Next, we presented a polynomial-time algorithm to solve the problem that the number of the simultaneous link failures is restricted to one. Furthermore, we presented a polynomial-time approximation algorithm with the approximation ratio of a constant, the number of the simultaneous link failures. In addition, we applied the approximation algorithm to the topology of actual networks and evaluated the approximation ratio. The results indicated that the approximation ratio in the actual networks is much smaller than the theoretical upper bound. In addition, the proposed algorithm considering the order of choosing hyper edges achieved the better approximation ratio than the previous algorithm, when the algorithms are applied to the topology of actual networks.

Acknowledgment. This work was partially supported by by the Japan Society for the Promotion of Science through Grants-in-Aid for Scientific Research (B) (17H01742).

References

1. Frank, A.: Augmenting graphs to meet edge-connectivity requirements. In: Proceedings of 31st Annual Symposium on Foundations of Computer Science, St. Louis, 22–24 October 1990
2. Ishii, T., Hagiwara, M.: Minimum augmentation of local edge-connectivity between vertices and vertex subsets in undirected graphs. Discrete Appl. Math. **154**(16), 2307–2329 (2006)
3. Arulselvan, A., Commander, C.W., Elefteriadou, L., Pardalos, P.M.: Detecting critical nodes in sparse graph. Proc. Comput. Oper. Res. **36**(7), 2193–2200 (2009)
4. Irie, D., Kurimoto, S., Miwa, H.: Detecting critical protected links to keep connectivity to servers against link failures. In: Proceedings of NTMS 2015, Paris, 27–29 July 2015, pp. 1–5 (2015)
5. Imagawa, K., Miwa, H.: Detecting protected links to keep reachability to server against failures. In: Proceedings of ICOIN 2013, Bangkok, 28–30 January 2013

6. Yamasaki, T., Anan, M., Miwa, H.: Network design method based on link protection taking account of the connectivity and distance between sites. In: Proceedings of INCoS 2016, Ostrava, 7–9 September 2016, pp. 339–343 (2016)
7. Irie, D., Anan, M., Miwa, H.: Network design method by finding server placement and protected links to keep connectivity to servers against link failures. In: Proceedings of INCoS 2016, Ostrava, 7–9 September 2016, pp. 439–344 (2016)
8. CAIDA. http://www.caida.org/

When Does Network Coding Benefit Store-Carry-and-Forwarding Networks in Practice?

Rabenirina Aina Tsiory[1(✉)], Masato Tsuru[1], and Agussalim[2]

[1] Kyushu Institute of Technologie, Fukuoka, Japan
aina@infonet.cse.kyutech.ac.jp, tsuru@cse.kyutech.ac.jp
[2] STMIK Handayani, Makassar, Indonesia
agoesalim@gmail.com

Abstract. For large content delivery in store-carry-and-forwarding networks, we study how and when the message delivery delay of the basic epidemic routing (ER) is improved with the introduction of message coding at the source (source coding: SC) and/or at the intermediate relay nodes (network coding: NC). A scenario is examined where a large file is divided into a number of messages and delivered from a stationary source node to a stationary destination in different islands being relayed by cars and ferry boats. In our scenario with limited network resources for exchanging and storing messages, message coding is expected to help ER but not in a straight-forward manner. Through quantitative and controllable evaluation by realistic simulations, we showed that ER+SC+NC is better than ER+SC when the communication link quality is good (e.g., with high wireless signal strength and low car moving velocity), but also that it is not always true as the link quality worsens.

1 Introduction

We study large content deliveries across multiple islands without high-speed infrastructural telecom networks in and between each island [1]. The purpose is to provide low-costs remote access means to televisions (e.g., regular news programs) and e-learning contents (e.g., educational videos). Cars and ferries are considered as mobile relay nodes in a store-carry-and-forward (SCF) routing fashion. This is a typical example of delay and disruption tolerant networking (DTN) where SCF routing is commonly used. Due to the intermediate nodes' high mobility, no end-to-end path is ever guaranteed. Therefore, a multi-copy transfer is often adopted to augment the probability of delivery to the destination. One such method is the Epidemic routing (ER) which is the simplest form of SCF [2]. In ER, based on flooding, messages are replicated before being sent to any node that is newly encountered and that does not have a copy of the messages. ER performs well if there are enough resources in any given node in the network without any global information. However, in real situations this is

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_39

rarely the case due to the nodes' high mobility and poor wireless communication conditions. Another issue of ER here is that the nodes are not dedicated for message relaying and thus, storage provided should be low-cost. These issues lead to a limited number of messages forwarded per contact and to the poor performance of ER.

In this study, we focus on the delivery of messages across a DTN scenario with multiple islands where intermediate nodes with limited resources will transport messages from a source to a destination. In our scenario, one large file is divided into smaller messages that are sent from a stationary source node to a stationary destination in different islands being relayed by cars and ferry boats. Due to the limited resources, ER takes too much time to deliver all the messages (i.e., the entire file).

Random Linear Network Coding (RLNC) in general provides robustness and reliability in relaying messages, and is considered as one of the solutions proposed to ameliorate the performance of ER in such situations. RLNC allows some nodes in the system to first linearly encode messages before replicating or sending them to the next encountered nodes. At the destination, the original messages will be decoded from the encoded messages once a sufficient number of linearly independent messages are received. In this paper, we distinguish message coding at the source (Source Coding: SC) and that at the intermediate relay nodes (Network Coding: NC). Through quantitative and controllable evaluation by realistic simulations. We will show how and when the message delivery delay of ER is improved by introducing SC at the source and/or NC at the intermediate relay nodes. We previously provided a preliminary result that the delivery delay of ER can be improved by SC+NC more than by only SC when the buffer is small and the communication link quality is good [3]. On the other hand, through extensive simulations in this paper, interesting results will be obtained when the communication link quality is worse.

We will use Scenargie Simulator version 2.0 for the simulations. Scenargie is a modeling and simulation framework that allows users to investigate end-to-end system performances under large-scale, complex yet realistic system scenarios [4].

2 Network Coding in Store Carry and Forward Protocol

2.1 Epidemic Routing

ER allows each node to replicate and then send messages to any newly encountered node that does not have a copy of those messages. ER does not perform well in case of limited resources due to flooding that can cause some intermediate nodes to have the same messages. Resulting in a bad use of the limited resources as more transportations will be needed in order to deliver all the original messages to the destination. Signal strength, buffer storages and contact duration between nodes are some of the resources that are limited in our scenario, which lead to a long delivery delay of the F sent by the source.

Note that, in our simulation, an acknowledgement (ACK) scheme is used in all methods, which releases copies of messages that have arrived at the destination but still remain in an intermediate node. However this paper does not investigate and discuss the ACK scheme and its improvement.

2.2 Network Coding

Network Coding paradigm allows the source node or intermediate nodes to not only do SCF but also to combine messages using Galois Field. This results in a more robust and reliable transfer and therefore a better use of the resources which will reduce the time needed for delivering all the original messages. Gauss elimination will be used by the destination to decode the original messages from the encoded messages. NC has a positive effect on ER at a cost of some overheads. Indeed, the encoding coefficients used during the encoding process must be sent with the encoded messages in order to decode. Encoding coefficients are saved as encoding vectors and each vector is inserted in its corresponding header.

We use Galois Field 2 (GF(2)) in message coding. GF is a field with a finite number of elements. A field is a set of elements under four operations with a set of properties governing these operations. Theoretically, even a small GF such as GF(2) offers a better performance than ER based on [5]. A study of NC with larger GF has been considered in [6].

Among different types of network coding, we use RLNC because of its simplicity. With RLNC, the coding coefficients used during the encoding process will be randomly generated from the elements of the chosen Galois Field. Authors in [7] studied and showed some benefits of RLNC.

- Encoding

  We have a large file F which we divided into n small original messages $M_1 ... M_n$. Let $c_1 ... c_n$ be a set of coefficients randomly chosen from a certain GF. An encoded message X_i will be defined as below.

$$X_i = \sum_{i \in N} c_i * M_i \tag{1}$$

  - c_i: coding coefficient of X_i
  - M_i: original Messages
  - N: Natural numbers
  - The arithmetic operations are defined by the GF chosen during the generation of the coefficients c_i.

  In the intermediate nodes we re-encode messages that has already been encoded using the same method. We can re-encode messages as Y_i, which can be expressed in terms of the original messages M_i as below.

$$Y_i = \sum_{i \in N} g_i * M_i \tag{2}$$

The coding coefficients c_i (or g_i) will be saved as encoding vectors C_i. They will be put in the header file and will be sent to the next nodes along with their respective encoded messages X_i (or Y_i). Here C_i is in the next form.

$$C_i = \begin{bmatrix} c_1 \; c_2 \; c_3 \; \ldots \; c_n \end{bmatrix}$$

- Decoding
  At the destination, the goal is to recover the original messages $M_1...M_n$ from the received encoded messages. The destination will use the encoding vectors C_i and encoded messages Y_i to recover the original messages M_i using Gauss elimination and subsequently retrieve the original file F. The decoding process in the destination leads to a solving of equations with n unknowns as below.

$$\begin{bmatrix} c_{11} \; c_{12} \; c_{13} \; \cdots \; c_{1n} \\ c_{21} \; c_{22} \; c_{23} \; \cdots \; c_{2n} \\ \cdots\cdots \\ c_{d1} \; c_{d2} \; c_{d3} \; \cdots \; c_{dn} \end{bmatrix} = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{d1} \end{bmatrix}$$

  – n: number of original messages
  – d: number of messages received by the destination
  In order to solve that kind of equation at the destination, we must, at the very least, receive n linearly independent messages. If the messages are linearly dependent, we will not be able to solve the system of equations and retrieve all the original messages. Thus, the file F will not yet be retrievable and we must wait for more messages, resulting in a long delivery delay.

3 Implementations

To compare the performances of the three methods: ER, ER+SC, and ER+SC+NC, we implement the different types of nodes as follows. The source node only sends messages. It divides a large file into n messages, encodes them to $n+k$ messages (as SC), and distributes them to encountered intermediate nodes. The destination node only receives messages. It stores only a new message (i.e., linearly independent to already received ones), and decode the original file when n independent messages are collected. All other nodes are intermediate nodes with two types. One is an encoding node and the other is a non-encoding node. In ER+SC, all intermediate nodes are non-encoding nodes. On the other hand, in ER+SC+NC, cars are encoding nodes but ferry stations and ferry boats are non-encoding nodes.

3.1 ER with Source Coding (ER+SC)

At the source, when the file F is received from the application layer, it will be divided into n original messages. From these messages, α number of encoded messages will be created using (1). The source will then contain $n + \alpha$ messages. To make an efficient use of SC, the newly encoded α messages must be linearly independent. This is to ensure that all encoded messages contain different information. To check the linear dependency of these messages, the rank of the encoding matrix will be calculated and must be equal to α. When meeting the intermediate nodes, the source will then randomly select and send messages from these $n + \alpha$ messages.

At the intermediate nodes, messages will only be stored, carried and forwarded to other nodes. The intermediate nodes will only replicate the messages they have before sending them to a newly discovered node that do not have any copy of the messages they contain.

At the destination, the goal is to recover the original messages from the delivered encoded messages. The main point is that the destination needs to at least receive n linearly independent messages in order to decode. The destination contains a decoding matrix. After a message is received, its encoding vectors are placed into that decoding matrix and the rank of the decoding matrix will be computed. If the rank of the decoding matrix augments after a message reception, that message contains new information for the destination and will be stored. However, if the rank does not augment after a message reception, that message will be considered as linearly dependent to the messages already within the destination. It will then be removed as it does not help in the decoding process. When the rank of the decoding matrix is full (rank $= n$), decoding process is done using Gauss elimination.

3.2 ER with Source Coding and Simple NC (ER+SC+NC)

At the source, the process is exactly the same as in ER+SC (described in Sect. 3.1). The only difference is that with ER+SC+NC the intermediate nodes are allowed to encode or re-encode messages.

At the intermediate nodes, let us suppose a node contains s number of messages and it receives a message from another node. In one hand, if the buffer storage is not yet full, the message will be stored and the number of messages within that node will be $s+1$. In the other hand, if the buffer storage is full, encoding will happen. At the time of encoding, the messages will be randomly encoded to one another using elements from GF(2). The encoding process is described in Sect. 2.2. s newly re-encoded messages will then be created. The older messages will be deleted and replaced with the newly re-encoded messages. Allowing the intermediate nodes to re-encode messages not only makes each message contain more information but also allows the creation of many different messages in ER+SC+NC. Indeed, there are much more variation of messages circulating within the intermediate nodes with ER+SC+NC compared to ER or ER+SC. With ER+SC+NC, the intermediate nodes likely do not have the same messages. Consequently, less transportations are needed in order to recover the original messages at the destination and to retrieve the original file F. Therefore, a better performance with ER+SC+NC can be expected.

At the destination, the behaviour is the same as in SC. As we know, the original messages are retrieved when enough linearly independent messages arrive at the destination. The challenge here is how to deliver those linearly independent messages as fast as possible. With NC, the presence of linearly dependent messages is delaying this delivery. These linearly dependent messages greatly affect the performance of ER+SC+NC. In some cases, there are so many linearly dependent messages that even ER+SC performs better than ER+SC+NC. In this study, we are going to see the advantages and drawback of NC in different cases.

4 Simulation

4.1 Simulation Settings

Figure 1 represents our simulation scenario containing two separate islands connected to each other by a ferry travelling in and out of the two islands. The role of the ferry is to deliver messages from one station in an island (station A) to another station in another island (station B). The first island (island A) includes one static source node, one static station node (station A) and 9 cars that are responsible for delivering messages from the source to station A. The second island (Island B) is a smaller island containing 1 static destination node, 1 station node (station B) and 5 cars. Note that in our simulations, when arriving at an intersection, each car will randomly select a road to go on. Moreover, a car does not have a constant velocity but instead is assigned a pre-defined range of velocity. When arriving at an intersection, each car will randomly pick a velocity within that range and keep it until arrival at the next road intersection.

A large file (5 MB) divided into 100 small (50 KB) original messages is delivered from the source in island A to the destination in island B. Since 5 MB may not be enough for a long video file, a larger content file is assumed to be divided into sub-files (of 5 MB size) and those sub-files are delivered separately and independently. At the source 20(α) redundant messages are sent with the original

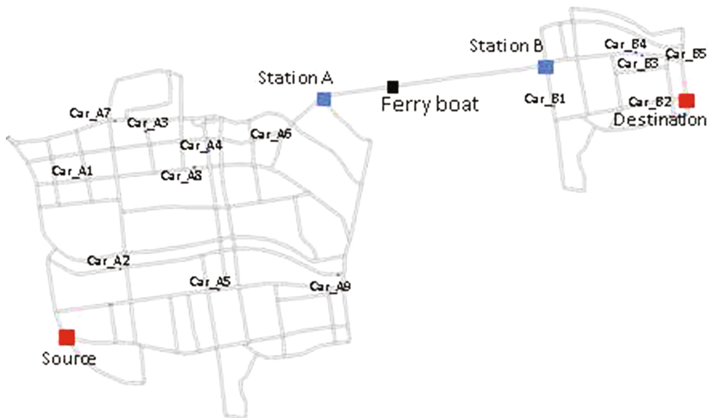
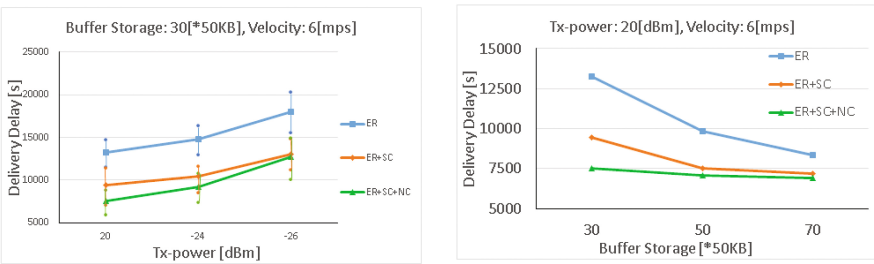


Fig. 1. Simulation scenario

messages with ER+SC and ER+SC+NC. Since we focus on the delivery delay time of the entire file (i.e., the time when the entire file is retrieved/decoded at the destination), we set the message TTL infinite. The buffer storage of the static nodes are unlimited but the cars' are limited and varied (30, 50, 70 messages) to simulate real situations. The Wi-Fi module used during the simulations is 802.11g with a frequency of 2.4 [GHz]. The signal strength of the Wi-fi module is varied from strong (20 [dbm]) to very weak (−26 [dbm]) in order to simulate a wide range of real situations in which diverse link quality conditions will happen.

4.2 Simulation Results and Consideration

Figure 2a represents the average delivery delay performances with different signal strengths. The center value indicates the delivery delay time averaged over 30 different mobility seeds, and the error bar indicates the range between 10-percentile and 90-percentile. We can see that the lower the signal strength is, the longer the delivery delay for all methods, while the absolute value of delay considerably varies according to mobility. With a strong signal, the quality of the communication link is good, e.g., without packet losses. With a good link quality, the nodes can exchange messages effectively, and thus a high number of messages can be exchanged during a contact between two moving nodes. In contrast, with a weak signal, less messages will then be exchanged per contact compared to the case with strong signal. Therefore, it will take more time to deliver the necessary number of messages from the source to the destination.



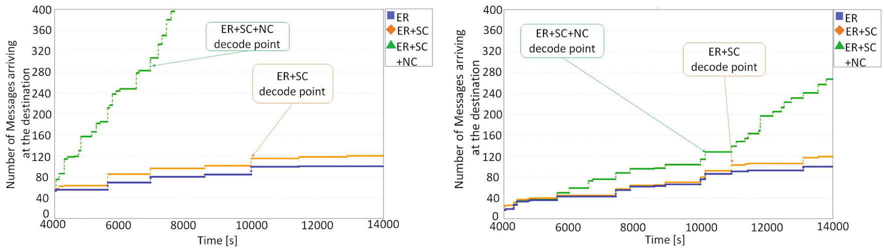
(a) Delivery delay with different signal strength (b) Delivery delay with different buffer sizes

Fig. 2. Average delay performances with different signal strengths and buffer sizes

Regardless of the signal strength, both ER+SC and ER+SC+NC clearly outperform ER due to the inefficiency of ER and the benefit of message coding explained in Sect. 2. Therefore, from here on, we focus on the comparison of ER+SC and ER+SC+NC. With strong signal (tx-power is 20), the performance of ER+SC+NC is clearly better than that of ER+SC. This is explained as follows. Note that drop-oldest algorithm is used as a mean to manage the buffer storage when it is full. In addition, encoding or re-encoding in the intermediate nodes only happens when the buffer storage is full in NC. A high signal strength, i.e., a good link quality, can ensure many messages be exchanged in each contact duration. In such situations, the buffer storages of the intermediate nodes become full fast, and a lot of re-coding are done in the intermediate nodes and messages from the source are well mixed in the intermediate nodes in ER+SC+NC. This means that ER+SC+NC fully leverages the capacity of good links to deliver as many different encoded messages as possible to the destination.

In contrast, with ER+SC, when its buffer storage is full, an intermediate node has to drop a message in order to be able to receive another one from a different node. This means that a node has to lose one information in order to gain another one, therefore the capacity of good links is not efficiently utilized in order to convey all the needed messages to the destination.

Figure 3a shows a case example of the total number of (different) messages delivered to the destination through time with a strong signal. The decoding time (i.e., for retrieving the entire file) of each method is also depicted. We can see that a large number of messages are created with NC and received by the destination in ER+SC+NC. Although a large number of messages may include many linearly dependent messages, the n linearly independent messages needed for decoding will arrive at the destination faster than ER+SC. In this case of ER+SC+NC, around 280 messages are needed for the destination to collect 100 independent messages.



(a) With a strong signal

(b) With a weak signal

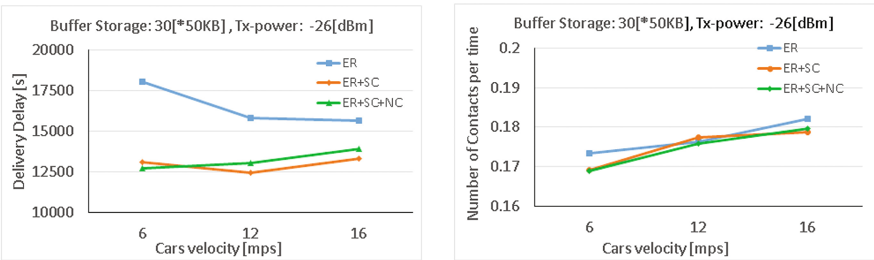
Fig. 3. Increase of the number of messages delivered to the destination by ER, ER+SC and ER+SC+NC through time with different signal strength

On the other hand, with a weak signal (tx-power is -26), the performance of ER+SC+NC and ER+SC are almost the same. This is because a low signal strength, i.e., a bad link quality, cannot ensure many messages be exchanged in each contact duration. In such situations, the buffer storages of the intermediate nodes will take a longer time to be full. Due to the fact that re-encoding only happens in the intermediate nodes when the storage is full, there are less re-encoding processes in the intermediate nodes compared to the case with strong signal. With less re-encoding in the intermediate nodes, the behaviour in ER+SC+NC and ER+SC are similar in an early stage. Figure 3b shows an example of the total number of messages delivered to the destination through time with a weak signal. Focusing on ER+SC+NC in green, we can see that a smaller number of new messages are received by the destination in each contact. Moreover, a smaller number of valid contacts to the destination happens, compared to the case with strong signal (Fig. 3a).

Figure 2b shows the average delivery delay performances of ER, ER+SC and ER+SC+NC with different buffer sizes when the link quality is good. We can see that all performances are better when the buffer size of the intermediate nodes augments. That is because with a large buffer storage, every node can bring a lot of messages at once. Therefore, the messages will arrive very fast at the destination.

The performance improvement of ER+SC+NC with a large buffer is small. This suggests that the increase of buffer size (from 30 to 50 and 70 messages) does not contribute much to increase the number of linearly independent messages in the buffer. In contrast, the performance improvement of ER+SC with a large buffer is large. This is because less messages are dropped in ER+SC compared to the case with small buffer. Hence the similar performances for ER+SC and ER+SC+NC.

Figure 4a shows the average delivery delay performances with different car velocities and a weak signal (tx-power: −26). With a low car velocity of 6[mps], the performance of ER+SC+NC and ER+SC are almost the same as previously explained.



(a) Average delivery delay performances

(b) Average number of contacts between the nodes per time

Fig. 4. Average delay performances and number of contacts with different velocities

When the intermediate nodes move with a higher velocity of 12[mps], while the performance of ER and ER+SC becomes better, that of ER+SC+NC clearly worsens. On one hand, a higher velocity can generally result in an increase of the number of contacts between two nodes per time, which may shorten the time to distribute messages. Figure 4b shows this tendency and also shows that the number of contacts per time does not essentially differ in all methods. On the other hand, a higher velocity worsens the link quality between encountered nodes, i.e., a decrease of the number of messages that can be exchanged in one contact. While the former positive impact contributes to the better performance of ER and ER+SC, the latter negative impact degrades the performance of ER+SC+NC. This trade-off makes things more complicated. In fact, when the intermediate nodes move even faster with a velocity of 16[mps], the performance of all methods becomes worse because the negative impact of high speed movement is stronger.

The poor performance of ER+SC+NC can be explained as follows. A low signal strength and a high velocity result in a very bad link quality with many packet losses, and thus only a small number of messages are exchanged in each contact duration. In particular, since ER+SC+NC tries to exchange messages more aggressively, we can verify that the packet loss rate of ER+SC+NC is larger than that of ER+SC. This means an ineffective NC leading to linearly dependent messages. Furthermore, the number of messages received by the destination is small in each contact. Therefore, more number of contacts are needed in order to deliver the n linearly independent messages with ER+SC+NC. Figure 5 shows a case example of the total number of messages delivered to the destination through time with high velocity and weak signal. We can see that a smaller number of new messages are received by the destination in each contact while a larger number of valid contacts to the destination happens, compared to the case with low velocity and weak signal (Fig. 3b).

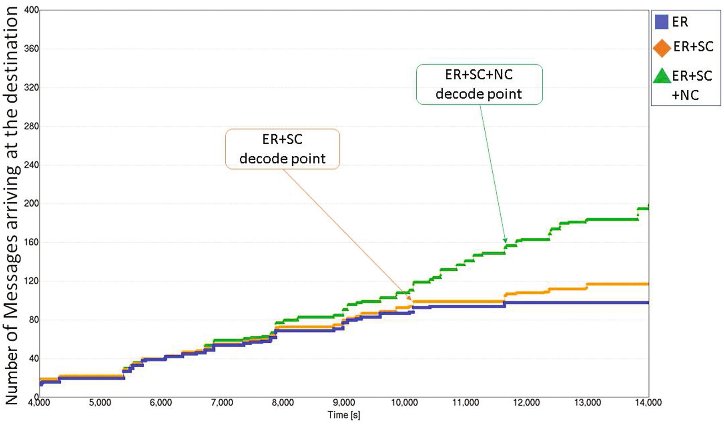


Fig. 5. Increase of the number of messages delivered to the destination by ER, ER+SC and ER+SC+NC with a car velocity of 12 [m/s]

5 Conclusion

In this paper, we studied how and when the message delivery delay of the basic ER is improved with the introduction of random linear coding at the source (SC) and/or at the intermediate relay nodes (NC). A scenario is examined where a large file is divided into a number of messages and delivered from a stationary source node to a stationary destination in different islands being relayed by cars and ferry boats. We evaluated the delivery delay time of the entire file (i.e., the time when the entire file is retrieved/decoded at the destination) by ER, ER+SC, and ER+SC+NC, through realistic network simulations.

The simulation results suggested the following. (i) SC and NC can improve the delivery delay performance of ER even if applied in a simple way and with a small GF(2). With a good quality link but without a large buffer, ER+SC+NC can perform significantly well compared to ER+SC. (ii) In ER+SC+NC, the performance decreases when the number of messages exchanged between nodes in contacts is not enough, e.g., with a low wireless signal strength and a high car moving velocity. In summary, ER+SC+NC is not always better than ER+SC with a bad quality link. This suggests a need for adaptive NC schemes according to link quality conditions.

Acknowledgement. This work is partly supported by JSPS KAKENHI (16K00130).

References

1. Agussalim, Tsuru, M.: Spray router with node location-dependent remaining TTL-based message scheduling in DTNs. IPSJ J. Inf. Process. **24**(4), 647–659 (2016)
2. Vahdat, S., Becker, D.: Epidemic routing for partially connected Ad Hoc networks, Duke University, Durham, NC, Technical report, April 2000
3. Tsuru, M., Takai, M., Kaneda, S., Agussalim, Tsiory, R.A.: Towards practical store-carry-forward networking: examples and issues. IEICE Trans. Commun. **E100-B**(1), 2–10 (2017)
4. Scenargie HP. https://www.spacetime-eng.com/en/. Accessed 02 Apr 2016
5. Matsuda, T., Noguchi, T., Takine, T.: Survey of network coding and its applications. IEICE Trans. Commun. **E94-B**(3), 698–717 (2011)
6. Zhang, Q., Jin, Z., Zhang, Z., Shu, Y.: Network Coding for Applications in the Delay Tolerant Network (DTN)
7. Ho, T., Koetter, R., Medard, M., Karger, D., Effros, M.: The benefits of coding over routing in a randomized setting. In: Proceedings of IEEE International Symposium on Information Theory, Yokohama, Japan, June 2003. doi:10.1109/ISIT.2003.1228459

Fairness Aspects of Secondary Radio Spectrum Access

Mario Köppen[1(✉)] and Masato Tsuru[2]

[1] Graduate School of Creative Informatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
mkoeppen@ieee.org

[2] Department of Computer Science and Electronics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
tsuru@cse.kyutech.ac.jp

Abstract. We consider the allocation of secondary radio spectrum under fairness constraints. Two neutrality aspects apply here: the provider allocates independent of the packet utility of the secondary user, and: all secondary users are given the same unbiased opportunity to access the available spectrum. The first means that the standard fairness models do not apply directly, the second that we are more interested in the chance of a secondary user to place a number of packets together, i.e. a numerical value from (0, 1]. This leads us to a formulation of the underlying combinatorial optimization problem as an "multi-user knapsack proble" (MUK). For short, it means that several users put their items into the same knapsack and the fairness is related to the share of packed items for each user. We discuss how common ideas from other fairness models can be applied to this new problem, esp. the concept of fairness as convex mean maximization. A number of example cases are studied to demonstrate feasibility of the proposed approach.

1 Introduction

When the Jovian-bound Galileo spacecraft failed to fully open the high-gain antenna on earth-flyby in December 1990, before gaining full momentum to start the final journey to Jupiter, all communication means had to resort to the low-gain antenna of much lower bandwidth that was used to maintain spacecraft communication during the gravitational push phase in the inner solar system. It wasn't enough bandwidth to stream valuable scientific data from all instruments, including an 800×800 pixel solid-state camera, spectrometer, radiometer, dust and energy particle detector, magnetometer, heavy ion counter and the plasma subsystem back to Earth. The scientific teams behind each of these instruments had to face this permanent congestion of the communication channels for the rest of the planetary mission. Further assessing the situation, it poses a number of efficiency and fairness-related question that will be, in an abstract way, further discussed here. For example, while in general the enabled spectrum access will be

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_40

organized in an unbiased way, in some cases we find the need of some "primary traffic"—such cases can be the transmission of the high resolution images, as they are to some not to underrate degree important for the public appeal of the whole mission, or the timely transmission of important scientific experiment results.

However, there is no need to look as far as Jupiter to find the same problem: it has been reported many times that in fact, also subscribed communication channels in terrestrial wireless infrastructures are for larger parts left unused (see [1] for a survey on the general problem, [10] for a more recent specific story). Also here, some means to allocate unused radio spectrum to secondary users will in total leverage the efficiency of the whole spectrum employment. Now we face the situation that there are a number of user packets of different size that must fit into whatever is available for such a secondary access, with the accompanying question how to implement a fair allocation scheme. Starting point will be to consider it as a variation of the famous knapsack problem. There, a number of items, each item having a weight and a value, have to be put into a knapsack of maximum total weight such that the value in the knapsack is maximized. In the same sense we want to put "items" (packets of given transmission size) into a "knapsack" (the available radio spectrum within some time frames). However, compared to the classical problem, two aspects have to be changed: at first, there is no value in that sense that it could directly serve as a cue to prefer the one item to be packed against the other. As for the above example of the transmission of scientific data from various instruments that naturally differ in size, from the receiver point of view (who essentially doesn't know what will arrive at all) there is no reason to make any difference among the kinds of scientific data with regard to allocated bandwidth, after the "primary tasks" have been served. Moreover, for requesting a valuation of items it would need some common standards of evaluation among the affected parties, which could tempt someone to report larger values than they really are in order to gain an advantage in the spectrum access—in other words, such a scheme could hardly be strategy-proof.

At second, it differs from the classical knapsack in the way that the sharing is represented per user, and not as a single value to be maximized. Assuming that each user who applies for secondary spectrum access puts in a total less than the total available, or equal, it also means that the direct benefit has to be measured on a per-user base and not as a total (which would just mean to fill the knapsack as much as possible, without even bothering if indeed each user can put at least one item into the knapsack).

In summary, we come up with a problem statement of a "multi-user knapsack" problem where the measure is the per-user share of items to be put into a single knapsack together. It will be introduced as a base class of fairness models, which is not yet much investigated. In the next subsection, some related work on shared knapsack problems will be discussed, while then Sect. 3 gives a formal definition of the problem. Section 4 will summarize some basic insights into the problem. Then, from Sect. 5 on, we will study a generic approach to fairness based on convex mean maximization that can be applied here, and it will be exemplified for a number of cases in Sect. 6. The paper concludes with a summary of our findings.

2 Related Work

There have been a number of studies on a "knapsack sharing problem" with a strong focus on the combinatorial optimization aspects. The publication [2] was probably the first to introduce and study this kind of problems. In this form, it requires that each item of a user put into the (same, shared) knapsack is associated with some profit, or value, of putting it in. Then, an optimization problem is linked to maximality of profits. For example, in [4] of Fujimoto and Yamada—who provided numerous other contributions to this discipline—a case is studied where there are also common items to select (i.e. items that have the same value for all users). An exact algorithm is provided that maximizes the least profit. Despite the common use of a maximin criterion, the problem hasn't been studied as a problem of fairness among the users until the recent [9], also paying attention now to issues like proportional fairness and the fairness-efficiency trade-off and considering the study case here as a subset sum problem. Main difference is that in the present approach, such a profit or value will not be taken into account, i.e. the only quality that matters in filling the knapsack is the volume, and the share of items packed. To see that the present analysis is not restricted to the case of secondary spectrum access, but also applies to other logistic components of communication networks like queues and buffers, see [11]. There, within the domain of delay-tolerant networking, the non-interruption of transmissions by packet drops within a queue is modeled as a knapsack problem as well, while not directly considering any fairness aspect.

3 The Multi-user Knapsack Problem

We formalize the problem in terms of a knapsack problem. Several users are putting items into the knapsack in a competing fashion. The items are assumed to be "shapeless"—so there is no issue of how to put them densely enough together into the knapsack (it isn't an issue for the classical knapsack problem as well...).

Assume that a knapsack has capacity (unite-less volume) V. Among n users, each user i has n_i items j of (not necessarily same) volume v_{ij} to pack into that knapsack. However, for each user i the total volume of items $V_i = \sum_{j=1}^{n_i} v_{ij}$ does not exceed the knapsack volume. For some specific packing of items into the knapsack, $\delta_{ij} = 1$ indicates that item j of user i is in the knapsack, and by $\delta_{ij} = 0$ if this is not the case. So, in order to not overload the knapsack it must hold

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\delta_{ij}v_{ij} \le V \tag{1}$$

where, as a result of the packing coordination, each user has a "payload" of

$$p_i = \sum_{j=1}^{n_i}\delta_{ij}v_{ij} \tag{2}$$

in the knapsack. This payload value will be seen in relation to her or his total item volume as an expression of "user satisfaction" and it can be represented as vector r with n components r_i as

$$r_i = \frac{\sum_{j=1}^{n_i}\delta_{ij}v_{ij}}{V_i} \tag{3}$$

Then, the goal of the Multiuser-Knapsack Problem (MUK) is to find an optimal user satisfaction in terms of allocation vector r.

A simple numerical example: there are 3 users and a knapsack of volume 1, user 1 has items of volumes $\{0.5, 0.5\}$, user 2 3 items $\{0.2, 0.4, 0.4\}$ and user 3 of $\{0.3, 0.3, 0.4\}$. We put one item of 0.5 of user 1 in the knapsack, an 0.2 item of user 2 and a 0.3 item of user 3. The knapsack is full. For the user loads this gives 50% for user 1, 20% for user 2 and 30% for user 3. Can we do better?

The example also allows to demonstrate a fairness issue. Assume the objective is to fill the knapsack as much as possible and we have only users 1 and 3. We can reach a total of 1 easily, like taking all items of one user, but we cannot reach a total of 1 from mixing both users. The most we can get is 0.9 (volume 0.5 item of user 1 and volume 0.4 item of user 3). Thus, the only way to maximize knapsack filling is by excluding one user.

4 Tractability

The set cover problem asks if from a set of numbers a subset can be found that is exactly half of the total sum of all numbers. It is an NP-complete problem. It can be the base for showing that a related MUK decision problem is NP-complete as well.

We can easily devise an algorithm for item selection that resembles a discrete version of Bottleneck Flow Control [5]. All users sort their items by size, from smallest to largest. Then, within each such group of k-th smallest user items, the items are sorted by size in non-decreasing order as well. Now we start to fill the knapsack, first the smallest item from the group of smallest items, then second-smallest etc. If the knapsack is not yet full, we continue with the smallest item of the group of user's second-smallest items, then second smallest etc. Either we can also fully place all items of that group into the knapsack as well, or at one place the knapsack is already full. In former case, we continue with the group of third-smallest items etc. We will surely reach a point where no further item can be put in.

Now the intriguing point is that such an allocation is not giving (Pareto) efficient solutions—which means that we can find another solution where for at least one user more items are placed in the knapsack without reducing the number of items of any other user.

As an example that we will also use in the following section, consider the following items for 3 users who want to fill a knapsack of volume 10.

User	Item volumes
User 1	1 2 3 4
User 2	1 1 2 2
User 3	3 3 4

The above procedure will lead to the following filling of the knapsack:

User	Item volumes
User 1	1 2
User 2	1 1
User 3	3

which gives a total of 8. In the following, we will write such allocations more compact as ((1 2) (1 1) (3)). It appears that in this case, user 1 has 3 of 10 total in the knapsack, i.e. 30%, user 2 has 2 of 6 (33.3%) and user 3 3 of 10, also 30% like user 1. So the payload of users is represented by the vector $(0.3, 0.33, 0.3)$. Can we do better?

Of course. It's because the knapsack is not yet full, 2 volume units are still vacant. For example, we can increase user 2's payload by adding one more volume 2 item. The knapsack is full, and user 2 now has 4 of 6, 50% in the knapsack. And we can achieve this without reducing any other user's payload.

5 Fairness by Convex Mean Maximization

Fairness is an inevitable aspect of joint resource usage. For a complete unconstrained set up, an additive resource (additive with regard to its replenishment, and not with regard to its distribution) has to be distributed in equal amounts. The idea of equity is then how to realize "equal amounts" in constrained and more complex scenarios. All existing fairness models can be linked to means in the one or other way. To explain this, we assume that there is a method to compare two different allocations by some binary relation. For some pairs there can be indifference, i.e. that for two allocations A and B, neither A is preferred to B nor B preferred to A. For such cases (for additive resources) the preference relation then should work such that any feasible convex combination is (strictly) preferred to both, A and B. Why do we assume this? Reason is that once a decision has to be made for either A or B, and assuming sufficiently risk-averse users, there is always a number of them that are worse off in A, and other users who are worse off in B. In a convex combination, both users will simply win.

For the basic case that there is a qualifying function F mapping allocations into the real domain, preferred allocations to larger values, indifference is expressed as $F(A) = F(B)$ and our basic requirement as $F(wA + (1-w)B) > F(A)$ for $w \in (0,1)$. It means that F is a convex function.

We can make some further requirements that appear reasonable. At first, the way how users are sorted shouldn't matter. Thus, F should be symmetric. At second, we expect a (weakly) monotone-increasing function: whenever there is more to distribute, no user should have a disadvantage. Third and last point is that F should be bounded, even more, bounded with some sense of the range of the mapped data, i.e. sense for their maximum and minimum. It doesn't mean to be exactly bounded between minimum and maximum, but a rescaling of F for example might always serve for this purpose.

Putting all together, these are all properties fulfilled by convex means, essentially these are the requirements for a mapping to be considered as a mean: symmetry, monotonicity, to be bounded between min and max.

Following this viewpoint, a fairness model can be based on the maximization of some mean. Before giving a few more details in the following, and without blowing up this theory to heavily here simply for space reasons, what is known is: maximization of the geometric means equals maximization of the product of the allocations, known for long as proportional fairness [6,8]. For power means with negative p (e.g. harmonic mean with $p = -1$) this represents α-fairness [7], for $p \to \infty$ it approximates lexicographic maximin. Moreover, once A and B approximate the maximum they will come infinitesimal close to each other, and the general preference of convex combination changes into a tangential form, like $\sum(B_i - A_i)/A_i \leq 0$ as a binary relation derived from the maximization of the geometric mean.

For the present case of MUK, we will consider application of the same fairness model of convex mean optimization. In difference to other fairness models a special feature is that the allocations are all from $(0,1]$, since they express the shares r_i of items each user can put into the knapsack. This allows to apply means that generally can't be used for general real values, like entropy or T-norms. It will be further discussed in the next section.

6 Study Cases

We saw before an example for an inefficient filling procedure of the knapsack that nevertheless appeared to be fair. We can compare this solution with solutions maximizing some mean and using above examples. Since in this example we have 11 items to pack, there are 2^{11} possible ways to select items for the knapsack, independent to which user they belong. However, many of them exceed the knapsack volume. Moreover, we only consider itineraries where each user has at least one item in the knapsack. At the end, there are 85 remaining.

6.1 Pythagorean Means

Given n values x_i, the arithmetic mean is defined as

$$M_a = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

The result maximizing the arithmetic mean is the allocation

((1) (1 1 2 2) (3))

with a payload of (0.1 1.0 0.3). Note that the maxima here has been found by exhaustive search over all 85 feasible knapsack packings where each user puts in at least one item. It can be seen that user 2 is able to put all items in the knapsack. This is possible since user 2 generally has the smallest total of item volumes. However, it appears as an aspect of unfairness against the other users, who can only put a small fraction of their items. We find this unfairness in coincidence with the arithmetic mean not being strictly convex.

Given n (non-negative) values x_i, the geometric mean is defined as

$$M_g = \left[\prod_{i=1}^{n} x_i\right]^{1/n} \tag{5}$$

Here we have several maximizing itineraries, all for the same value of maximum geometric mean of $\sqrt[3]{0.06}$. Here and in the following tables, one solution per row is shown.

Payload (0.3 0.5 0.4) for:	((1 2) (1 2) (4))
	((3) (1 2) (4))
Payload (0.4 0.5 0.3) for:	((4) (1 2) (3))
	((1 3) (1 2) (3))
Payload (0.3 0.6$\bar{6}$ 0.3) for:	((3) (2 2) (3))
	((1 2) (1 1 2) (3))
	((1 2) (2 2) (3))
	((3) (1 1 2) (3))

Note that none of them is related to the allocation that maximizes the arithmetic mean. Some entries appear multiple times due to alternatives way to put the same volume in items of some user. For example, in the third solution, user 1 can either put the item of volume 3, or two items of volumes 1 and 2.

As further "classical" Pythagorean means there is the harmonic mean, defined for positive numbers x_i as

$$M_h = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{-1}\right]^{-1} \quad (6)$$

Here the maximal allocations are:

Payload (0.3 0.5 0.4) for:	((1 2) (1 2) (4)) ((3) (1 2) (4))
Payload (0.4 0.5 0.3) for:	((4) (1 2) (3)) ((1 3) (1 2) (3))

appearing the same as for the geometric mean except the 3rd solution.

6.2 Power Means

The Pythagorean means are special cases of the so-called power means (also sometimes called Hölder means). For some value $q \neq 0$ the q-th power means is defined

$$P_q = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{q}\right]^{1/q} \quad (7)$$

and for $q = 1$ it equals the arithmetic mean and $q = -1$ the harmonic mean. It can also be shown that for $q \to 0$ the power means approximates the geometric mean. Moreover, for $q \to -\infty$ the power means approximates the minimum of the p_i while for $q \to \infty$ the maximum. Also remember that for $q_1 > q_2$ it is $P_{q_1} > P_{q_2}$. Here we show itineraries for other values of q than $-1, 0$ or 1. For $q = -2$ we obtain the same result as for $q = -1$ (i.e. the harmonic mean). For $q = -5$ the situation changes:

Payload (0.4 0.3$\bar{3}$ 0.4) for:	((4) (2) (4)) ((1 3) (1 1) (4)) ((4) (1 1) (4)) ((1 3) (2) (4))

Further decrease of q continues to give the same maximizing solution. Actually, it appears as the (only) solution that has the largest minimum payload among the users. In the following we will call this solution as the **maxmin** solution.

On the other hand, if we increase q beyond 1, we always get the same solution as for the arithmetic mean. It also appears as the solution with the maximum payload for at least one user of 1.0. Simply, more than 1.0 is not possible.

6.3 Entropy

As a complete different approach, we consider the Shannon entropy of the payloads, imaging an underlying interpretation of the payload as the probability for the user to have an item in the knapsack. Formally we want to maximize

$$E_s = \sum_{i=1}^{n} -x_i \log x_i \tag{8}$$

Maximizing this operator we again get the minmax solution as the one maximizing the Shannon entropy, i.e. a payload of $(0.4, 0.\bar{3}, 0.4)$.

6.4 Mean Indizes

Mean indizes are a little bit different from means in the sense that they are aggregating information on a per-user-basis. Thus, it are values between 0 and 1 and express the idea that index value a means that a percentage a of users shares a common advantage or disadvantage. The famous Jain index however is of limited value since it is maximized as soon as all values are equal. In our example, this can be achieved by an allocation ((3) (1 2) (3)) for example, with a Jain index of 1. But this is far from being efficient. Much better suitable and known is the Gini coefficient as such an index.

The Gini coefficient expresses the balance of a distribution. It is 0 for a perfectly equal distribution. A convenient way of its calculation is:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 M_a} \tag{9}$$

(also sometimes called "relative mean difference"). Here is the solution that minimizes the Gini coefficient.

Payload (0.3 $0.3\bar{3}$ 0.3) for:	((3) (1 1) (3))
	((1 2) (2) (3))
	((3) (2) (3))
	((1 2) (1 1) (3))

We note that the solution on the last line is already known to us. It is the solution that we have got from packing the knapsack in the manner described in Sect. 4, where smallest items were packed first. It appears as a solution where in fact the distribution is most balanced.

6.5 Symmetric Means

A number of means is defined by unfolding over all choices of $k \leq n$ elements from the available values, for example over all pairs, triples etc.

6.5.1 Symmetric Polynomial Means

The symmetric polynomial means has a single parameter k (the number of elements chosen from n). We define the k-th **symmetric polynomial** as

$$\sigma_k = \sum_{1 \le i_1 \le i_2 \le \cdots \le i_k \le n} x_{i_1} x_{i_2} \ldots x_{i_k} \tag{10}$$

and then the **symmetric polynomial means** as:

$$S_k = \left[\frac{\sigma_k}{\binom{n}{k}} \right]^{1/k} \tag{11}$$

Another variant is the **symmetric polynomial ratio means**, given as

$$S_k^r = \frac{\sigma_k / \binom{n}{k}}{\sigma_{k-1} / \binom{n}{k-1}} \tag{12}$$

For the case of our example problem, there are not so many choices for k since for $n = 3$ the symmetric polynomial means with $k = 1$ is the arithmetic means and with $k = 3$ the geometric means. Note that this is generally the case that for increasing k the symmetric polynomial means moves from arithmetic means to geometric means. For $k = 2$ the symmetric polynomial means is maximized with payloads $(0.3\ 0.\bar{6}\ 0.3)$, a vector that we already know as solution to maximizing the geometric means.

6.5.2 Gini and Lehmer Means

The **Gini mean** is generally defined for pairs of exponents p and q as:

$$G_{p,q} = \left[\frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} x_i^q} \right]^{1/p-q} \tag{13}$$

and the **Lehmer mean** is most easily defined as a special case where $q = p - 1$:

$$G_{p,q} = \frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} x_i^{p-1}} \tag{14}$$

The Lehmer mean evolves similar to the power mean. For $p = 1$ it is equal to the arithmetic mean, for $p = 0$ it is the harmonic mean. For $p \to \infty$ it approximates the maximum, and for $p \to -\infty$ the minimum. For $p = 1/2$ there is the main difference to the power mean: for two values it is the same as the geometric mean, however, for 3 and more values it defines a different mean.

In case of our example, we do not get anything new with regard to selection from all possible packings. For example, the Lehmer mean with $p > 1$ is maximized by the same solution that maximizes the arithmetic mean. For $p = 1/2$ it is the same as for the geometric mean, even while being a different mean. Another example: the Gini mean for exponents $1/2$ and $-1/2$ is maximized by a solution with payloads $(0.3\ 0.\bar{6}\ 0.3)$.

7 Summary and Conclusions

In order to study fairness aspects of secondary spectrum access, as proxy for a number of other allocation problems in communication networks, the multi-user knapsack (MUK) problem was stated. In contrary to other knapsack sharing problems, profits of stored items are not assumed, and the allocation's effectiveness is measured in terms of the share per user that can be put into the shared knapsack. Maximization of some convex means (see [3] for an extensive discussion of many other means) has been studied as the generic method to solicit fair allocations. We could find that allocations per maximization of non-convex means show characteristics of unfairness, while other values like fairness indizes lead to non-efficient solutions. Thus, there is some preference to focus indeed on convex means.

Acknowledgement. The work presented in this paper was partially supported by JSPS KAKENHI (16K00130).

References

1. Akyildiz, I.F., Lee, W.Y., Vuran, M.C., Mohanty, S.: Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)
2. Brown, J.R.: The Knapsack sharing problem. Oper. Res. **27**(2), 341–355 (1979)
3. Bullen, P.S.: Handbook of Means and Their Inequalities, Mathematics and Its Applications, vol. 560. Kluwer Academic Publishers, Dordrecht (2010)
4. Fujimoto, M., Yamada, T.: An exact algorithm for the Knapsack sharing problem with common items. Eur. J. Oper. Res. **171**(2), 693–707 (2006)
5. Jaffe, J.: Bottleneck flow control. IEEE Trans. Commun. COM-29 **1**(7), 954–962 (1981)
6. Kelly, F.: Charging and rate control for elastic traffic. Eur. Trans. Telecommun. **8**(1), 33–37 (1997)
7. Mo, J., Walr, J.: Fair end-to-end window-based congestion control. IEEE/ACM Trans. Netw. **8**(5),556–567 (2000)
8. Nash, Jr., J.F.: The bargaining problem. Econometrica J. Econometric Soc. **18**(2), 155–162 (1950)
9. Nicosia, G., Pacifici, A., Pferschy, U.: Price of fairness for allocating a bounded resource. Eur. J. Oper. Res. **257**(3), 933–943 (2017)
10. Valenta, V., Maršálek, R., Baudoin, G., Villegas, M., Suarez, M., Robert, F.: Measurements, analyses and observations. In: 2010 Proceedings of the Fifth International Conference on Cognitive Radio Oriented Wireless Networks & Communications (CROWNCOM), pp. 1–5. IEEE (2010)
11. Wang, E., Yang, Y., Wu, J.: A Knapsack-based message scheduling and drop strategy for delay-tolerant networks. In: Abdelzaher, T., Pereira, N., Tovar, E. (eds.) EWSN 2015. LNCS, vol. 8965, pp. 120–134. Springer, Cham (2015). doi:10.1007/978-3-319-15582-1_8

The 5th International Workshop on Frontiers in Intelligent Networking and Collaborative Systems (FINCoS-2017)

Research on Security and Interoperability in Middleware

Cheng Zhou[1], Mingwen Tuo[2], Mingfu Tuo[1(✉)], Xin Zhao[1], and Lei Wang[1]

[1] Air Force Engineering University, Xi'an, China
mftuo@163.com
[2] Troops 93115 of PLA, Shanghai, China

Abstract. The interoperability between security services of different middleware platforms directly affect the degree of resource sharing in the distributed environment and business interoperability. The architecture and mechanism of CSIv2 are analyzed, an implementation scheme for security and interoperability based on interceptor technology is proposed. Finally, we implement the scheme on StarBus, a middleware platform. Test result shows correctness of the implementation.

1 Introduction

In order to meet the security communication requirements of specific applications, many CORBA product suppliers add security services in their object request broker (ORB). Interoperability of different security service provided by these CORBA products can directly influence degree of resource sharing and business interoperability. It is also one of the key issues in the distributed application system integration technology [1–3].

In order to realize the interoperability of different CORBA products security services that is security interoperability, the OMG has defined several security protocols for secure interoperability (including SECIOP, SSLIOP, etc.) and recommend several security mechanisms (including SSL, SPKM, Kerberos, original, etc.) in its security standard CORBASec, Furthermore, it determine the three levels of security interoperability [4–6]:

(1) CSI Level 0: security policy based on the identity is supported, while delegation is not supported.
(2) CSI Level 1: security policy based on the identity is supported, and unrestricted delegation is not supported.
(3) CSI Level 2: security policy based on the identity and privilege (such as role, group, ability, etc.) are supported, and controlled delegation is supported.

Delegation here means: In a distributed environment, when a user or an entity cannot complete one request alone, it need to entrust other user or entity to help it to complete request. The assist entity should use some security attributes of the initiate entity in the process of complete the request, such as the status and role, etc. It is a kind of common behavior in distributed object systems. If the initiate entity can control the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_41

delegated authority of assistant entity, it is called unrestricted delegation, otherwise, it is called uncontrolled delegation.

According to the above security interoperability level judgment standard, it puts forward all kinds of security mechanism for different layers in CORBA Sec [7, 8]. SECIOP protocol can reach to level 2. However, it is not widely supported by security products supplier. SSL is the most widely used security mechanism, this mechanism supports authentication and message protection (i.e., the message integrity and confidentiality) based on X509 certificate, and it can meet the CSI Level 0, but in the system still exists some deficiencies, such as:

(1) Delegation is not supported.
(2) Security strategy based on privilege attribute is not supported.
(3) Clients usually have not certificates in a specific environment.

In order to support a more extensive and complete security interoperability, the OMG and Java Community formulate the general security interoperability protocol second edition (CSIv2) jointly. The OMG has added them to CORBA3.0 specification [9].

The architecture and security interoperability principle of CSIv2 is analyzed. A realization scheme based on interceptor is put forward and this scheme is implemented on StarBus, a middleware platform following the CORBA3.0 specification. This make StarBus can provides end-to-end security service and access control. The security service conforms to the CSIv2 interoperability specification.

2 Related Terms

To facilitate understanding, some main concepts and terminology involved are introduced briefly.

Principal: A user or entity that has registered and been certified in the system. It has a unique name to identify itself.

Security Attribute: A set of attributes Associated with a principal. It mainly includes identity attribute and privilege attribute. The identity attribute refers to the name of a principal. Privilege attribute refers to the role, group, capability and other features. Security attribute is typically used for access control.

Credential: A set of security attributes.

Delegation: A principal authorizes other principal to use its identity or privilege. In delegation scenario, Principals call one by one, so an invocation chain is formed. The first node invocation chain is called Initiator. The intermediate node is called Intermediate. The end node is called Target.

Identity Assertion: An access entity initiates a request using other identity. It is the basis of impersonation and it is the important way to realize unrestricted delegation.

Proxy Endorsement: A principal allows another principal to use its some security attributes. Initiator decides whether intermediate can initiate a request using the identity or privilege of initiator. It is an important way to realize controlled delegation.

3 CSIv2 Analysis

CSIv2 is mainly used to solve the interoperability of certification, delegation and privilege attribute among different CORBA products. At the same time, it also can realize the security interoperability between EJB and CORBA, two classical middleware. This paper mainly discusses its support for CORBA product security interoperability [10–12].

From an architectural point of view, CSIv2 is divided into transmission layer, client authentication layer and security attribute layer, as shown in Fig. 1. The transport layer provides message protection, target-to-client authentication (that is, target certificate client) and client-to- target authentication. SECIOP or SSL/TLS, which are recommended by CORBASec, can be used to this layer. Client authentication layer is needed when the transport layer for client authentication is not sufficient or it cannot implement the necessary authentication mechanism (such as SSL client have no certificate), The other authentication mechanism is applied to make supplementary certification. Attribute layer is used to transfer security attributes for client to server. These layers cooperate with each other and relative independence. At least one layer cannot be empty.

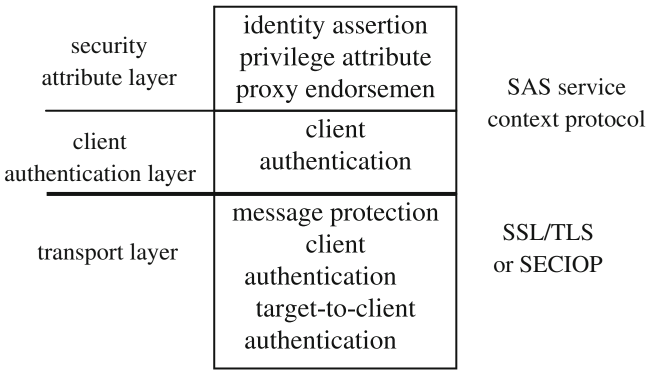


Fig. 1. The architecture of CSIv2

4 Design and Implementation of Security and Interoperability

4.1 Implementation Framework

The design scheme is based on interceptor technology [13, 14]. Interceptor is a bridge between the ORB and the ORB services, it provides highly flexible method to add a portable ORB services into ORB following CORBA specification. The interceptor technology makes the clear separation of the ORB kernel with the ORB, realizing independence and coexistence of different ORB services. So it is one of the most common way to add ORB service into the ORB core. Security interoperability is a kind of the ORB services in essence, so the interceptor technology implementation is very appropriate.

In this framework, client request interceptor, server request interceptor and IOR interceptor are registered to ORB during the CSIv2 initialization time. The ORB will call every intercept points of these interceptors at the appropriate time by its interceptor framework. Intercept point refers to methods of interceptor that can be invoked by interceptor framework. The concrete implementation of CSIv2 is realized in these interceptors point. A security context is constructed in the client interceptor and verified in the server interceptor. The server-side IOR is configured in the server IOR. The overall design framework is shown in Fig. 2.

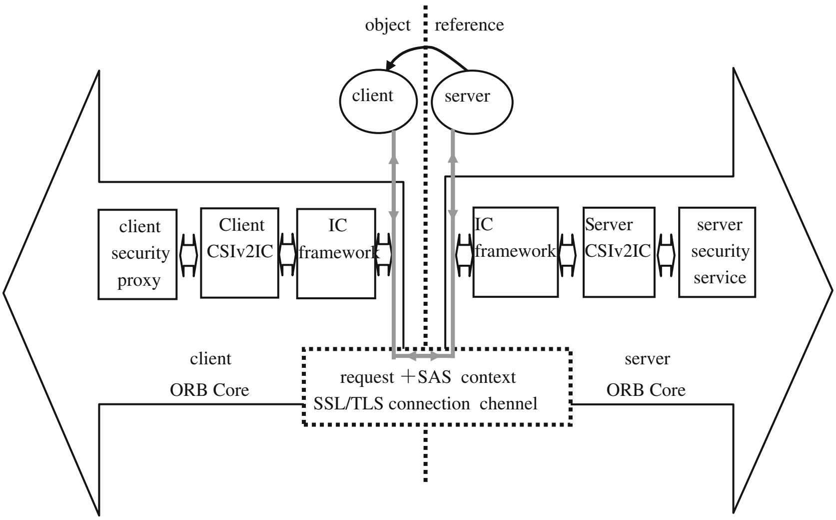


Fig. 2. The design framework

4.2 Security Interoperability Request Processing

Here we give a delegation scenario for example to describe how the request is processed after CSIv2 plugin is added, as shown in Fig. 3. The solid line shows the request flow, while dotted line shows security information flow in the figure.

(1) Initiator, intermediate and target initialize security services, including setting up security strategy, etc.
(2) The initiator credentials information and initiate the request.
(3) The client interceptor of initiator intercepts request, select the appropriate composite security mechanism of intermediate IOR according to its security policy, establish a security context using the application credentials information, then add it to the GIOP message header.
(4) The request with security context is send from initiator to intermediate by SSL security connection channel.

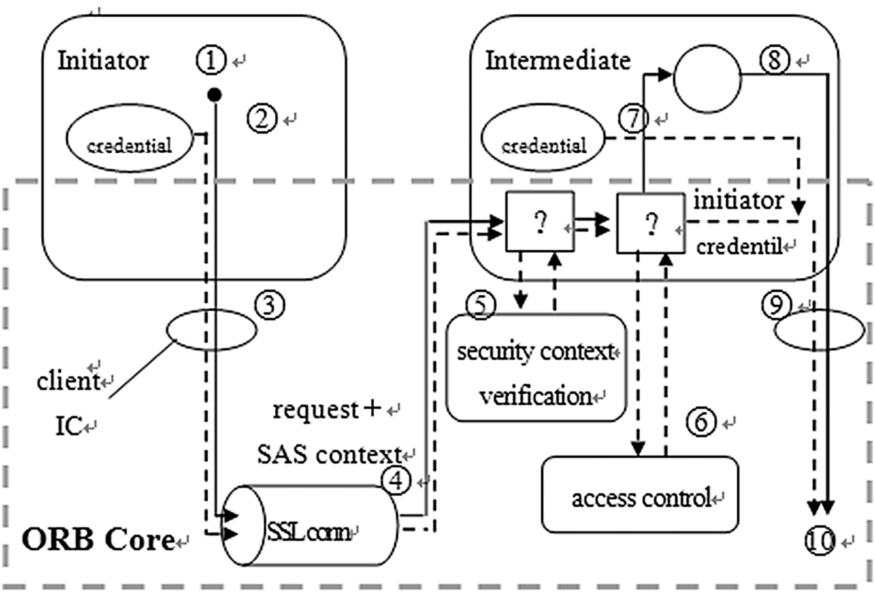


Fig. 3. The request processing in delegation scenario

(5) The server-side interceptor of intermediate intercepts requests, get the security context from the GIOP message header and verify it, including mechanism check and certificate verification.
(6) If the verification is successful, then security attributes are into the access control module and saved at the same time. Otherwise, the request is refused and exception is returned.
(7) If access control is successful, the request will be delivered to the application of intermediate. Otherwise, the request is refused and exception is returned.
(8) If the intermediate can meet the request independently (that is no delegation), then it provide the corresponding services, and returns the success message of the request and the corresponding CSIv2 response security context. Otherwise (that is it need delegation), it provides its own credentials, and send request to target that can provide the service.
The client interceptor of intermediate intercepts request, select the appropriate composite security mechanism of target IOR according to its security policy, establish a security context using the its own credentials and received credentials, then add it to the GIOP message header.
(10) The subsequent processing steps are similar to the former.

4.3 Security Context Verification Algorithm

Unreasonable security context validation logic will lead to legal requests are not being met, while some illegal request is happed unauthorized. It will make security interoperability lost meaning. Therefore, the security context verification algorithm plays an important role for the target. A security context verification algorithm is given in this paper as follows by analyzing the CSIv2 specification verification:

security context verification algorithm

```
TSS get security context from customer request
if there is SAS context in request {
  if  CAT != NULL {
     Analyze the content of CAT add verify it;
     if verification is successful
       get client id (CId) from CAT;
     else refuse request and raise exception;
  else {   // client authentication layer is empty
       verify the transport layer context;
       if verification is successful
         get client id (CId) from transport layer;
       else
         refuse request and raise exception;
  }
  if   AT !=NULL or IT != NULL {
   get  security  attributes(SecAttr  )from  attribute
   layer;
 run access control by two-tuples(CId, SecAttr);
  }
  else
     run access control by CId;
} // end of if at line 1
else {  //has no SAS context
   verify transport layer context ;
   if verification is successful
      run access control by Cid of transport layer;
}
if access control is passed
   process request and send reply; //request success
else
   refuse request and raise exception;
```

5 Security and Interoperability Test

In order to verify the correctness of the scheme in this paper, we do a detailed security and interoperability test. The test has two steps. The first step is security and interoperability test among objects of StarBus. The second test is security and interoperability between StarBus and ORBacus. ORBacus is CORBA product developed by IONA company. It is also one of typical ORB implementation in the world. Its security services can be provided by ORBASec, which is a CSIv2 plugin developed by the Adiron company. The plug-in follows CSIv2 specification, and it is one of the typical implementation of CSIv2.

Test result shows that not only the objects of StarBus have security and interoperability, but also the StarBus objects and ORBacus objects have security and interoperability. Table 1 is the security interoperability test of StarBus objects. In Table 2 is shown the security interoperability test between StarBus object and ORBacus object. Three rows of each scenario in the table are the security mechanism of three layers in CSIv2. The first row is security mechanism of the transport layer. The second row is the security mechanism of client authentication layer. The third row is the mechanism of security attribute layer. Blank represents no corresponding mechanism.

Table 1. Security interoperability test among StarBus objects

Scenario	Initiator CSS	Intermediate TSS	Intermediate CSS	Target TSS	Memo
1	SSL			SSL	Transport layer, SSL, no delegation
2	IIOP			IIOP	SAS protocol, no delegation
	GSSUP			GSSUP	
3	IIOP	IIOP	SSL	SSL	SSL+SAS, identity assertion, delegation
	GSSUP	GSSUP			
		Principal	Principal	Principal	
4	SSL	SSL	SSL	SSL	Complete CSIv2 interoperability
	GSSUP	GSSUP	GSSUP	GSSUP	
	Principal	Principal	Principal	Principal	

Table 2. Security interoperability test between StarBus and ORBacus

Scenario	Initiator CSS	Intermediate TSS	Intermediate CSS	Target TSS	Memo
1	SSL			SSL	target: ORBacus client: StarBus
2	SSL			SSL	target: Star Bus client: ORBacus
3	IIOP	IIOP	SSL	SSL	Intermediate: ORB acus others: StarBus
	GSSUP	GSSUP			
		Principal	Principal	Principal	
4	IIOP	IIOP	SSL	SSL	Intermediate: Star Bus others: ORBacus
	GSSUP	GSSUP			
		Principal	Principal	Principal	
5	SSL	SSL	SSL	SSL	Intermediate: ORB acus others: StarBus
	GSSUP	GSSUP	GSSUP	GSSUP	
	Principal	Principal	Principal	Principal	
6	SSL	SSL	SSL	SSL	Intermediate: Star Bus others: ORBacus
	GSSUP	GSSUP	GSSUP	GSSUP	
	Principal	Principal	Principal	Principal	

6 Conclusions

The security and interoperability of middleware services directly affect the degree of resource sharing in the distributed environment and business interoperability. It is also one of the key issues in the distributed application system integration technology. In order to support more complete security and interoperability, OMG and Java groups jointly launched CSIv2.

This paper firstly analyzes the architecture and mechanism of CSIv2, then proposes a implementation scheme for security and interoperability based on interceptor technology. We implement the scheme on StarBus, a middleware platform that conforms to CORBA3.0. This provides the ability of security interoperation with other CORBA products to StarBus.

Test results verify the validity of the proposed sheme. It also shows that the implementation of this solution in StarBus conforms to the CSIv2 specification. This scheme adopts the technology of the interceptor. It makes the clear separation of the ORB kernel with the ORB, realizing independence and coexistence of different ORB services.

References

1. Object Manage Group: Common Object Request Broker Architecture: Core Specification v 2.4.0-v 3.0.2. OMG, 2015.12 20-35
2. Common Secure Interoperability Version 2 Request For Proposal, OMG, 2016.11 56-69
3. Klein, R., Rilling, S., Usov, A., Xie, J.: Using complex event processing for modelling and simulation of cyber-physical systems. Infrastructures **9**(1/2), 148–172 (2013)
4. Mola, L., Rossignoli, C., Carugati, A., Giangreco, A.: Business intelligence system design and its consequences for knowledge sharing, collaboration, and decision-making: an exploratory study. Int. J. Technol. Hum. Interact. **11**(4), 1–25 (2015)
5. Mezghani, K., Ayadi, F.: Factors explaining is managers attitudes toward cloud computing adoption. Int. J. Technol. Hum. Interact. **12**(1) 1–20 (2016)
6. Yuriyama, M., Kushida, T.: Integrated cloud computing environment with IT resources and sensor devices. Int. J. Space Based Situated Comput. (IJSSC) **1**(2–3), 174–182 (2011). doi:10.1504/IJSSC.2011.040342
7. Boyinbode, O., Le, H., Takizawa, M.: A survey on clustering algorithms for wireless sensor networks. Int. J. Space Based Situated Comput. (IJSSC) **1**(2–3), 137–150 (2011). doi:10.1504/IJSSC.2011.040339
8. Di Stefano, A., Morana, G., Zito, D.: QoS-aware services composition in P2PGrid environments. Int. J. Grid Util. Comput. **2**(2), 139–147 (2011). doi:10.1504/IJGUC.2011.040601
9. Lee, E.A.: Cyber physical systems: design challenges. In: The 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC-2008), Orlando, Florida, USA, pp. 363–369, 5–7 May 2008
10. Lee, E.A.: The past, present and future of cyber-physical systems: a focus on models. Sensors **15**, 4837–4869 (2015)
11. Guyeux, C., Makhoul, A., Atoui, I., Tawbi, S., Bahi, J.M.: A complete security framework for wireless sensor networks: theory and practice. Int. J. Inf. Technol. Web Eng. (IJITWE) **10**(1), 47–74 (2015)

12. Saripalle, R.K., Demurjian, S.A., Blechner, M., Agresta, T.: HOD2MLC: hybrid ontology design and development model with lifecycle. Int. J. Inf. Technol. Web Eng. (IJITWE) **10** (2), 16–42 (2015)
13. Xingshe, Z., Yalei, Y., Gang, Y.: A method for cyber-physical system dynamic behavior modeling. Chin. J. Comput. **37**(6), 1411–1423 (2014)
14. Zhou, C., Yin, Z., Feng, Y.: Events algebra of triggers. Comput. J., 1–10 (2017)

Home Energy Management Using Enhanced Differential Evolution and Chicken Swarm Optimization Techniques

Muhammad Awais, Zain Ul Abadeen, Taimur Bilal, Zafar Faiz, Muhammad Junaid, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. Energy optimization is important aspect of smart gird (SG). SG integrates communication and information technology in traditional grid. In SG there is two-way communication between consumer and utility. It includes smart meter, Energy Management Controller (EMC) and smart appliances. Users can shift load from on peak hours to off peak hours by adapting Demand Side Management (DSM) strategies, which effectively reduce electricity cost. The objectives of this paper are the minimization of power consumption, electricity cost, reduction of Peak to Average Ratio (PAR) using Enhanced Differential Evolution (EDE) and Chicken Swarm Optimization (CSO) algorithms. For the calculation of cost Critical Peak Pricing (CPP) is used. The simulations result show that proposed schemes reduce electricity cost, reduce power consumption and PAR.

1 Introduction

The Smart Grid (SG) is the future of electric grid. It is different from traditional grid in many ways. There is energy loss and more carbon emission in traditional grid. There is no load management which is big issue of traditional grid. There is no concept of energy storage in traditional grid. SG revolutionize the electric grid by incorporating communication and load management technology. It is robust and self-healing system. There is also backup of electric energy in SG. In Fig. 1 features of smart grids are shown. SG features smart meter and Energy Management Controller (EMC), which manage the energy consumption. EMC schedules all appliances in such a way that reduce overall load and cost of electricity, by transferring load to off peak hours. For the reduction of load and cost and maximization of user comfort scheduling algorithms and Demand Side Management (DSM) strategies are deployed. DSM helps to balance between demand and supply [1]. DSM include automated energy management system which control electricity consumption in real time. When an action is taken to change the demand of electricity in response to change in Electric Price (EP) is called Demand Response (DR). EMC uses both

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_42

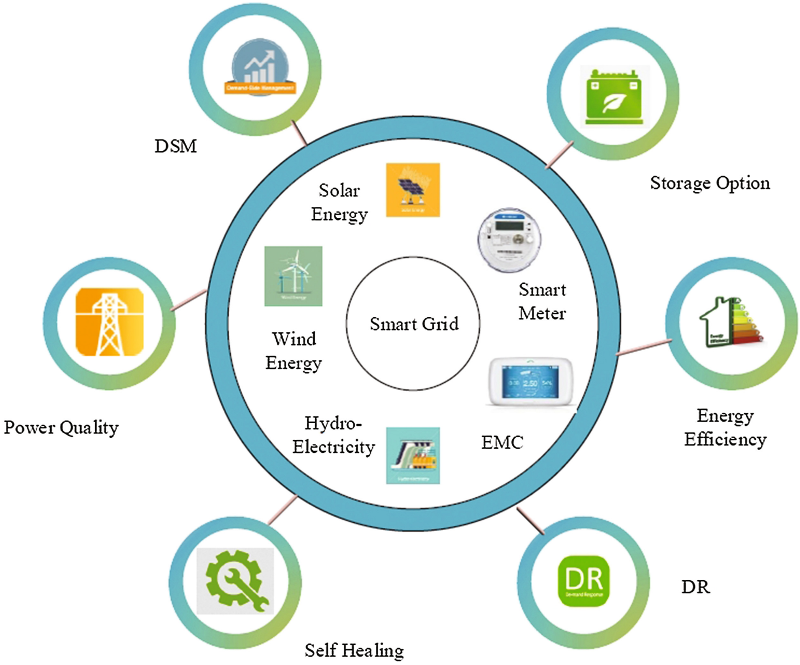


Fig. 1. Smart grid

user preference and price signals for scheduling of appliances. DR includes Real Time Pricing (RTP), Time of Use (TOU), Critical Peak Pricing (CPP). In RTP scheme EP changes every hour. In TOU the price remains same for fixed period of time. In CPP price is previously determined for peak hours. In this paper a Home Energy Management System (HEMS) is presented which reduce electricity cost, electricity load and Peak to Average Ratio (PAR). Enhanced Differential Evolution (EDE) and Chicken Swarm Optimization (CSO) optimization technique are used for HEMS. The rest of paper is organized as follows: Sect. 2 is about related work, in Sect. 3 problem statement is discussed, Sect. 4 is about proposed methodology. In Sect. 5 simulation results are discussed. Finally, in Sect. 6 conclusion is presented.

2 Related Work

SG consists of smart meter, smart appliances and (EMC). SG allow two-way communication between consumer and grid station. Consumer can schedule their appliances with Home Energy Management System (HEMS) to reduce cost, load, Peak to Average Ratio (PAR) and maximize user comfort. Research has been done in the field of SG. Different scheduling algorithms have been proposed which mainly reduce the cost and electricity load. In [1], authors discuss HEMS for one and multiple homes. Different heuristic algorithms are used for optimization such Genetic Algorithm (GA), Binary Particle Swarm Optimization (BPSO) and Aunt Colony Optimization (ACO). Electricity load is categorized into fixed appliances, shiftable appliances and elastic appliances. GA algorithm perform better than other optimization algorithm. In this work reduction of electricity bills, power consumption and maximization of user comfort are considered. Renewable Energy Sources are integrated. Achievement of this work is reduction of cost minimizing PAR and maximization of user comfort. Cost of RESs is not calculated that user have to pay for the maintenance. An efficient home energy management system is presented in [2] that use integer linear programming technique for the residential area. In this work user preference and requirement of individual appliances are considered. Appliances are divided into non shiftable, power shiftable and time shiftable. Achievement of this research are balanced load in each hour by scheduling optimal power and operational time for power shiftable appliance and time shiftable appliances that is according to optimum power consumption pattern of each appliance. The proposed scheduling algorithm reduce peak hour load and satisfy user comfort and preference of devices. Electricity bills increases when user comfort is considered which is the limitation of this work. Consumer can reduce electricity bills, reduce PAR by scheduling their appliances in smart gird. Authors in [3] aims to optimize the electricity usage by efficient scheduling algorithm. For this purpose, GA scheduling technique is used. Real Time pricing (RTP) and Inclining Block Rate (IBR) pricing schemes are used for optimization of appliances. Appliances are categorized into automatically Operated appliances and manually Operated appliances and their time of operation is also considered. The objectives of this work is to reduce electricity cost and minimizing PAR by GA scheduling algorithm. After applying scheduling technique waiting time of appliances and cost is reduced. By reduction of PAR utilities companies also get benefit which result in stability of grid station. Deficiency of this work is ignorance of user comfort. Scheduling algorithm perform only better if RTP pricing scheme is combined with IBR.

A heuristic algorithms based EMC is developed in [4] for home energy management system. Five heuristic algorithms GA, binary particle swarm optimization (BPSO), bacterial foraging optimization algorithm (BFOA), wind-driven optimization (WDO) and hybrid genetic wind driven (GWD) are used for scheduling. Scheduling is applied for single and multiple home. RTP is used as pricing scheme. Load is shift from peak hour to off peak hours using shiftable and non-shiftable appliances in residential area. The objective of this work is to optimize energy consumption, maximization of user comfort and reduction of cost. In simulation result the proposed GWD algorithm perform better than other heuristic algorithm in term of cost and PAR reduction. Limitation is non integration of RESs and compromising user comfort when cost is reduced. In [5] authors describe power scheduling strategy using day ahead pricing scheme for achieve desired tradeoff between cost and user comfort. Three operation modes are discussed for different parameters settings. In first operation mode consumer care about cost and does not consider user comfort. In second operation mode consumer consider discomfort and does not care about cost. In third operation mode consumer cares about both discomfort and cost of electricity. In simulation optimal scheduling strategy is achieved for three operation modes. A desired tradeoff is achieved between user comfort and electricity bills. Limitation of this work is increase in waiting time when cost is reducing and there is no integration of renewable energy sources which can reduce carbon emission. Environment friendly renewable energy resources (RERs) are focused in this research work [6]. Authors have used dynamic programming to schedule appliances which are categorized as must run and controllable. Objective of this paper is to develop a model in which users can buy and sell locally generated electricity. The achievements are of research are that users are able to sell their excess power generation and choose their offered price and to maximize revenue. Users have to adjust cost for battery and photovoltaic system replacement which are drawbacks of this work.

Genetic algorithm (GA) is used as load satisfaction algorithm for minimizing cost [7]. Three budget scenario are performed. User satisfaction in term of minimum cost is the main problem address in this paper. After applying GA scheduling algorithm in simulation result a pattern of energy consumption is achieved to meet predefined budget of consumer. The results are compared with randomly used appliances. The results show that GA algorithm have minimum electric cost when compared to randomly used appliances. The main purpose of this work is to reduce cost which compromise on user comfort. Binary backtracking search algorithm is used for home energy management system (HEMS) in this study [8]. It limits the demand of total load and schedule the appliance to shift load. A most common home appliances are used for the simulation. Two case of appliances are considered first one 4 to 11pm and second case is different time in day. Scheduling of home appliance to shift load to off peak hours are the main objective of this study. Simulation result shows that total energy consumption and electricity cost is reduced and energy is saved in peak hours. User priorities are compromised which is deficiency of this work. In [9] authors present price based HEM scheduler which is used for optimization. TOU and IBR pricing schemes are used. The motivation of this paper to formulate operational priority of appliances from consumer perspective. The achievement after applying price based HEM scheduler are that it adds priorities in HEM scheduler which lead to minimum cost. And IBR tariff lead to minimum cost and reduction of load in peak hours. In this paper there no integration of RESs. Coordinated scheduling of home appliances is presented in [10] considering multiple home with PV and battery system. In this work there is use of RESs. The purpose of this research is to formulate a home energy system model that reduce inconvenience, cost of electricity and carbon emission. The achievement of this work is the reduction of power consumption from gird and consumer cost. Consumer have to pay Extra maintenance cost for PV and battery system.

In [11] authors describe Real Time Appliance based power management system (Ab-HPMS) that manage power consumption. Appliance-based Rolling Wave Planning (Ab-RWP) used as control algorithm. The main objective of this research is integration of smart appliance with HEM. Comparison is performed of smart appliances with conventional appliances with respect to power saving. In simulation result energy efficiency is improved of smart appliances. Cost is reduced and user comfort is maximized. The limitation of this paper is that appliance usage frequency is ignored. Appliances Coordination (ACORD) scheme and Appliances Coordination with Feed In (ACORD-FI) scheme are used for HEM [12]. The purpose of this work is to reduce energy consumption, cost and interruption period. To reduce peak demand. Proposed scheduling scheme reduces energy consumption and reduce total cost of electricity bills. User comfort is compromised in this work.

3 Problem Statement

The main purpose of DSM is to manage the electric energy so that it can be saved and environment can be protected by less carbon emission. DSM can be customized according to user requirements and reducing electricity cost by shifting load to off peak hours. DSM effectively manage the electric energy that decrease load during peak hours and reduce cost of electricity. In literature many optimization techniques have been proposed for HEMS. Authors in [15] present HEMS that forecast electric energy consumption at different time of day to balance electricity load. If load is high during peak hours, it increases electricity bills. This paper presents a EMC system model which reduce electricity cost, minimize load during peak hours and reduce PAR by using CPP as pricing scheme (Fig. 2).

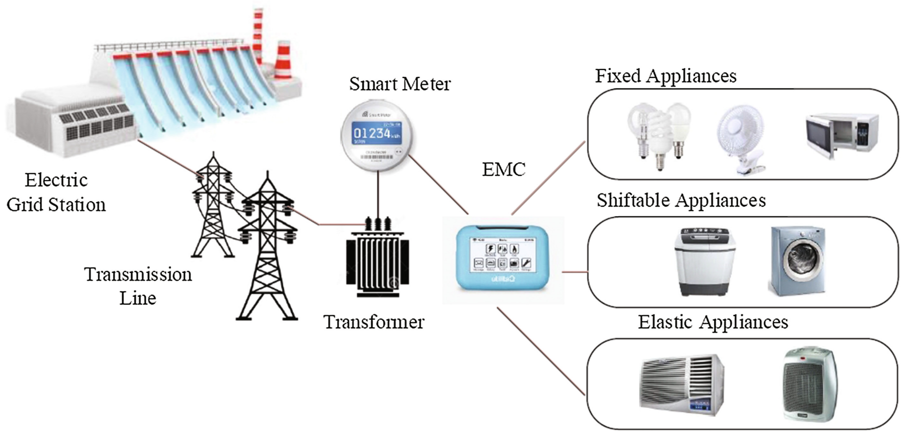


Fig. 2. System model

4 Proposed Solution

The proposed schemes for optimization are EDE and CSO. EDE is enhanced version of DE. Mutation, crossover and selection are main operation of DE. From a randomly chosen vector a mutant vector is generated which is added in target vector [13]. In EDE five trial vectors are generated. A trial vector which have minimum objective values is selected. Chicken Swarm Optimization (CSO) is bio inspired algorithm which is proposed by Meng et al. [14]. CSO is based on behavior of chicken (chicks, hens, rosters). There are several groups in chicken swarm. In each group there is dominant roster. Groups in chicken swarm are based on fitness of chicken themselves. Rosters have best fitness value. Chicken with worst fitness values will be chicks while other will be hens. Hens randomly choose a group to live. CSO algorithm have these steps. Population is initialized randomly of N chicken. Fitness value is calculated of N chickens. Relationship is determined between chicks, hens and rosters. For new solution a fitness value is calculated. Solutions are compared with old ones and new one. Better solution is replaced with new ones. Best solution is calculated. Critical Peak Pricing (CPP) schemes is used for the calculation of cost. The appliances are classified into following three categories [1]. Fixed appliances includes lighting, fans, clothes iron microwave oven, toaster and coffee maker. Shiftable appliances includes washing machine, cloth dryer and dish washer. Elastic appliances includes air conditioner, refrigerator water heater and space heater. Parameters of appliances are summarized in Table 1.

Table 1. Parameters of appliances

Group	Appliances	Power rating (kW/h)	Length of operation time (hour)
Fixed appliances	Lighting	0.6	7
	Fans	0.75	13
	Clothes iron	1.55	5
	Microwave oven	1.18	10
	Toaster	0.5	6
	Coffee maker	0.8	4
Shiftable appliances	Washing machine	0.78	5
	Cloth dryer	4.40	4
	Dish washer	3.6.	6
Elastic appliances	AC	1.44	10
	Refrigerator	0.73	13
	Water heater	4.45	11
	Space heater	1.50	13

Algorithm 1. Chicken Swarm Optimization

```
Initialize a population of N chickens and define the related parameters
Evaluate the N chickens fitness values, t=0;
while t < MaxGeneration do
    if t%G == 0) then
        Rank the chickens fitness values and establish a hierarchal order in the swarm
        Divide the swarm into different groups, and determine the relationship between
        the chicks and mother hens in a group
    end if
    for i = 1 : N do
        if i == rooster then
            Update its solution/location
        end if
        if i == hen then
            Update its solution/location
        end if
        if i == chick then
            Update its solution/location
        end if
        Evaluate the new solution
    end for
end while
```

5 Simulations and Discussions

This section provides detail of simulation results. Simulations are conducted in MATLAB for evaluation of EDE and CSO. A single home with 13 appliances is

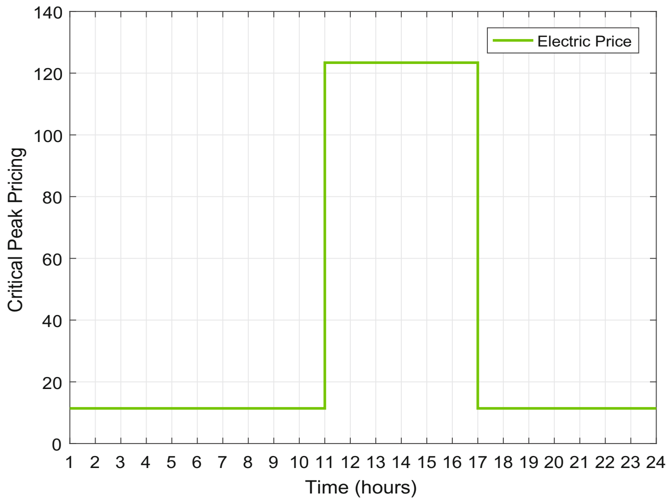


Fig. 3. Critical peak pricing

considered for simulation. There are 3 categories of appliances, fixed appliance, shiftable appliances and elastic appliances. CPP pricing scheme is used of the calculation of electricity cost which is shown in Fig. 3.

In Fig. 4 electricity cost per hour is represented. Results demonstrate that cost reduced during peak hours by shifting load to off peak hours. Figure 5 illustrates total cost of unscheduled and EDE, CSO scheduled. Total cost of CSO is low as compared to EDE and unscheduled. Figure 6 represent per hour load. Scheduling algorithms shift load to off peak hours to reduce cost. PAR is

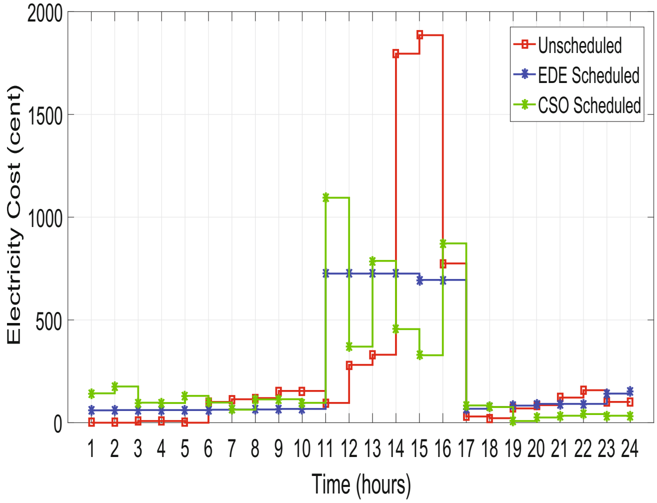


Fig. 4. Per hour cost

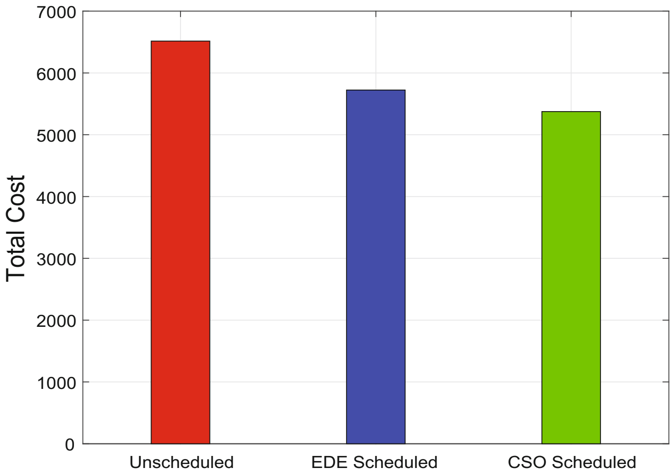


Fig. 5. Total cost

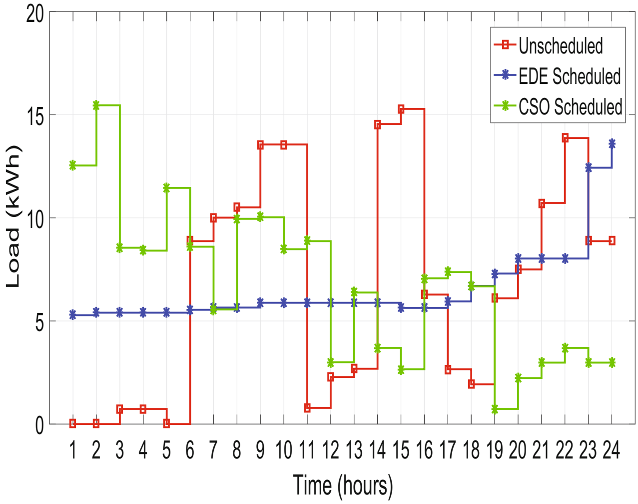


Fig. 6. Per hour load

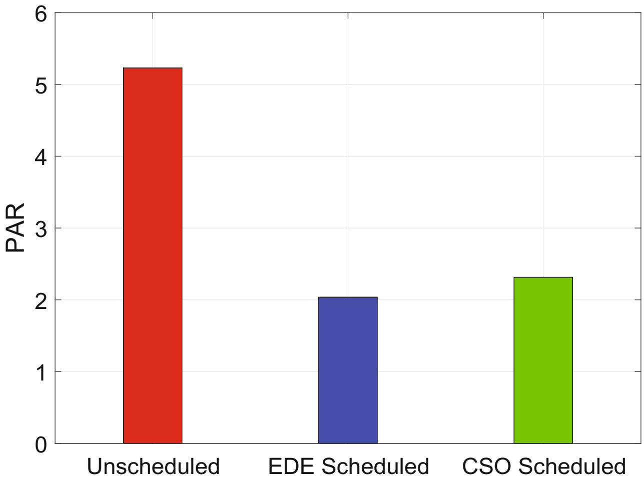


Fig. 7. PAR

illustrated in Fig. 7. PAR of EDE is low as compared to CSO and unscheduled. In Fig. 8 waiting time is represented. Waiting time of EDE is low as compared to CSO.

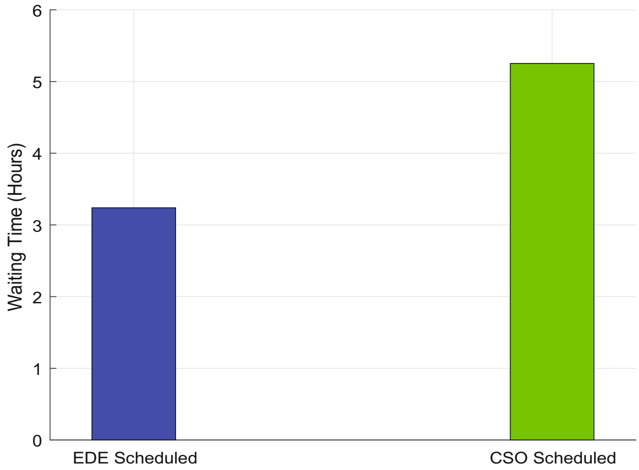


Fig. 8. Waiting time

6 Conclusion

In this paper HEMS model is presented for house home appliances. The performance of EDE and CSO is evaluated in terms of minimizing cost, load and PAR. CPP pricing scheme is used for the calculation of electricity cost. User comfort is also calculated in term of waiting time of appliances to turn on. Waiting time of EDE is low as compared to CSO. SG is the need of modern world which have integrated new technology of communication and management. Energy is a valuable thing which should be properly managed and saved.

References

1. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
2. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), p. 15. IEEE (2012)
3. Zhao, Z., Lee, W.C., Shin, Y., Song, K.-B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
4. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electric. Power Energy Syst. **78**, 320–325 (2016)
6. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)

7. Ogunjuyigbe, A.S.O., Ayodele, T.R., Akinola, O.A.: User satisfaction induced demand side load management in residential buildings with user budget constraint. Appl. Energy **187**, 352–366 (2017)
8. Maytham, S., Ahmeda, B., Mohameda, A., Khatibc, T., Shareefd, H., Homode, R.Z., Ali, J.A.: Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. Energy Build. **138**, 215–227 (2017)
9. Rastegar, M., Fotuhi-Firuzabad, M., Zareipou, H.: Home energy management incorporating operational priority of appliance. Electric. Power Energy Syst. **74**, 286–292 (2016)
10. Setlhaolo, D., Xia, X.: Combined residential demand side management strategies with coordination and economic analysis. Electric. Power Energy Syst. **79**, 150–160 (2016)
11. Ozkan, H.A.: Appliance based control for home power management systems. Energy **114**, 693–707 (2016)
12. Albalas, F., Khamayseh, Y., Mardini, W.: Improved appliance coordination scheme with waiting time in smart grids. Int. J. Adv. Comput. Sci. Appl., April 2016
13. Arafa, M., Sallam, E.A., Fahmy, M.M.: An Enhanced Differential Evolution Optimization Algorithm. IEEE (2014) ISBN 978-1-4799-3724-0/14/2014
14. Meng, X., Liu, Y., Gao, X., Zhang, H.: A new bioinspired algorithm: chicken swarm optimization. In: Advances in Swarm Intelligence, ICSI 2014, pp. 86–94 (2014)
15. Collotta, M., Pau, G.: An innovative approach for forecasting of energy requirements to improve a smart home management system based on BLE. IEEE Trans. Green Commun. Networking, March 2017

A Novel Meta-heuristic Technique for Energy Optimization in Smart Grid

Shaista Bibi[1], Mahnoor Khan[1], Bushra Abbasi[1], Muhammad Fawad[2], Ayesha Anjum Butt[1], and Nadeem Javaid[1](✉)

[1] COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
[2] University Institute of Information Technology, PMAS - Arid Agriculture University, Rawalpindi 46000, Pakistan
https://www.njavaid.com

Abstract. The increase in the energy consumption causes a serious crisis, especially during on-peak hours when the demand of energy consumption is high. Consequently, the peak to average ratio and electricity cost will be increased. This issue can be overcome by integrating Demand side management (DSM) with traditional Smart grid (SG), so that electricity utilization can be minimized during on-peak hours by efficiently distributing them into off-peak hours. In this paper, Crow search algorithm (CSA) is proposed to schedule the appliances for DSM and the performance of Home energy management system (HEMS) is assessed by two meta-heuristic techniques; Enhanced differential evolution and CSA. The reduction in cost and peak to average ratio along with increase in user comfort is mainly focused in this paper. Moreover, the electricity cost is based on real time pricing scheme. The main objective is to provide a comparative analysis of the aforementioned techniques for energy optimization using simulations in HEMS. The simulations results show that our proposed technique outperformed as compared to the existing meta-heuristic technique.

1 Introduction

To overcome the growth in demand of energy consumption is one of the main concern nowadays. The International energy outlook 2016 [1] shows the exponential increase in demand of power consumption over the next three decades as shown in Fig. 1 [1]. Apart from this issue, the inefficiencies of traditional grid such as lack of energy management system and its inefficiency are the key reasons to bring in the Smart grid (SG). SG is the environment friendly, elastic, consistent and controllable concept which combine the bidirectional communication and information across the traditional grid and the utility along with smart meters, smart appliances, Renewable energy resources (RES) and Energy efficient resources (EER) as shown in Fig. 2. The term smart differentiate the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_43

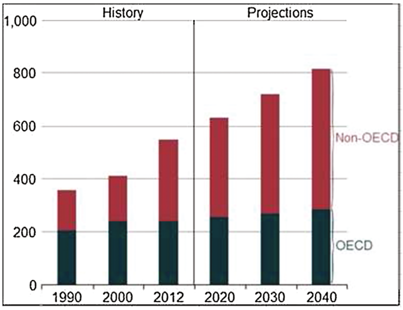


Fig. 1. World energy consumption, 1990–2040 (quadrillion Btu) [1]

aforementioned features from other traditional grid by their capability of computation, exchanging information and storing it for future use. This stored information is used for energy management to fulfill the requirements of the customer and the utility. This feature also provides the facility to the customer to get informed of the decisions regarding the usage of energy [2].

Home energy management system (HEMS) is one of the most important aspect of the DSM. As it provides a reduction in electricity cost, Peak to average ratio (PAR) and an increase in User comfort (UC) to some extent. Load management and Demand response (DR) are the two main functions of DSM, where the DR offers incentives to the customer to shift their load from high peak hours to low peak hours. However, in load management the optimality in energy management is objective. It tries to maximize the efficient use of energy and reduces the blackouts etc. According to article [2], the energy consumption can be reduced up to 15% by adapting DSM strategies. The aim of DSM is also reducing PAR, energy consumption and electricity bill along with affordable increase in UC.

The energy optimization is trying to be achieved by using different meta heuristic techniques such as Binary particle swarm optimization (BPSO) based on the flock behavior, Genetic algorithm (GA) based on natural selection, Harmony search algorithm (HSA) based on harmony improvisation techniques. Similarly, Bacterial foraging (BF), Ant colony optimization (ACO), and EDE etc. All these modern nature inspired techniques are called meta-heuristic algorithms that use certain trade offs of local search. These algorithms does not guarantee the 100% accuracy. However, it provide optimal solution. Similarly, we cannot get the 100% accuracy in all performance parameters such as PAR reduction, UC maximization, Electricity cost minimization and optimal scheduling by minimizing the load. There is always a trade-off between some of the UC and cost. However, these different techniques show better results in some parameters while lagging behind in comparison with others. Similarly, we are also proposing a novel meta heuristic technique for energy optimization. Since they work very efficiently.

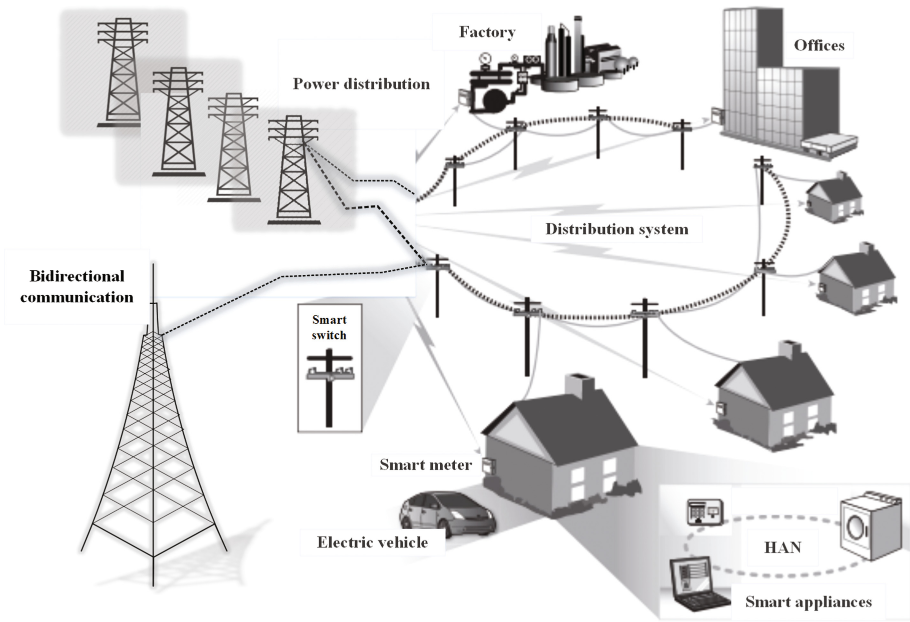


Fig. 2. Smart grid components

The classification of appliances is divided into two categories; non shiftable appliances and shiftable appliances according to the user requirement and their individual requirements. The former category includes those appliances which have fixed power consumption rate and operational time such as hob/oven and heater. Moreover, the latter category is further divided into two categories; Time shiftable appliances and power shiftable appliances.

The rest of the paper is organized as follows: Sect. 2 discussed the related work. Sect. 3 presents the system proposed model and Sect. 4 illustrate the simulations and results. Finally, conclusion and future work are discussed in Sect. 5.

2 Related Work

Researchers has been proposed various methods have for optimal scheduling in HEMS. These optimization techniques are mostly used to find an optimal solution. Scheduling of appliances with the goal of reducing cost and maximizing UC is achieved using GA is [7,12]. In article [12], a real-time HEMS system with scheduler is proposed, that is, using Inclining block rates (IBR) in combination with RTP to control the instability of the system or blackouts. This proposed hybrid pricing scheme achieved better results and satisfy all benefits for both residential and utility company. Similarly a comparative analysis of GA, BPS and ACO techniques is also made in [7]. The pattern of Scheduling scheme (SS) by using GA with consideration of RES and Battery storage system (BSS) is focused. The objective is designed to minimized the cost and maximize the UC. The use of RES reduced the load on utility and provides the opportunity to the consumer to store their produced electricity for use in critical time and also can sell it back to utility in case of excess energy. Although the simulations shows better results in the other schemes as compared to GA in the perspective of PAR, UC and cost, however the execution time complexity is not considered also the ideas of using BSS could not be effective in real life scenarios because of its high installation and maintenance cost.

Power trading and scheduling of load along with integrating RES is achieved using Dynamic programming (DP) [10]. DP is used to schedule the appliance according to its required time. Game theoretic (GT) approach is used for the excess power generation which provide facility for users that sell out their excess power generation, they can offered that too local customers at lower price than the utility price of the company. However the initial installation cost of RES is not ignored.

In DSM in SG, shifting load to off-peak hours from on-peak hours, minimizing PAR, cost reduction and increase in UC is achieved in [4,5], using Hybrid genetic wind driven (HGWD), knapsack problem - WDO (K-WDO) hybrid techniques. The single and 50 homes cases are considered in HEMS [4]. They merged GA with WD for optimized results. All initial steps of WD are used whereas the velocity uploading for overall air pressure value is changed to crossover and mutation of GA. The results show that proposed algorithm outperformed as compared to GA BF WD and BPS in the perspective of cost, Energy consumption (EC), UC and PAR parameters for single home. The paper could not present the comparative analysis of results of proposed schemes with other optimization schemes on the defined parameter for fifty homes. The use RES could further optimize the results up to maximum level. According to [8], an optimization model of HEMS to optimally schedule the residential power loads along with maximum UC. Electrical appliances are classified into three categories based on EC and UC. The WDO algorithm is used for maximizing the UC along with minimizing the cost with TOU in off-peak hours and on-peak hours. Likewise KWDO is used for PAR reduction. Moreover, RES is incorporated during critical hours to avoid system instability, electricity cost reduction and UC. This solution outperformed after extensive simulations as compared to other techniques in perspective of fast convergence rate. However the explicit fitness (pressure) values reduce the overall performance.

An optimal load scheduling is achieved amongst appliance utility and cost effectiveness in [6,11] using BPSO with TOU. An optimal approach for scheduling in residential area based on the TOU pricing scheme is elaborated in [11]. Power users are divided based on criteria of scheduling of the appliances, that is, Traditional users, Smart users, Smart Prosumers. The targeted area is HEMS, and the achievement gained is Cost minimization. However, UC and RES are compromising. In [6], to reduce user discomfort in the pricing scheme a Realistic scheduling mechanism (RSM) is proposed. This is achieved by classifying the appliances according to their preferences. BPSO algorithm is also used for scheduling the appliances optimally. A balanced load scheduling is achieved for all appliances with their operational time, utility and cost effectiveness.

Dynamic scheduling scheme (DSS) is proposed that from the idea of optimal range selection to create a customers EC history that is called a Weighted graph (WG) [3]. The scheme can provide a possible need of power load closely to optimal load by using the WG. Cost, peak-demand and demand variation are the different metrics that are used to measure the achievements of the proposed scheme. Simulation results showed that the improved performance of proposed Dynamic demand scheduling (DDS) scheme as compared to the other schemes.

In [9], the framework of decentralized system proposed DR for the HEMS to reduce electricity cost, increase the UC and security. In this approach, user's Smart meters integrates Home load management (HLM) for exchanging the load related information. This process is repeated until no more optimal solution is achieved. Results are measured after extensive simulations. Fast convergence and optimal scheduling are achieved without compromising UC and cost. The proposed approach converges after few seconds independently of input size.

Energy efficiency for SG and Smart home security (SHS), [3] represented a case study that explains the demand of electricity in rural area of Colombia - Narino. In future the derived model from categorical and non-categorical regressions can be used for electricity generation, consumption minimization strategies, energy substitution strategies, especially for RES. Another main problem focused in a paper [13] is, some appliances required a fixed power rating, which means that once the operational time is started, it has to work with same power rating throughout until the end. The appliances are categorized into non-shift able, time shift able and power shift able appliances based on specifics requirements. The proposed solution can be applied in residential environment and local area. In residential environments the HEMS unit makes the scheduling for all appliances whereas central control unit is used to make the scheduling in local area. The proposed mechanism balanced the load scheduling in 24 h and improves the performance of the power grid by reducing the PAR that is PAR is just 19% higher as compared to daily average. However the UC and RES are not considered. In [5], mainly EC is focused. The Distributed energy resources (DER) is integrated with smart house that provides the major effect on the power consumption cost efficiency. A new concept is proposed that is the cost efficiency. The customer can optimize the consumption scheduling since cost efficiency is a marker to amend its consumption pattern. For providing the better understanding of their consumption history to the customers, the cost efficiencies of different residential appliances consumption behavior and examine the effect of the load shifting according to cost efficiency. A novel power scheduling scheme is developed to optimize the cost efficiency by using the advanced fractional programming tools. The DER are also integrated in algorithm design and optimization is achieved that proposed algorithm is well effective in the improvement of the consumption cost efficiency. Significant results are achieved of the consumption cost efficiency by DERs. Overall in this paper, cost efficiency can be improved by optimal scheduling since indicator of previous consumption behavior is provided. However, by simulation cost minimization is not considered.

3 Proposed Scheme

In SG, EMS enables the end users to communicate with utility, i.e. provides bi-directional communication. Smart homes are equipped with smart meters having EMCs to provide the more trustworthy grid operations. Consumers are able to advise about their preferred pattern of energy consumption to utility. On the other hand, HAN and scheduler are used to get requirement information and power consumption pattern of all appliances to provide the optimal scheduling to customer. One of the main function of DSM is to control demand side activities for customers and motivate them to minimize their cost by shifting their load from on peak hours to off peak hours. Reliable networks are needed for the communication. Wireless sensor networks (WSN) is a promising technology for residential communication system. WSN combines communication and sensing for effective communication [3]. This paper considered residential area, with a home and p number of appliances as shown in Fig. 3. These appliances are divided into different categories based on the preferred schedule. Here first the energy consumption model is presented for p appliances.

In this paper, crow search algorithm (CSA) is used for efficient energy management, which is a user friendly technique, implemented in a simple way to get optimal results by comparing them with EDE. This technique is based on the intelligence of crows. CSA is a population based technique which mainly works

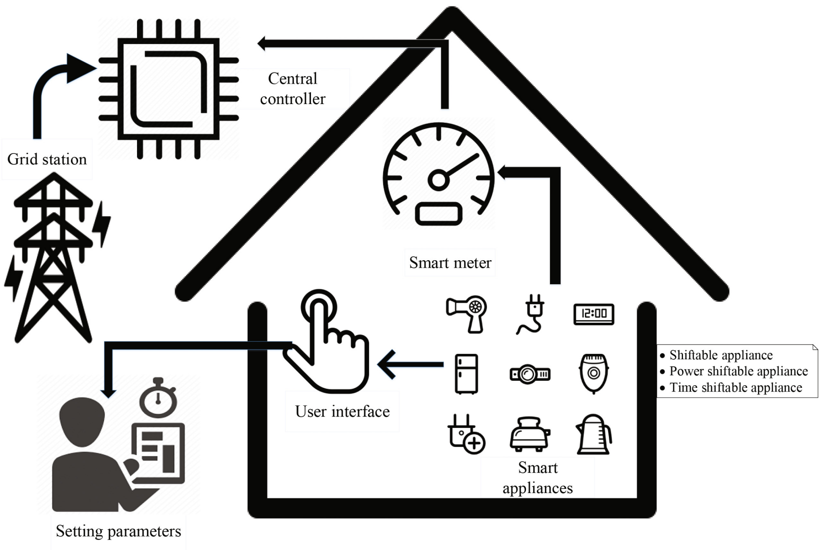


Fig. 3. System model

on the concept of hiding excess food and use it in the future when needed. Crows are considered to be the most intelligent animals among the world. Storing their excessive food in their hiding places and memorizing that is the key phenomenon used here for excessive energy storage and used that information for further use. Apart from it, obtaining best possible hiding place is also the feature used for getting optimal power scheduling. Crows are the greedy creatures they always try to find the hiding place of other birds, and if any crow finds other crows following it, they will acquire a new position. In our case here, the crows are to be acted as a searcher, the whole environment is the sample search space in which we have to find the optimal solution. Where the searching of hiding place is the fitness criteria. CSA tries to get the energy optimization by the natural behavior of crows.

3.1 Energy Consumption Model

Let the set of appliances used here for single home is A, whereas p is the total number of appliances such that A = {a1, a2, a3, ap}. Similarly, let t be the single static time slot used for scheduling of a day T, such that $t \subset T$, where T = {1, 2, 3,..., N}. Where N is 24 in our case. The appliances are divided into three different categories according to its required load and user preferences to get the optimal results of appliance scheduling. Therefore, p number of appliances are categorized into non-shiftable, power shiftable and time shiftable appliances. The details are given in upcoming sections.

3.2 Load Categorization

Based on the collected information by scheduler, all appliances as shown in Table 1 with power consumption are scheduled for 24 h.

Table 1. Classification of appliances

Name	Type	Operational time period	Power rating
Hob and oven	Non shiftable	7 pm–8 pm	1 kWh
Heater	Non shiftable	9 pm–10 pm, 3 am–5 am	1 kWh
Fridge and freezer	Non shiftable	24 h	0.12 kWh
Water boiler	Power shiftable	Need to complete the required energy consumption	0–1.5 kWh (hourly consumption) Requirement: 3 kWh
Electric vehicle	Power shiftable	Preferred hours: (8 pm–8 am)	0.1 kW–3 kW (charging power) Requirement: 5 kWh
Washing machine	Time shiftable	2 h	1 kWh for the 1st hour 0.5 kWh for the 2nd hour
Dishwasher	Time shiftable	1 h	0.8 kWh

3.2.1 Non Shiftable Appliances

For non shiftable appliances Ns, which require non interruptible operational time and power consumption. The scheduling will ensure the continuous power consumption for required operational time.

3.2.2 Time Shiftable Appliances

The scheduling is mainly carried out for shiftable appliances, such as washing machine, which are also known as burst load appliances; the smart meter will control and ensure the flexible time and the required power rating.

3.2.3 Power Shiftable Appliances

For power shiftable appliances, such as water boiler and electric vehicle. The scheduler will schedule elastic power rating in preferred time periods and ensure the total power supply.

3.3 Scheduling Techniques

The main objectives are to reduce the PAR, electricity cost of the customer by load shifting from on peak hours to off peak hours and increasing the user comfort. Energy is optimized by using EDE and CSA algorithms. The results of both schemes are better than unscheduled load as shown in the upcoming section. The brief explanation of CSA is described as follows.

3.3.1 CSA

Crows are known as the most intelligent birds among the animals. They have largest brain as compared to their body. They are using a greedy strategy for accessing food and a better hiding place for storing their excessive food. Another characteristic of crows is they remember other birds by face and alert their flock in case of any danger. Crows are also known by their thief behavior. They watch other birds and their hiding places of storing food and steal it once owner leaves. In our paper, based on above mentioned characteristics of crows the energy is optimized.

4 Simulations and Results

In this section the simulation results of underlying meta-heuristic techniques are discussed. We used EDE and CSA for energy optimization in HEMS. The simulations are done in MATLAB and both of these techniques are evaluated on the basis of performance parameters such as PAR, electricity cost, total energy consumption/load and waiting time of appliances. The input parameters used are 7 appliances divided into three categories; non shiftable appliances, power shiftable appliances and time shiftable appliances on the basis of user requirement and the appliances power consumption requirements etc. In non shiftable appliances such as hob/oven, heater and freezer/refrigerator needs uninterruptable operational time along with fixed power consumption rate, so these appliances can be

Algorithm 1. CSA

Randomly initialize the position of a flock of N crows in the search space
Evaluate the position of the crows
Initialize the memory of each crow
while $iter \leq iter^{max}$ **do**
  for $i = 1N(forallNcrows)$ **do**
    randomly get a crow j to follow i
    Define awareness probability
    if $r^j \geq AP^{j,iter}$ **then**
      $x^{i,iter_1} = x^{i,iter} + r^i X(m^{j,iter} - x^{i,iter})$
    else
      $x^{i,iter+1} = arandompositionofsearchspace$
    end if
  end for
  Check the feasibility of new positions
  Evaluate the new position of the crows
  Update the memory of crows
end while

scheduled. The shiftable appliances are further categorized into power shiftable appliances and time shiftable appliance. The flexible time scheduling can be used for time shiftable appliances such as washing machine and dish washer, however it needs fixed power consumption. On the other hand, flexible power rate can be used for power shiftable appliances such as electric vehicle and water boiler however it need to be scheduled within the preferred time period of the user. Moreover, 24 h scheduling is done on the basis of 6 pm to 10 pm hours are considered as on peak hours in the evening. Similarly, 7 am to 10 am are considered as on peak hours in the morning according to the most accepted statistics. For load scheduling some load of on-peak hours are shifted is to off peak hours to get optimal performance parameters results. In this paper the RTP pricing scheme is used for electricity cost calculation. The results of performance parameters are discussed in detail along with their graphical representation achieved during simulations in next section.

The Fig. 4 shows, the hourly load distribution of unscheduled, CSA and EDE in 24 time slots. The load is optimally distributed in case of both meta-heuristic techniques EDE and CSA as compared to the unscheduled. Moreover, 80% and 85% maximum load during on-peak hours is abridged in CSA and EDE respectively. It is clear from the graph that our proposed technique CSA distributed the load optimally and beats the EDE scheduling by 5%.

The Fig. 5 shows the PAR achieved from simulation results for CSA and EDE along with unscheduled. It is clearly shown in the graph that PAR of unscheduled is relatively higher than CSA and EDE. Since both of the techniques reduced the PAR by optimal load scheduling. However, CSA beats the EDE by 8%.

The Fig. 6 expresses the electricity cost for unscheduled along with CSA and EDE scheduling for 24 time slots. We have used RTP price signal for electricity bill calculation. The electricity cost for unscheduled is comparatively higher than

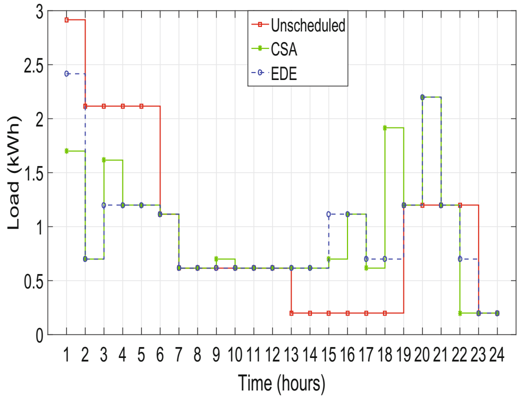


Fig. 4. Load graph

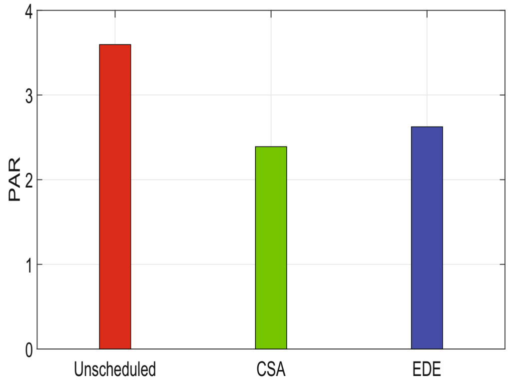


Fig. 5. Peak to average ratio

CSA and EDE. It is clearly noticeable in the figure that CSA get ahead having minimum cost as compared EDE. Since the load is shifted from on-peak hours to off-peak hours in scheduling which consequently reduced the electricity cost. Moreover, this will increase the waiting time or disturb the UC as shown in Fig. 7. Because there exists a trade-off between waiting time and electricity cost. The waiting time can be describe as the time the user preferred time of an appliance lags behind to its scheduled time. However, if user comfort cannot be compromised at any cost then electricity cost will be increase consequently. Both meta-heuristic techniques, scheduled the appliance in an optimal way to pay compensation to both UC and electricity cost.

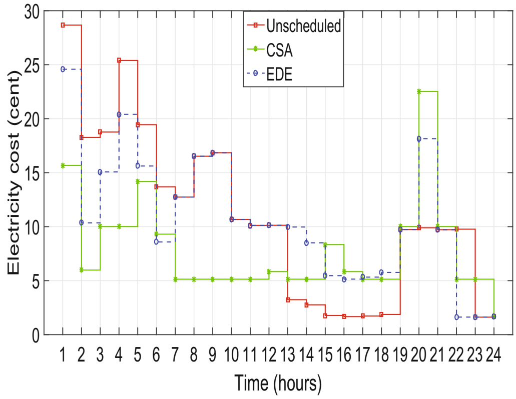


Fig. 6. Electricity cost

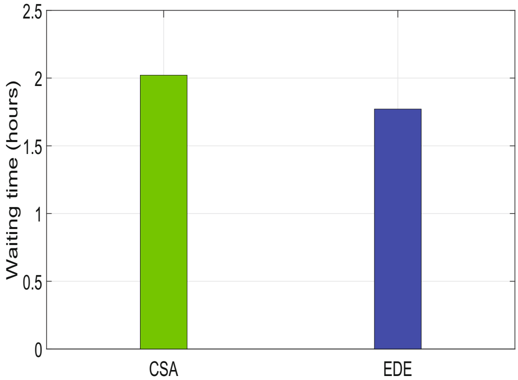


Fig. 7. Waiting time

5 Conclusion and Future Work

In this paper, we evaluated residential area energy management for smart home for various appliances. Appliances are categorized based on their energy consumption requirements and users' lifestyle. The performances of meta-heuristics techniques are based on performance parameters such as reduction in PAR, cost minimization, UC maximization, and load shifting from on peak hours to off peak hours. The results show that CSA outperformed as comparable to EDE and unscheduled. However, there is a trade-off between waiting time and load minimization. In CSA the load is minimized along with cost, PAR reduction as compared to the results of EDE. Moreover, the waiting time is reduced as compared to CSA. In future, RES will be integrated to get better results.

References

1. International Energy Outlook 2016: Executive Summary - Energy Information Administration
2. Palensky, P., Dietrich, D.: Demand side management: demand response, intelligent energy systems, and smart loads. IEEE Trans. Ind. Inf. **7**(3), 381–388 (2011)
3. Wilson, O., Achicanoy, M., Jimenez, J.B.: Electricity demand modeling for rural residential housing: a case study in Colombia. In: 2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM), pp. 614–618. IEEE (2015)
4. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Ma, J., Chen, H., Song, L., Li, Y.: Residential load scheduling in smart grid: a cost efficiency perspective. IEEE Trans. Smart Grid, 1 (2015)
6. Mahmood, D., Javaid, N., Alrajeh, N., Khan, Z., Qasim, U., Ahmed, I., Ilahi, M.: Realistic scheduling mechanism for smart homes. Energies **9**(3), 202 (2016)
7. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
8. Rasheed, M., Javaid, N., Ahmad, A., Khan, Z., Qasim, U., Alrajeh, N.: An efficient power scheduling scheme for residential load management in smart homes. Appl. Sci. **5**(4), 1134–1163 (2015)
9. Safdarian, A., Fotuhi-Firuzabad, M., Lehtonen, M.: Optimal residential load management in smart grids: a decentralized framework. IEEE Trans. Smart Grid **7**(4), 1836–1845 (2016)
10. Samadi, P., Wong, V., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
11. Ullah, I., Javaid, N., Khan, Z.A., Qasim, U., Khan, Z.A., Mehmood, S.A.: An incentive-based optimal energy consumption scheduling algorithm for residential users. Procedia Comput. Sci. **52**, 851–857 (2015)
12. Zhao, Z., Lee, W.C., Shin, Y., Song, K.-B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
13. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–5. IEEE (2012)

Privacy Preserving Data Mining on Big Data Computing Platform: Trends and Future

Gao Zhiqiang(✉) and Zhang Longjun

Department of Information Engineering, Engineering University of PAP, Xi'an, Shaanxi, China
1090398464@qq.com, 15891741749@163.com

Abstract. Data mining is becoming increasingly important in the data-driven society in recent years. Unfortunately, privacy of the individuals fails to be protected and considered deliberately. It's a significantly challenging question that outputs of data mining models can be applied to preserve privacy while simultaneously maintaining analyzing capability. With advancements in big data, series of big data computing platforms have evolved into widely utilized paradigms for data mining. However, users' sensitive data which are outsourced on the cloud and mined on open-sourced computing platform. It poses such severe threats that measures must be taken to protect the privacy of individuals' data. Regarding this issue, much fruitful work has been done on designing privacy preserving data mining approaches for improving big data computing platform security and privacy of individuals. In this paper, a systematic investigation of a wide array of the state-of-the-art privacy preserving data mining (PPDM) techniques has been performed from different aspects on threat model, anonymity, secure multiparty computation (SMC), differential privacy. We are focused on improving data privacy in these sensitive areas on big data computing platforms. Hopefully, our work aims to highlight the urgent need for applying privacy preserving data mining approaches on big data computing platforms. Moreover, a better understanding of this research area may benefit the usage of big data and future exploration.

Keywords: Big Data · Spark · MapReduce · Data mining · Privacy preserving · Differential privacy

1 Introduction

With the advent of Big Data and Internet+, multi-sourced data stored in various distributed forms are ubiquitous and accessible ranging from different application domains such as digital devices to Internet of things at a extremely high speed. According to the statistic report of International Data Corporation (IDC), the total of amount of data has reached 1.8 ZB in 2011 [1] and will be doubled every two years. There will be a total of 35 ZB in 2020, which will nearly increase 20 times. However, This booming increasing phenomenon has caused extensive concerns and challenges all around the world in the past few years.

Historically, Jim Gray, the winner of Turing Award, is the pioneer of Big Data analyzing and propose the fourth paradigm in scientific research. In 2008, "Nature"

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_44

launched a special issue on a Big Data. In succession, "Science" also launched a similar special issue on data processing in 2011 [2]. Notably, the storm of Big Data is sweeping all over the IT industry, in which IT giants such as Google, Amazon and Baidu all have established their own Big Data computing and storage platform. At the same time, governments all over the world also pay unprecedented attention to Big Data. In 2012, the U.S. government invested $200 million to start "Big Data Research and Development Program". In 2015, our State Council approved the "China's Big Data Development Program" aimed at comprehensively promoting the development and application of Big Data.

It is becoming increasingly important that mining individual's information in the data-value-driven society. However, the development of Big Data is still facing many problems at present, where security of individual's private data is one of the key problems that people generally concern [3, 4]. Especially, the security of analysis and processing on massive data has captured increasing attention globally. In particular, the original users' data may be intercepted or damaged during the process of data mining by malicious attackers. In the situation of social networks, data privacy security is seriously threatened. Moreover, traditional data mining methods can not satisfy the growing demand of Internet and overwhelming complexity of cloud services. Therefore, in the era of big data, privacy-preserving data mining (PPDM) on Big Data computing platform has become a new attractive and promising area of cross disciplinary research since it was introduced.

Scholars home and abroad have done a lot of fruitful research work in recent years. Yet, existing solutions adopt either random disturbance (e.g. [5, 6]) or cryptographic techniques (e.g. [7–11]) to protect privacy. For example, [12] built three new systems that consist of hiding small details in pictures by rotating some pixels to deal with the problem of image perturbation for privacy preserving. The focus of [13] is on exploring private computation about vector addition and its applications in privacy-preserving data mining, which supports a large number of popular data-mining algorithms verifying zero-knowledge (ZK) protocols for user data. According to [9], we note that cryptography provides primitives with various levels of efficiency. However, [14] points out that the assumption that passive adversary never deviates from specified behavior is clearly not realistic when the data from individual users are corrupted by hackers. Except for sacrificing accuracy, randomization has been proved to provide very little privacy protection in many cases unless sufficient amount of noise is introduced. While classical cryptographic approaches typically fall into the following two major drawbacks: (1) not adequate for large-scale systems, especially in the Big Data situation due to complex operation of expensive cryptographic performance, and (2) no effective countering mechanism to cope with actively cheating attackers. Regarding the issues above, we note that implementing data mining algorithms using a general approach to parallelize the models and run them in a distributed fashion may be beneficial. For example, [15, 16] showed that many popular data mining models have a aggregated performance that can be executed in form of MapReduce framework. It is a widely used distributed programming over large amounts of distributed clusters that is being successfully deployed in application. In this paper, we make an systematic survey on privacy-preserving data mining on Big Data computing platform. Our contributions can be concluded as follows:

- We provide a detailed review of classical scalable and privacy-preserving data mining approaches and strategies that can modify the weaknesses of corresponding models. Especially, we illustrate a proof-of-concept data privacy preserving method, differential privacy in detail.
- we introduce the applicability of existing state-of-the-art efficient computing platforms, using Apache Hadoop and Spark as concrete examples. As for Spark, its solution goes beyond a simple parallelization with less I/O communication with HDFS. In addition, two main privacy-preserving MapReduce-based scheme for security are presented.
- We summarize the trends and directions of privacy preserving data mining on Big Data computing platform, as well as point out the problem of privacy-preserving data mining.

The reminder of this paper is organized as follows. In Sect. 2, preliminaries and related definition are illustrated. Then, we provide a review of privacy-preserving data mining approaches in Sect. 3. In Sect. 4, we demonstrate the state-of-the-art efficient computing platforms, Apache Hadoop and Spark. Section 5, two main privacy-preserving MapReduce-based scheme for security are presented in detail. Section 6 points out the trends and directions of privacy preserving data mining on Big Data computing platform. Section 7 concludes the paper.

2 Preliminaries

2.1 Anonymity

k-anonymity [20] is one of the most widely used approaches, However, it does not consider the distribution of sensitive values in all records. In succession, two varieties were developed to cope with this shortcoming. *l*-diversity [21] and *t*-closeness [22] are the two examples (Table 1).

Table 1. Approaches of anonymity

Method	Definition	Drawback
k-anonymity	Each record in the published data set needs to be indistinguishable from at least $k - 1$	The distribution of sensitive values is not considered when all records in an equivalence class contain the same sensitive value
l-diversity	Every equivalence class should contain *l* well-represented values	Difficult to determine variable *l*
t-closeness	Distance between the distribution of sensitive values in an equivalence class and the whole data should differ by no more than a threshold *t*	Distribution of sensitive values needs to be provided

2.2 Data Perturbation

Except for the solutions of privacy preserving above, we can build a centralized warehouse to perform distributed data mining which can be performed with provably data privacy-preserving. On the other hand, data perturbation is also a satisfied approach, which can rectify the original data so that malicious attackers can not identify the individual even with the background knowledge they obtain. Basic perturbation techniques include data swapping [23], randomization [24], etc. (Table 2).

Table 2. Approaches of data perturbation

Method	Definition	Feature
Data swapping	Combinations of attributes that are effective at predicting a target class value	Limited swapping has a minimal impact on the results
Randomization	Adding noise to hide actual data values	Construct models that generalize the data

2.3 Secure Multiparty Computation

As a traditional privacy-preserving tool which is suitable in the distributed mode, secure multiparty computation (SMC) can protect participants obeying SMC protocols. Table 3 is the general outline of SMC.

Table 3. Approaches and history of secure multiparty computation [25–27]

Method	Application	Author	Idea
Secure two-party computation	Privacy-preserving, distributed data mining	Andrew Yao (1986)	Based on the premise that every piece of private information is validly known to one or more parties
Security multiparty computation		Oded Goldreich, Matt Franklin and Modi	Any function with inputs distributed between two parties could be securely computed without revealing any information to any party other than its local input and output
Excellent survey		Yung	Provide an excellent survey of secure distributed computation

2.4 Differential Privacy

Differential privacy [28] is a rigorous privacy guarantee on publishing statistical information when considering statistical control of the privacy of individual. Additionally, differential privacy is implemented by adding noise into data sets to achieve the effect of privacy protection and overcome the shortcoming of the traditional security model. In statistics view, we can define differential privacy as follows: In a statistical database is composed of a set of rows or tuples, when distance between databases D and D_0 is 1, they are neighbors or adjacent. That is, they differ by only a record at most.

Definition 1. If for all pairs of adjacent databases D and D_0 randomized function $\boldsymbol{K}$ is (ε, δ)-differential privacy,

$$\Pr[\boldsymbol{K}(D_0) \subseteq Range(\boldsymbol{K})] \leq e^{\varepsilon} \cdot \Pr[\boldsymbol{K}(D) \subseteq Range(\boldsymbol{K})] + \delta \tag{1}$$

Where all $S \subseteq Range(\boldsymbol{K})$ and when $\delta = 0$, function $\boldsymbol{K}$ is ε-differential privacy.

Specifically, global sensitivity of query function $\boldsymbol{F}$ can be used to measure the impact of change on the query output from a single record.

Definition 2. L_1-sensitivity of query function $\boldsymbol{F}$: $D \rightarrow R^d$ can be defined as follows, where D and D_0 are neighbors.

$$\Delta \boldsymbol{F} = \max_{D, D_0} \| \boldsymbol{F}(D) - \boldsymbol{F}(D_0) \|_1 \tag{2}$$

In addition, *Laplace* mechanism is one of the most widely accepted realization of ε-differential privacy, (3) gives the probability density function of *Laplace*:

$$\boldsymbol{Lap}(x) = \frac{1}{2\Delta \boldsymbol{F}/\varepsilon} exp(\frac{|x|}{\Delta \boldsymbol{F}/\varepsilon}) \tag{3}$$

Moreover, we can compose a sequence of ε_1, ε_2,…, ε_m-differential privacy mechanism to satisfy $\varepsilon = \sum_{i=1}^{m} \varepsilon_i$-differential privacy, where ε denotes privacy budget in privacy-preserving approaches. When data mining tasks can be executed in multiple steps, each step just need to apply a partition of ε. Guaranteed by composable mechanism, the sum of privacy budget ε can be assigned reasonably.

3 Privacy-Preserving Data Mining Approaches

3.1 Privacy-Preserving Decision Tree

As we all know, privacy of data can be leaked via the process of describing tree nodes while building a decision tree. Especially, when decision tree is making decision on how to split each node with attributes. That is, publishing the information of leaf node can be an neglectable way of privacy breach. Notably, implementing decision tree algorithm with differential privacy is a powerful and potential method to prevent privacy breaches [29].

On the other hand, the integration of a certain number of random decision trees, namely, Multiple Trees may be a good classifier (i.e., a decision forest). Additionally, differential privacy decision forests which are built greedily and randomly using random subsets and bootstrapped samples, are beneficial in reducing over-fitting.

However, it remains an open question in optimizing multiple differential privacy decision trees with lower cost of privacy budget division unfortunately when trees are joint in the private scenario.

3.2 Privacy-Preserving *k*-means

Clustering [11] is a useful data mining tool and *k*-means has advantage on dividing data into *k* clusters according to different metrics. In Table 4, we summary the pseudo code of classic *k*-means. Generally speaking, *k*-means can achieve expected output when (1) preset iteration parameters are met or when (2) the similarity in a optimized cluster has reached a tolerable prefixed threshold.

Table 4. Pseudo code for classic *k*-means clustering

Algorithm 1: *k*-means clustering
Input: *k, the number of centroid* ***Output:*** *entities assignments to final clusters*
Select k entities as initial centroid randomly ***Repeat*** *Assign each entity to the nearest centroid* *Recompute the centroid of each cluster* ***Until*** *termination criterion is achieved*

Traditional *k*-means is challenged by two main outstanding problems: (1) clustering quality highly depends on initial centroid selection with poor local minimal. (2) privacy of data to be analyzed is threatened during clustering. Regarding these issues, our early work in our team had proposed a privacy-preserving hybrid *k*-means, which is capable to satisfy differential privacy by combining swarm intelligent in the phase of center point selection. And then in the second phase of clustering performed under Spark, we verified our approach on public UCI big data sets with promising and satisfactory results.

4 Big Data Computing Platform

4.1 Apache Hadoop

Traditional computing architectures have been overwhelmed by Google, Baidu, and other IT giants by the advent of Big Data. Significantly, Apache Hadoop [17] has rapidly emerged as the outstanding flag for managing, computing and storing large volumes of unstructured data in a distributed form. MapReduce (MR) is a powerful computing framework to sort and shuffle between mappers and reducers tasks.

Meanwhile, Hadoop Distributed File System (HDFS), is able to provide high scalability and availability on scaled clusters consisting of inexpensive hardware. NameNode and DataNode are two main components in HDFS, as illustrated in Fig. 1.

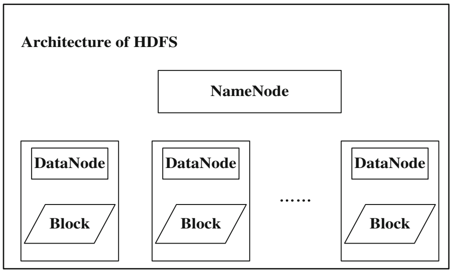


Fig. 1. Architecture of Hadoop Distributed File System

4.2 Apache Spark

Spark [19] is generated by the UC Berkeley RAD Lab in 2009, namely the famous AMPLab later, which is suitable in the situation of interactive queries and iterative in-memory storage with effective fault recovery. From what we have learned, features of Apache Spark can be illustrated as Table 5.

Table 5. Features of Apache Spark

	Feature
Apache Spark	Interoperable, fast enterprise-grade large-scale data processing engine
	Written in Scala, object-oriented, runs in a JVM
	In-memory processing
	Suitable for iterative algorithms through memory
	100 times faster than MR in-memory or 10 times faster on disk
	A stack of libraries, including Spark SQL, MLlib, Graph X, and Spark Streaming
	Support for Java, Scala, Python, and R
	Runs on Hadoop, Mesos, standalone cluster managers, on-premise hardware, or in the cloud

5 Typical Solutions for PPDM on Big Data Computing Platform

5.1 Airavat

Airavat [30, 31] is MapReduce-based platform which is capable of performing untrusted code of data mining, especially *k*-means on original users' sensitive data. Further, aimed at computing and aggregating the input data sets accurately, the output security of Airavat can be guaranteed by differential privacy. Explicitly, the structure of Airavat and privacy preserving mechanism it adopted is depicted as Fig. 2.

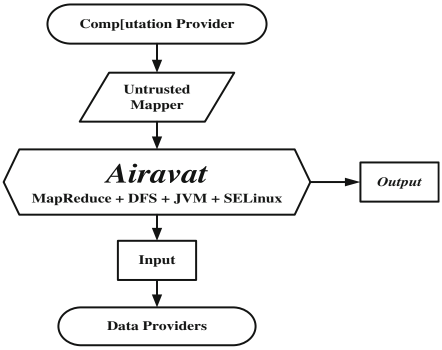


Fig. 2. Key structure of Airavat

5.2 PRISM: Privacy-Preserving Search in Map Reduce

PRISM [32], a privacy-preserving scheme specified for searching engine in cloud environment. Faced with untrusted cloud providers, the main challenge of security scheme is to achieve trade-offs between privacy and efficiency. Solutions range from encrypted keyword search to Private Information Retrieval (PIR). Significantly, PRISM, with the practical idea of transforming word search matters into subset can satisfy the requirement of real-world cloud environment. Further, contributions of PRISM can be summarized as follows in Table 6.

Table 6. Contributions of PRISM

	Contributions
PRISM	Suited to cloud computing
	Preserves privacy in the face of potentially malicious cloud providers
	Compatible to standard Map Reduce
	Provides flexible search
	Not limited to searching for a fixed set of predetermined keywords to be known in advance, but offers flexible search for any words

5.3 Privacy-Preserving Learning Analytics (PPLA)

The architecture of Privacy-preserving Learning Analytics (PPLA) [33] is depicted in Fig. 3. Further, PPLA adopts data publishing and statistical disclosure control as privacy protection mechanisms. Notably, statistical disclosure control is guaranteed by ε-differential privacy.

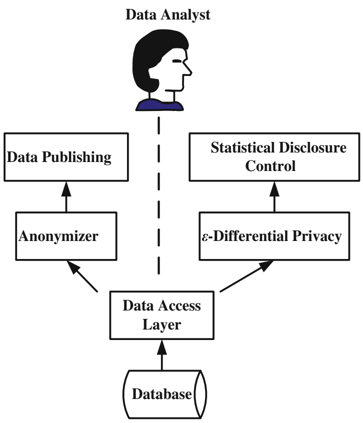


Fig. 3. Architecture of privacy-preserving learning analytics system

6 Future for PPDM on Big Data Computing Platform

Despite that privacy preserving approaches on Big Data computing platform [34] are increasingly accepted, PPDM still is faced with unprecedented challenges. Thus, we summarize challenges, application scenario and future trends of PPDM on Big Data computing platform in Table 7.

Table 7. Challenges, application scenario and future trends

	Challenges
PPDM on Big Data computing platform	The adversary could be an insider
	Anonymization cannot guarantee hiding the fact that some event has happened faced some disciplinary action
	Having multiple records per data subject in a database is complicating, since these records are often correlated
	Continuous and sequential releases based on anonymization are problematic
	Application Scenario
	Surveys and Data Collection
	Monitoring for Emergencies
	Product Traceability, such as manufacturing, packaging, transportation, storage, and sale
	Medical Research
	Social Networks
	Future Trends
	Make PPDM scale and achieve higher accuracy while guarantees privacy
	Proving complexity lower bounds and accuracy upper bounds

(*continued*)

Table 7. (*continued*)

	Secure co-processors form trusted computing can be leveraged to enable privacy-preserving operations over data sets much larger than their storage capacity
	Bridging game theory and PPDM could lay the theoretical foundation for a market of private data
	Fully understand SMC, validate all the assumptions or tailor the protocol to the situation and then reevaluate the proof of protocol's privacy
	Based on secret sharing, new trend in e-commerce is open to the general public at a very low cost, like Amazon
	Protocols does not scale well, namely, real scalability bottleneck
	It can be an interesting area for future work concerning ε-differential privacy can be solved via implementing a readily available, privacy-integrated tool

7 Conclusion

Privacy-preserving data mining is proposed to cope with the tradeoffs between data analysis and guaranteeing the privacy of individual. In this paper, an systematic overview of the PPDM approaches was illustrated, namely, threat model, anonymity, secure multiparty computation (SMC), differential privacy. Each approach was demonstrated in detail from aspect of advantages and disadvantages. Finally, the challenges, application scenarios, future trends of PPDM were outlined. In spite of our comprehensive survey, we are not meant to be conclusive. Instead, it is our objective that our work can spark interest in PPDM on Big Data computing platform with the state-of-the-art privacy protection mechanisms in the near future.

Acknowledgments. We will thank for National Natural Science Foundation funded project 61309008 and Shaanxi Province Natural Science Funded Project 2014JQ8049. Also, we would also like to thank our partners at our Research Lab and their generous gifts in support of this research.

References

1. Chasaki, D., Mansour, C.: Security challenges in the Internet of Things. Int. J. Space Based Situated Comput. **5**, 141–149 (2015)
2. Beldad, A.: Sealing one's online wall off from outsiders: determinants of the use of Facebook's privacy settings among young Dutch users. Int. J. Technol. Hum. Interact. (IJTHI) **12**, 21–34 (2016)
3. Barhamgi, M., Benslimane, D., Ghedira, C.: PPPDM–a privacy-preserving platform for data mashup. Int. J. Grid Util. Comput. **3**, 175–187 (2012)

4. Li, X., He, Y., Niu, B.: An exact and efficient privacy-preserving spatiotemporal matching in mobile social networks. Int. J. Technol. Hum. Interact. (IJTHI) **12**, 36–47 (2016)
5. Petrlic, R., Sekula, S., Sorge, C.: A privacy-friendly architecture for future cloud computing. Int. J. Grid Util. Comput. **4**, 265–277 (2013)
6. Duan, Y., Canny, J.: How to deal with malicious users in privacy-preserving distributed data mining. Stat. Anal. Data Min. **2**, 18–33 (2009)
7. Khan, N., Al-Yasiri, A.: Cloud security threats and techniques to strengthen cloud computing adoption framework. Int. J. Inf. Technol. Web Eng. (IJITWE) **11**, 50–64 (2016)
8. Zhang, W., Jiang, S., Zhu, X.: Cooperative downloading with privacy preservation and access control for value-added services in VANETs. Int. J. Grid Util. Comput. **7**, 50–60 (2016)
9. Almiani, M., Razaque, A., Al, D.A.: Privacy preserving framework to support mobile government services. Int. J. Inf. Technol. Web Eng. (IJITWE) **11**, 65–78 (2016)
10. Yang, Q., Wu, X.: 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Making **5**, 597–604 (2006)
11. Lloyd, S.: Least squares quantization in PCM. IEEE Trans. Inf. Theory **28**, 129–137 (1982)
12. Su, D., Cao, J., Li, N.: Differentially private k-means clustering. In: Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy, pp. 26–37 (2016)
13. Samet, S., Miri, A., Orozco-Barbosa, L.: Privacy preserving k-means clustering in multi-party environment. In: SECRYPT, pp. 381–385 (2016)
14. Doganay, M.C., Pedersen, T.B., Saygin, Y.: Distributed privacy preserving k-means clustering with additive secret sharing. In: Proceedings of the 2008 International Workshop on Privacy and Anonymity in Information Society, pp. 3–11 (2008)
15. Upmanyu, M., Namboodiri, Anoop M., Srinathan, K., Jawahar, C.V.: Efficient privacy preserving k-means clustering. In: Chen, H., Chau, M., Li, S.-h., Urs, S., Srinivasa, S., Wang, G.A. (eds.) PAISI 2010. LNCS, vol. 6122, pp. 154–166. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13601-6_17
16. Chen, H., Hu, Y., Lian, Z.: An additively homomorphic encryption over large message space. Int. J. Inf. Technol. Web Eng. (IJITWE) **10**, 82–102 (2015)
17. Hadoop. http://hadoop.apache.org
18. Mllib. http://spark.apache.org/mllib
19. Spark. http://spark.apache.org
20. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **10**, 571–588 (2002)
21. Machanavajjhala, A., Kifer, D., Gehrke, J.: l-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data (TKDD) **1** (2007)
22. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: privacy beyond k-anonymity and l-diversity. In: IEEE 23rd International Conference on Data Engineering, ICDE 2007, pp. 106–115 (2007)
23. Guo, Y.: Reconstruction-based association rule hiding. In: Proceedings of SIGMOD 2007 Ph. D. Workshop on Innovative Database Research, pp. 51–56 (2007)
24. Verykios, V.S., Pontikakis, E.D., Theodoridis, Y.: Efficient algorithms for distortion and blocking techniques in association rule hiding. Distrib. Parallel Databases **22**, 85–104 (2007)
25. Yao, A.C.: How to generate and exchange secrets. In: 27th Annual Symposium on Foundations of Computer Science, pp. 162–167 (1986)
26. Goldwasser, S., Micali, S., Wigderson, A.: How to play any mental game, or a completeness theorem for protocols with an honest majority. In: Proceedings of the Nineteenth Annual, vol. 87, pp. 218–229 (1987)
27. Franklin, M., Yung, M.: The varieties of secure distributed computation. In: Capocelli, R., De Santis, A., Vaccaro, U. (eds.) Sequences II, pp. 392–417. Springer, New York (1993)

28. Dwork, C.: A firm foundation for private data analysis. Commun. ACM **54**, 86–95 (2011)
29. Fletcher, S., Islam, M.Z.: Decision tree classification with differential privacy: a survey. arXiv preprint arXiv:1611.01919 (2016)
30. Zhou, M., Zhang, R., Xie, W.: Security and privacy in cloud computing: a survey. In: 2010 Sixth International Conference on Semantics Knowledge and Grid (SKG), pp. 105–112 (2010)
31. Roy, I., Setty, S.T., Kilzer, A.: Airavat: security and privacy for MapReduce. In: NSDI, pp. 297–312 (2010)
32. Blass, E.O., Di Pietro, R., Molva, R., Önen, M.: PRISM – Privacy-Preserving Search in MapReduce. In: Fischer-Hübner, S., Wright, M. (eds.) PETS 2012. LNCS, vol. 7384, pp. 180–200. Springer, Heidelberg (2012)
33. Gursoy, M., Inan, A., Nergiz, M.E.: Privacy-preserving learning analytics: challenges and techniques. IEEE Trans. Learn. Technol. (2016)
34. Kong, W., Lei, Y., Ma, J.: Virtual machine resource scheduling algorithm for cloud computing based on auction mechanism. Optik-Int. J. Light Electron Opt. **127**, 5099–5104 (2016)

The 3rd International Workshop for Collaborative e-Business Systems (IWCBS-2017)

Pigeon Inspired Optimization and Enhanced Differential Evolution in Smart Grid Using Critical Peak Pricing

Zunaira Amjad, Saadia Batool, Hafsa Arshad, Komal Parvez, Mashab Farooqi, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. In this paper, we have evaluated the performance of heuristic algorithms; Enhanced Differential Evolutionary (EDE) and Pigeon Inspired Optimization(PIO) for Demand Side Management (DSM). Moreover, Critical Peak Pricing (CPP) is used as a price traffic. The main purpose of this paper is to reduce Peak to Average Ratio (PAR) and electricity cost by scheduling appliances according to categories and constraints. Simulation results demonstrate that PIO outperforms in terms of user comfort.

1 Introduction

Electricity is becoming an integral part of our lives. A reliable supply of energy is required in residential and commercial areas to satisfy their needs, as well as to improve lifestyle. Demand for electricity is increasing day by day and expected to be double in coming years 2020–2030 [1]. It is also expected that total demand of electricity will increase by 75% at the end of 2020 in comparison to 2000 [2]. This rapid increase in energy demands, enforces the utility to think about electricity generation and distribution; to avoid electricity shortage and unexpected challenges. The utility, thus try to fulfill the demand with smart generation of electricity. For this purpose, traditional grids are converted into Smart Grid (SG) [3].

In SG, customers have the flexibility to interact with the utility using advanced information system and technology [4]. SG allows the energy generation to eliminate effects of CO2 on environment and reduces the consumption of electricity [5,6].

In several cases, utility company has to bear the price fluctuations for electricity and consumers are charged fixed prices. As consumers are not aware of wholesale rates of electricity, they show fluctuations in demand of electricity and complete load is shifted in off peak hours, peaks are being generated. These fluctuations effect the system efficiency and decrease the utility profit. Nowadays

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_45

countries have decided to restructure their industry. Therefore, established countries use the Demand Side Management (DSM) to control peak-load-reduction and energy consumption patterns [7–10].

According to [11–18] Energy consumption patterns can be controlled by using Dynamic Pricing Signals (DPS) and different techniques in DSM. Different heuristic algorithms have been proposed in literature as they result in controlling energy patterns, electric bill reduction, providing user comfort and peak load reduction by creating Peak to Average Ratio (PAR).

In this paper, we present two different heuristic techniques to minimize cost, PAR and enhancing the user comfort by applying DPS. These signals help the utility to maintain the electric price rates and communicate with user to control the energy consumption patterns. As the author of paper [19], proposed Integer linear Programming (ILP) technique to schedule the appliances on basis of their power rate and time. However, cost reduction and PAR were ignored by the author. Section 1 gives a brief introduction of this research work. Some related work in this field have been discussed in Sect. 2. Section 3 explain the system model of smart meter. However Sects. 4 and 5 deal with proposed technique and simulation results respectively. In last section conclusion and futile work of applied techniques have been elaborated.

2 Related Work

Many researchers schedule appliances or a single appliance which is on pre-established demand charts. Most of the articles from Smart grid domain speak with dealing problems of load scheduling, power consumption, cost reduction, minimizing PAR and User Comfort (UC) in smart grid. A brief view of changing trends of research in SG are discussed here.

The authors of paper [19, 20, 22, 30] deal with problem of scheduling the load from Peak Hours (PHs) to Off Peak Hours (OPHs). These authors propose some efficient approaches i.e. Integer Linear Programming (ILP), Dynamic Programming (DP) and Linear Programming (LP), Hybrid Generic-Wind Driven (GWD) Algorithm and Binary Particle Swarm Optimization (BPSO) respectively to deal with the problem. Cost minimization, reduction of energy expenses and maximizing the UC are the major achievements of paper [20, 22, 30]. However, author of paper [19] only deals with load scheduling as non-adaptability of Dynamic Pricing Models (DPM) and lack of Renewable Energy Sources (RES) integration are the major drawbacks of his technique. On Contrary, due to trade-off between UC and cost, prior is compromised in other techniques.

In research articles [23, 25, 27], optimization of energy is the major concern. DPM, Mixed Integer Non-Linear Programming (MINLP) and optimal scheduling techniques are proposed to schedule the power consumption. The objectives of these techniques include: to encourage the consumer to reduce the peak load and the desired trade-off between UC and cost. However, multiple users are not considered in DSM applying proposed techniques.

General structure of energy management systems is proposed in a residential area which is based on smart grid. Authors proposed efficient techniques to minimize the cost and PAR. Peak formation must be avoided by strengthening the complete electricity system. An objective function is used to solve these problems in an optimized way. These kind of optimization problems are usually non-linear, hence the authors use Genetic Algorithm (GA) to solve their problem. The simulation results of paper [21,29] shows that using Real Time Pricing (RTP) combined with Inclined Block Rate (IBR) pricing model reduces cost and PAR more effectively. However, M.B. Rasheed in [28] proposed Intelligent Programmable Communication Thermostat (IPCT) and Conventional Programmable Communication Thermostat (CPCT) by using GA; to minimize PAR and electricity cost.

In research article [26], authors proposed scheduling mechanism to reduce end user frustration, intensifies UC and minimizing electricity cost. For optimal operational time of different appliances and power usage optimization, heuristic algorithm BPSO is used. Simulations results prove that proposed scheme is cost effective and maintains an equal ratio between appliance utility and cost effectiveness up to user demand. Integration of renewable energy source management is major drawback of the proposed technique.

Ant Colony Optimization (ACO), GA and BPSO are proposed by the author by which performance evaluation of Home Energy Management System (HEMS) is carried out in paper [24]. Problem formulation is carried out via multiple knapsack problem, i.e., minimize the electricity bill, minimize the aggregated power consumption, minimize PAR, maximize UC and efficient integration of RESs. Time of Use (ToU) and IBR is used for cost estimation. This technique results in cost effective solution to increase sustainability. Reduction of cost and PAR whereas user UC maximization are objectives of this research. However, due to trade-off between cost and user comfort, maximizing the user comfort also increases the cost.

3 System Model

Figure 1, depicts the structure of Energy Management System (EMS). As shown in Fig. 1 smart meter is the basic component of EMS. It connects with user interface to collect the user requirements and preferences for scheduling appliances. Whereas, Smart Meter directly connect with appliances to provide them electricity, as well as to obtain the energy consumption schedule for each appliance. On the basis of all the information obtained using smart meter will schedule the appliance into three categories:

1. Non-Shiftable Appliances
2. Time Shiftable Appliances
3. Power Shiftable Appliances

The system described here is used for a single home. However, it can be extended further to multiple users where different smart meters are connected to achieve a scheduling mechanism.

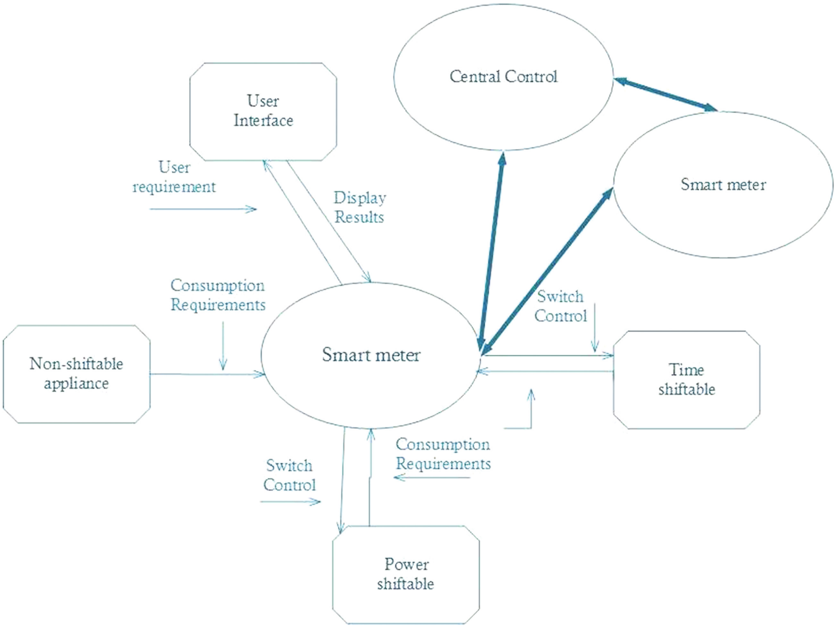


Fig. 1. Energy management system

4 Proposed Technique

A new technique in smart grid; PIO is proposed by authors to deal with electric consumption, electric bill and peak-load-reduction. PIO is inspired by homing behaviour of pigeons based on swarm intelligence algorithms. PIO resembles to ACO, Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO) and Differential Evolution (DE) algorithm in their control parameters i.e. compass factor and map. Pigeons sense the earth with and map the location in brain using magneto reception. Pigeons fly close to their destination relying on their landmarks. If pigeons are away from destination, they take landmark of their nearest pigeon. This is how they find food on basis of their location. Moreover, working of PIO is explained by Algorithm 1.

Algorithm 1. PIO Algorithm

Inputmaximum iterations
Initialization pigeonnum, D, map and compass factor, T1, T2, X_g
Specify LOT of appliances and power ratings
Randomly initialized the population
Set initial path X_i and velocity V for each appliance
St X_p=X_i
calculate the fitness of individual appliances
find the optimal solution
map and compass operator
for $t = 1 \rightarrow T1$ **do**
    for $i = 1 \rightarrow pigeonnum$ **do**
        while X_i *is beyond the search range* **do**
            calculate X_i and V_i
        end
    end
    for *j=1 to D* **do**
        while X_P *is beyond the search range* **do**
            sort all the appliances according to their fitness values
            pigeonnum=pigeonnum/2
            keep half of the appliances with better fitness value and discard the other half
            X_c= average of the remaining appliances
            calculate X_i
        end
    end
    Output: X_g is output as the global optima of fitness function
end

Algorithm 2. EDE Algorithm

Randomly initialize the population
while *termination criteria is not satisfied* **do**
    Perform mutation
    if $G \leq 100$ **then**
        Perform crossover
        Find best member in each group of trial vectors
        compare trial vectors
        choose trial vector with best values
    end
    Perform selection
    set G=G+1
end

5 Simulation Results

Performance of the proposed models i.e. PIO and EDE in terms of PAR is shown in Fig. 6. It shows that almost 60% of PAR is reduced using EDE and PIO as compared to PAR of unscheduled load as the main purpose of these models is to avoid peak formation. However, 3% more PAR is reduced using PIO than that of EDE. PAR is reduced as electric consumption is shifted from on peak to off peak hours. By shifting the on-peak hour load to off-peak hours, PAR is reduced, as no peaks are created and load is balanced. However, in unscheduled load peaks are generated as load is not scheduled. Due to scheduling and shifting of load peak average ratio is minimized and our models efficiently dealt with the peak formation problem.

Figure 4, shows the hourly cost of PIO and EDE as compared to unscheduled load. In order to achieve maximum user comfort both electricity cost and waiting time are needed to be minimized. However, the described attributes consist of inverse relation between them. In order to reduce bill, user must use their appliance according to EMC. Waiting of appliance is reduced with increase in cost. Figure 3, shows that waiting time of appliances is reduced using heuristic algorithm PIO as compared to EDE. So the results clearly depict that more user comfort is achieved using PIO in terms of waiting time.

Figure 3, result shows that the cost of scheduled load is less as compared to unscheduled load because the load is shifted from peak hours to off-peak hours. By Scheduling the load, the overall load is shifted from on peak hours to off-peak hours. So, no peaks are created and cost is minimized. As described earlier, power rate of each appliance is fixed and there is no change in their values which effects the reduction of cost. Pricing signals also reduced the cost because the charges in off-peak hours are less as compared to peak hours. However, electricity

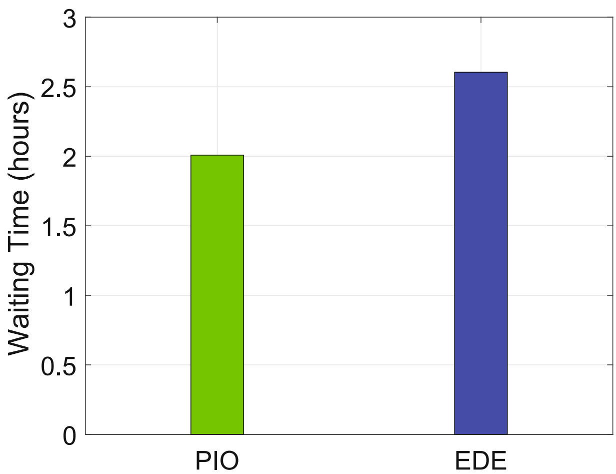


Fig. 2. Waiting time

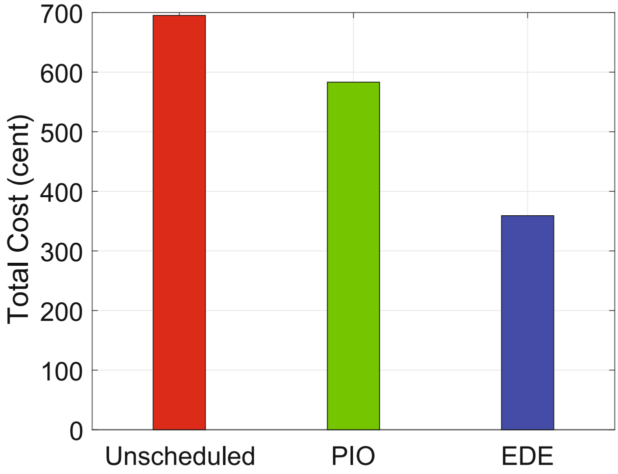


Fig. 3. Total cost

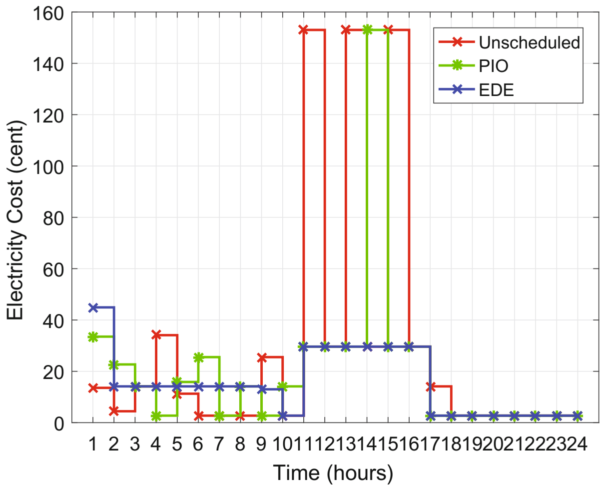


Fig. 4. Hourly cost of PIO and EDE

cost using PIO technique is greater than that of EDE, because user comfort is maximized in terms of waiting time. Trade off between cost and waiting time is clear from Figs. 3 and 2. Moreover, the total amount of load consumed remains the same as that of in EDE algorithm and PIO as shown in Fig. 5.

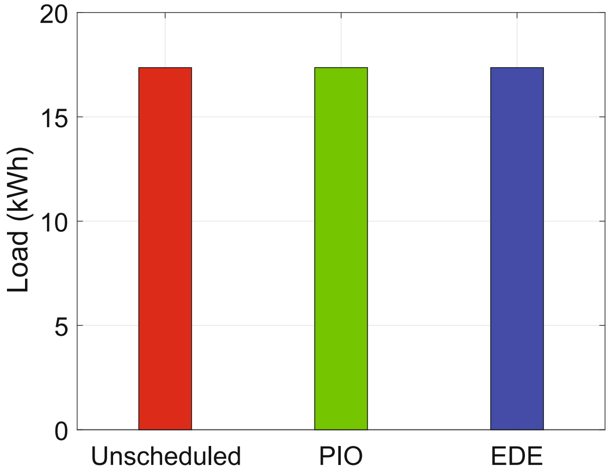


Fig. 5. Total load

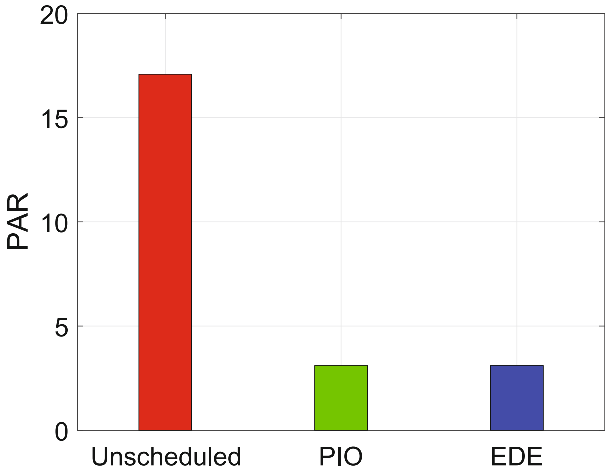


Fig. 6. Peak average ratio

6 Conclusion and Future Work

In this paper, we have proposed two heuristic techniques to minimize the electric cost and maximize the user comfort. The appliances have been scheduled on the basis of their power-shiftable, non-shiftable and time-shiftable behaviour, which deals with CPP. The results show that EDE works more effectively than PIO in terms of cost reduction. Less cost is to be paid in EDE, however, user comfort is being compromised. Simulation results show that PIO is identified as the best technique as it facilitates the user by reducing the waiting time of appliances and more PAR is minimized. Additionally, simulation results are shown in Sect. 5 to validate our work. In future, we will focus on multiple users management system. We will also work on integration of renewable energy resources in EMS.

References

1. Bobmann, T., Staffell, I.: The shape of future electricity demand: exploring load curves in 2050s Germany and Britain. Energy **90**(2), 1317–1333 (2015)
2. Bouvier, S.: Deploying a smarter grid through cable solutions and services. Accessed 18 May 2017
3. Yan, Y., Qian, Y., Sharif, H., Tipper, D.: A survey on smart grid communication infrastructures: motivations, requirements and challenges. IEEE Commun. Surv. Tutor. **15**, 5–20 (2013)
4. Ma, R., Chen, H.H., Huang, Y.R., Meng, W.: Smart grid communication: its challenges and opportunities. IEEE Trans. Smart Grid **4**, 36–46 (2013)
5. Ipakchi, A., Albuyeh, F.: Grid of the future. IEEE Power Energy **7**, 52–62 (2009)
6. Nguyen, H.K., Song, J.B., Han, Z.: Distributed demand side management with energy storage in smart grid. IEEE Trans. Parallel Distrib. Syst. **26**, 3346–3357 (2015)
7. Goran, S.: Demand side management: benefits and challenges. Energy Policy **36**(12), 4419–4426 (2008)
8. Yao, E., Samadi, P., Wong, V.W.S., Schober, R.: Residential demand side management under high penetration of rooftop photovoltaic units. IEEE Trans. Smart Grid **7**(3), 1597–1608 (2016)
9. Chiu, T.C., Shih, Y.Y., Pang, A.C., Pai, C.W.: Optimized day-ahead pricing with renewable energy demand-side management for smart grids. IEEE Internet Things J. **PP**(99), 1–10 (2016)
10. Jagruti, T., Basab, C.: Demand side management in developing nations: a mitigating tool for energy imbalance and peak load management. Energy **114**(1), 895–912 (2016)
11. Mohsenian-Rad, A.H., Wong, V.W.S., Jatskevich, J., Schober, R., Leon-Garcia, A.: Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. IEEE Trans. Smart Grid **1**(3), 320–331 (2010)
12. Thillainathan, L., Dipti, S., Zong, S.T.: Demand side management in smart grid using heuristic optimization. IEEE Trans. Smart Grid **3**(3), 1244–1252 (2012)
13. Logenthiran, T., et al.: Multi-Agent System (MAS) for short-term generation scheduling of a microgrid. In: 2010 IEEE International Conference on Sustainable Energy Technologies (ICSET). IEEE (2010)
14. Ghazvini, M.A.F., Faria, P., Ramos, S., Morais, H., Vale, Z.: Incentive-based demand response programs designed by asset-light retail electricity providers for the day-ahead market. Energy **82**, 786–799 (2015)
15. U.S. Department of Energy: Benefits of demand response in electricity markets and recommendations for achieving them (2006)
16. Boisvert, R., Cappers, P., Neenan, B.: The benefits of customer participation in wholesale electricity markets. Electricity **15**(3), 44–51 (2002)
17. Hong, S.H., Yu, M., Huang, X.: A real-time demand response algorithm for heterogeneous devices in buildings and homes. Energy **80**, 123–132 (2015)

18. Gao, D., Sun, Y., Lu, Y.: A robust demand response control of commercial buildings for smart grid under load prediction uncertainty. Energy **93**(1), 275–283 (2015)
19. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–5. IEEE (2012)
20. Samadi, P., Wong, V.W.S., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
21. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
22. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
23. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
24. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
25. Srinivasan, D., Rajgarhia, S., Radha Krishnan, B.M., Sharma, A., Khincha, H.P.: Game-theory based dynamic pricing strategies for demand side management in smart grids. Energy **126**, 132–143 (2016). Elsevier
26. Mahmood, D., Javaid, N., Alrajeh, N., Khan, Z.A., Qasim, U., Ahmed, I., Ilahi, M.: Realistic scheduling mechanism for smart homes. Energies **9**(3), 202 (2016)
27. Ahmad, A., Javaid, N., Alrajeh, N., Khan, Z.A., Qasim, U.: Khan, A.: A modified feature selection and artificial neural network-based day-ahead load forecasting model for a smart grid. Appl. Sci. **5**(4), 1756–1772 (2016)
28. Rasheed, M.B., Javaid, N., Awais, M., Khan, Z.A., Qasim, U., Alrajeh, N., Iqbal, Z., Javaid, Q.: Real time information based energy management using customer preferences and dynamic pricing in smart homes. Energies **9**(7), 542 (2016)
29. Khan, M.A., Javaid, N., Mahmood, A., Khan, Z.A., Alrajeh, N.: A generic demand-side management model for smart grid. Int. J. Energy Res. **39**(7), 954–964 (2015)
30. Karimi-Nasab, M., Modarres, M., Seyedhoseini, S.M.: A self-adaptive PSO for joint lot sizing and job shop scheduling with compressible process times. Appl. Soft Comput. **28**, 137–147 (2015)

Energy Optimization Techniques for Demand-Side Management in Smart Homes

Syeda Aimal, Komal Parveez, Arje Saba, Sadia Batool, Hafsa Arshad, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. Due to increase in population, demand of energy is increasing. To make energy demand efficient and reliable many techniques are integrated in home areas. We implemented Grey Wolf Optimization (GWO) using Time of Use (TOU) pricing scheme, to achieve an optimal balanced load and to minimize user comfort, then we compared the results of GWO and Bacterial Foraging Algorithm (BFA). The scheduling mechanism is capable of achieving the optimal operational time. Simulation results are presented to demonstrate the effectiveness of optimization techniques.

1 Introduction

Smart grid is an electrical grid which includes a variety of operational and energy measures including smart appliances, smart meters, energy efficient resources, and renewable energy resources. In current years, people are facing significant issues related to utility management due to increase in population. There is a need for fundamental development of the electricity power system. The changeover makes the next generation of power grid an intelligent and reliable system that is commonly known as a smart grid as shown in Fig. 1.

In smart grids, new technologies improve the productivity and efficiency of the power system. Information and communication technologies are applicable to the traditional grid to increase its efficiency, flexibility, and reliability. Smart grid is becoming an interesting area for research. In our scheme, we propose Demand-side Management (DSM) using BFA and GWO optimization techniques by applying TOU as a pricing model, inspired by the scheduling procedure in [1]. It is applicable to HEM unit through embedding it in the smart meter. The smart meter is capable of providing the fittest Scheduling to the connected appliances in smart homes.

The purpose of DSM is to reduce the energy demand for the customer, its main focus is managing and controlling the loads from on-peak to off-peak hours, and to keep the energy use in an efficient and sustainable way. DSM is applicable to electricity loads as well as is also used for the changes that has been made to demands for all types of energy. Our emphasis is to achieve an optimal balanced load and to minimize user comfort. The paper is organized in a following way:

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_46

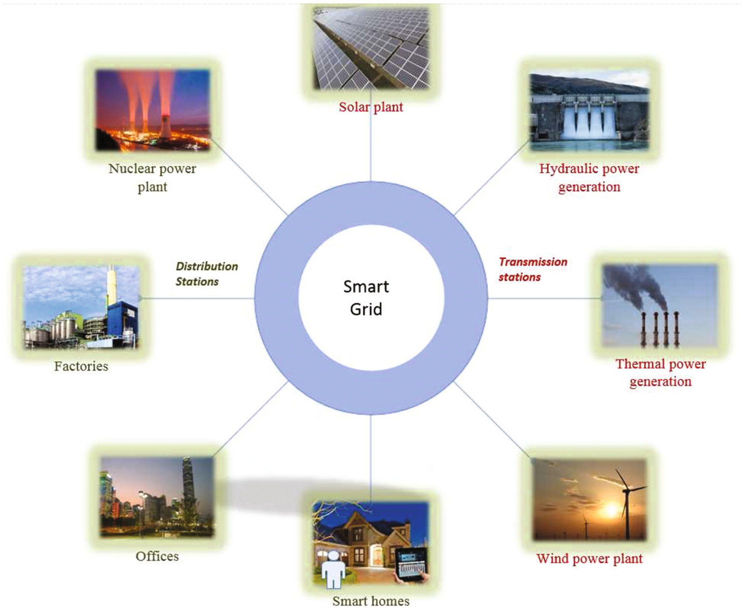


Fig. 1. Smart grid

Related Work is described in Sect. 2, Proposed Model is presented in Sect. 3, Simulation Results are discussed in Sect. 4. In the end, Conclusion is given in Sect. 5.

2 Related Work

Smart grids are the advance network, that are need to meet the challenges like robustness, stability, reliability. It allows us the implementation of Energy Management systems at home premises and also to the areas like business, commercial etc. Currently, many researches have been done in this regard; some of them are as follows:

In [1], the authors proposed a technique named Integer Linear Programming (ILP) Based Optimization. The objective mentioned in this paper is to reduce Peak to Average Ratio (PAR) for balancing daily load schedule. For that the author took an example of home area appliances by categorizing the appliances into three groups such as non-shift able, time-shift, and power-shift able. The scheduling mechanism is able to achieve minimum cost, reduction in peak to average load. We also see that if the scheduling is applied to the multiple houses with same scenario, a more balanced load is achieved. However, the waiting time might effects the scheduling scenario.

The authors addressed the problem of the electricity cost reduction and increasing waiting time [4]. For that purpose five heuristic algorithms are used.

The load of single home and then for multiple homes are tested. The author used ANOVA test for statistical analysis for measuring the variation in algorithms performance metrics. However, RTP pricing model may results in the instability which may root to blackouts due to increase in the electricity demand and peak to low price time.

In paper [5], the authors focuses to achieve the tradeoff that exists between price and user comfort through scheduling appliances. Day-ahead pricing model is used and the author proposes a scheduling scheme that is used to apply home area for demand response. However, the second type of appliances, to reduce payment may lead to user discomforts were not handled properly.

Problems regarding load scheduling and power trading [2] are being addressed and for resilient and distributed generation control of renewable energies in Micro grids [12] focused by the authors. The game theoretic approach is used to model the interaction between users that can buy and sell excess generated power to other local users with an offered/selected price. The achievements show that results in the reduction in cost as compared where trading is not applied. Although the Electricity cost is reduced; On the other hand while power trading; to improve the quality of power results in additional cost of energy storage resources. To improve the quality of power, maintainability cost, and installation cost may effects the whole scenario.

In paper [3] An Optimal scheduling method is used to lower electricity rate cost and the PAR ratios. Author classifies two kinds of home appliances; such as automatically and manually operated appliances to manage energy utilization with GA. The author combines RTP with IBR. In this paper the user has mentioned manually operated appliances, from them no one can surely say in advance that how time it will take to run as they are operated manually.

In paper [6], the author have considered three heuristic techniques; i.e. GA, Ant colony Optimization [ACO], BPSO. For electricity cost TOU and IBR pricing models are used. However, the author integrates RESs with GA-EMC, BPSO-EMC, and ACO-EMC to reduce the electricity bill, which may lead to further cost of RESs.

In paper [7], the authors have proposed the online energy management as a stochastic optimal power flow (SOPF) problem and thus uses the Lyapunov optimization to design an online algorithm to solve it in real scenario. The results show online EMS may leads to the security and reliability problems as well as it may not guarantee the quality of service for users.

In [8], for the better operation of micro grids, distributed EMS has been designed. This is being compared with the existing distributed approaches. The authors formulates the micro grid energy management as an Optimal Power Flow (OPF). The authors applies distributed EMS on the basis of IEC 61850. The proposed EMS is applied to the real scenario micro grid in Guangdong province, china. The result shows that EMS is functional for islanded as well as grid-connected mode. It shows its higher convergence rate. Consequently, the whole scenario may lead to uncertainty in processing times, prices, changes in demands. A spare coding approach to household electricity demand forecasting in smart grid [10]. The authors uses the spare code to model and forecast the electricity Demand for individual and 5000 household. However, demand forecast could be expensive both in terms of time and resources [9]. The framework for scheduling the households are able minimizes the cost without disturbing the devices that are non-schedulable. User comfort is neglected in this scenario. [11], the paper considers and plans expected and unfavorable consequences for Power quality. The mentioned assumption cannot be practically implemented.

3 System Model

Many techniques are proposed to manage the load and cost to achieve an optimal solution in smart home area network, the proposed mechanisms are also able to reduce cost as well as less user comfort level is compromised. In Fig. 2, the smart homes are integrated with smart meters in which the appliances are controlled by Home Energy Management (HEMS). Through HEMS, the smart meter users are able to manage their energy consumption. HEMS controller is used to control and operate the appliances according to their specified time period. In our work, we classify the appliances into three categories: Non-Shiftable, Power-Shiftable, and Time-Shiftable on the basis of their operating period, hourly consumption and daily requirement of each appliance in [1].

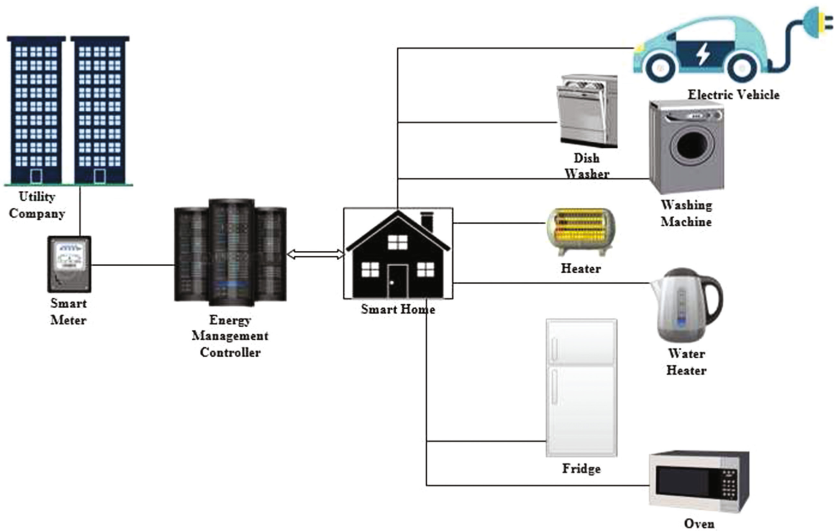


Fig. 2. HEMS in smart homes

Appliances and their energy usage patterns are given in Table 1.

Table 1. Power scheduling according to appliances

Group	Classification	Power rating
Washing machine	Time-shiftable	Hourly consumption: 1–0.5 kWh, Operating period: 2 h
Dish washer	Time-shiftable	Daily consumption: 0.8 kWh, Operating period: 1 h
Electric vehicle	Power-shiftable	Charging power: 1–3 kWh, daily req: 5 kWh
Water boiler	Power-shiftable	Daily consumption: 3 kWh
Fridge and freezer	Non-shiftable	Hourly consumption: 0.12 kWh, Operating period: 24 h
Oven and heater	Non-shiftable	Hourly consumption: 1 kWh, Operating period: 7pm–8pm (oven), 9pm–10pm (heater)

We schedule our appliances according to two optimization techniques. i.e. BFA and GWO using TOU as shown in Fig. 3, on the basis of their operating period, hourly consumption and daily requirement of each appliance in [1]. BFA was firstly introduced by Kevin Passino in 2002, which is nature inspired optimization technique; the key idea for this technique is to maximize the energy (obtain per unit time) by searching the nutrients. Each bacterium generates signals to communicate with others. The process of movement in each bacterium takes small step in the search of nutrients is called chemotaxis. The main impression of BFA is to mimic chemotactic motion of bacteria in the search space. BFA consists of the following steps:

(1) Chemotaxis,
(2) Swarming,
(3) Reproduction,
(4) Elimination and Dispersal.

In step 1: Provoke the motion of the bacteria through swimming and tumbling, it may swim or tumble for their whole period of life.

step 2 and 3: Motion of cell releases signal to communicate with other whereas the reproduction step include only those cells that can play their part to generate new population.

step 4: Where the previous cells are eliminated and new randomly selected samples are implemented.

GWO was put forward by Mirjalili *et al.*, in 2014 [13]. Grey wolf are meant to be the apex predators, as they are at the top of food chain. They like to live in group. The average group size is 5–12. The Hierarchy of grey wolf consists of alpha, beta, omega, and delta.

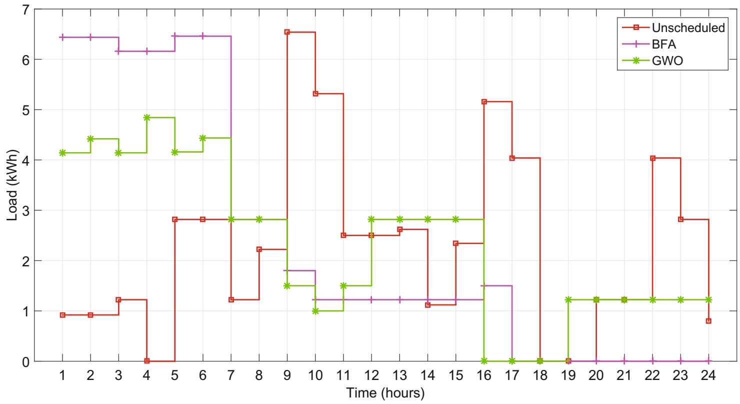


Fig. 3. Load

alpha: The wolves (male or female) that are responsible for decision-making about hunting, sleeping, etc.

beta: Wolves that helps alpha in decision-making. Delta: are the junior wolves that control the omega. They are divided into: scouts, caretakers, hunters, elders, sentinels.

omega: Considered as the lowest rank, their job is to serve the pack as a scapegoat and are the last that are permit to eat. They are not the important part of the group, but the pack will face internal fight and hurdles if they lose omega. The hunting tacts of grey wolf are as follows:

Dogging, following, and approaching prey.
Encircling, harassing unless the prey stops moving.
Finally Attacking towards prey.

To miniature the social hierarchy of the wolves. GWO is presented in which we considered alpha, as the fittest results. Accordingly beta and delta are the best solutions. The rest of the results are considered as omega.

4 Simulations and Discussions

We classified home appliances into three categories:

(a) Non-Shiftable, (b) Power-Shiftable, and (c) Time-Shiftable appliances on the basis of their operating period, hourly consumption and daily requirement of each appliance. Results are evaluated in matlab, in which we see the load as shown in Fig. 3, is balanced and the appliances that creates the peak are shifted to operate in off-peak as compared to BFA. In this scheme, although the electricity cost is increasing by 5% but the PAR is reduced approximately 8% in comparison to BFA.

The total cost as shown in Fig. 4, is 450 in unscheduled case while in case of BFA it is reduced to almost 320 in comparison with GWO. We see that the

Algorithm 1. GWO scheduling

```
Randomly initialize the position of search agents
Evaluate the position search agents
while iter < iter_max do
   for i = 1:size(Positions, 1) do
      fitness/obj function
   end for
   a value linearly from 0 to 2
   for i = 1:size(Positions, 1) do
      for j = 1:size(Positions, 2) do
         randomly initialize the value between 0 to 1
         calculate the value of A and C co efficient factors
         update the position of Alpha (X1), Beta (X2), Delta (X3)
         get the best positions
      end for
   end for
end while
```

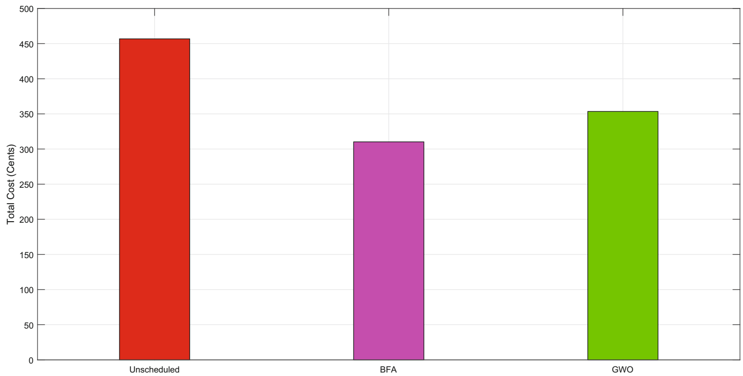


Fig. 4. Total cost

scheduled BFA is constantly decreasing as compared to unscheduled and GWO is also decreasing as compared to unscheduled. Here, BFA is good in terms of cost as compared to GWO.

On the other hand, if we see the waiting time as shown in Fig. 5, it is reduced by approximately 11% and the user comfort is achieved. There is always a trade-off between user waiting and cost; if the cost will increase than the waiting time will decrease and if the waiting time increases than the cost will decrease.

The PAR is reduced up to 75% as compared to unscheduled and approximately by 10% in comparison with BFA. The PAR as shown in Fig. 6, shows the GWO effectively reduced the peak load as compared to BFA and unscheduled (Figs. 7 and 8).

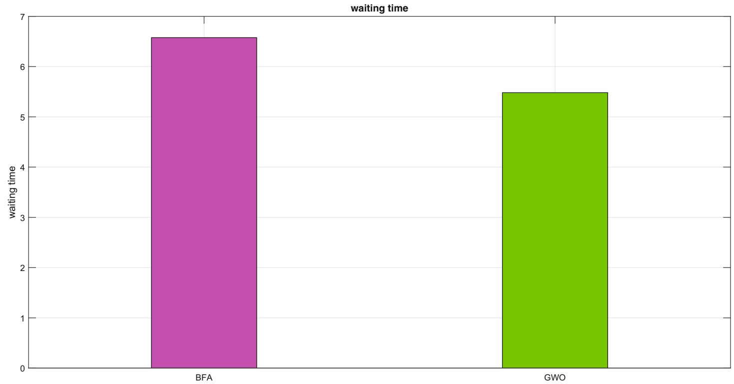


Fig. 5. Waiting time

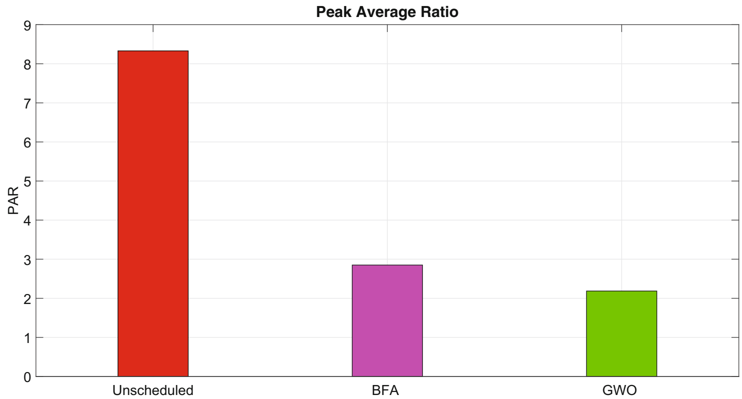


Fig. 6. PAR

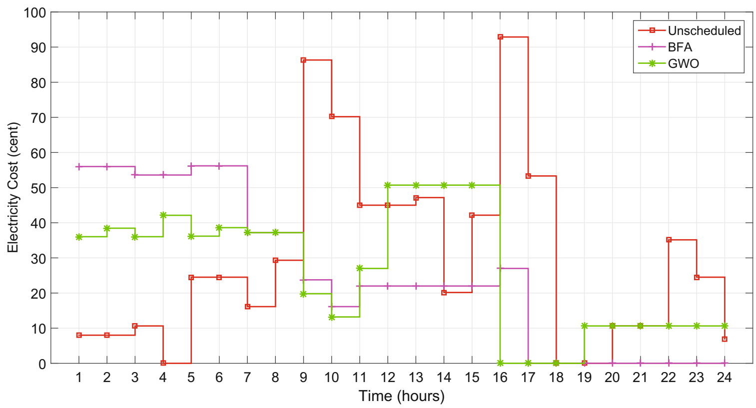


Fig. 7. Cost

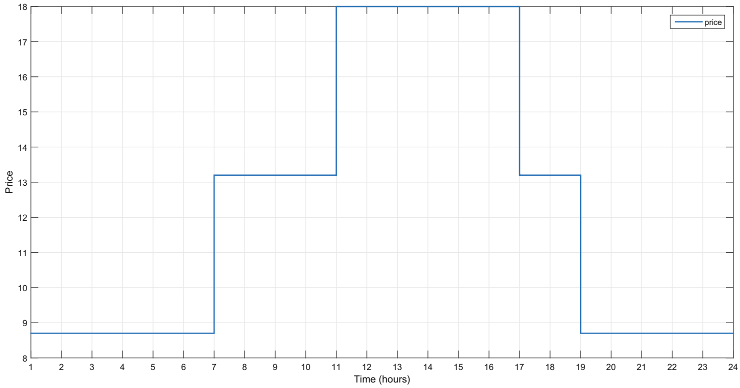


Fig. 8. TOU

5 Conclusion

In this paper, we have implemented GWO technique with TOU pricing scheme and used to schedule the appliances to attain an optimal balanced load and also to minimize the waiting time. We classified home appliances into three categories:

Non-Shiftable, Power-Shiftable, and Time-Shiftable appliances on the basis of their operating period, hourly consumption and daily requirement of each appliance. The scheduling mechanism is capable of achieving the optimal operational time for time-shiftable and power-shiftable appliances according to their energy consuming patterns for each appliance.

References

1. Zhu, Z., et al.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT). IEEE (2012)
2. Samadi, P., et al.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
3. Zhao, Z., et al.: An optimal power scheduling method for demand response in home energy management system. Autom. Auton. Syst. **8**(6), 176–178 (2016)
4. Javaid, N., et al.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
5. Kai, M., et al.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
6. Rahim, S., et al.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
7. Shi, W., et al.: Real-time energy management in microgrids. IEEE Trans. Smart Grid **8**(1), 228–238 (2017)

8. Shi, W., et al.: Distributed optimal energy management in microgrids. IEEE Trans. Smart Grid **6**(3), 1137–1146 (2015)
9. Basit, A., et al.: Efficient and autonomous energy management techniques for the future smart homes. IEEE Trans. Smart Grid (2015)
10. Yu, C.-N., et al.: A sparse coding approach to household electricity demand forecasting in smart grids. IEEE Trans. Smart Grid **8**(2), 738–748 (2017)
11. Bollen, M.H., et al.: Power quality concerns in implementing smart distribution-grid applications. IEEE Trans. Smart Grid **8**(1), 391–399 (2017)
12. Chen, J., et al.: A game-theoretic framework for resilient and distributed generation control of renewable energies in microgrids. IEEE Trans. Smart Grid **8**(1), 285–295 (2017)
13. Mirjalili, S., et al.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
14. Li, F., et al.: Smart transmission grid: vision and framework. IEEE Trans. Smart Grid **1**(2), 168–177 (2010)
15. Zhangt, P., et al.: Next generation monitoring, analysis and control for the future smart control centre. IEEE Trans. Smart Grid **1**(2), 186–192 (2010)
16. Ayesha, Z., et al.: A meta-heuristic home energy management system

DSM Using Fish Swarm Optimization and Harmony Search Algorithm Using HEMS in Smart Grid

Shahzeb Haider, Hafiz Muhammad Faisal, Zenab Amin, Haq Nawaz, Kaleem Akram, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
https://www.njavaid.com

Abstract. Proliferation in smart grid gave rise to different Demand Side Management (DSM) techniques, designed for type of sectors i.e. domestic, trade and commercial sectors, very effective in smoothening load profile of the consumers in grid area network. To resolve energy crises in residential areas, smart homes are introduced; contains Smart Meters, allows bidirectional communication between utilities and customers. For this purpose, different heuristic techniques are approached to overcome state of the art energy crisis which provide best optimal solution. The purpose of our implementation is to reduce the total cost and Peak to Average Ratio value while keeping in mind that there is a trade-off of these with waiting time up to an acceptable limit. Our proposed scheme uses heuristic technique Harmony Search Algorithm with Fish Swarm Algorithm to achieve the defined goals. Real time prizing signal is used for bill calculation in Advanced Metering Infrastructure.

1 Introduction

Demand side management (DSM) techniques are used to modify the consumer demand to meet the energy demand through different methods like financial and behavioral changes through education. In order to avoid blackouts and major distress, load management focuses on efficient use of energy management. Load management involves efficient energy consumption, reduces number of peak power plants, electricity bill reduction and improvement in power grid performance in terms of flexibility and reliability [1]. Demand Response (DR) is an action, engaged by a consumer against dynamic price models. It offers many financial and operational benefits for electricity utilities, end users and grid operations. Demand response is important to tackle the uncertainties because high volatile nature of load may threaten the integrity of grid within seconds as it provides flexibility at comparatively lower rates.

Traditional power grid system is insufficient to meet the requirements of power grid such as stability, flexibility, accountability, reliability and robustness. Thus, an advanced metering infrastructure is needed to meet the state

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_47

of the art challenges smartly and reduce pressure on global grid system environment. Increasing development in information and communication technology gives rise to the need of new change in traditional grid system. In this concern, concept of smart grid (SG) is introduced which is the integration of information and communication technology in old fashioned power grid system, enables bi-directional communication between utility and the customer. To maintain the balance b/w demand and supply, demand side management (DSM), one of the important aspects of SG is used. DSM further contains demand response (DR) and load management. We use heuristic techniques to find the solution of our problems. Heuristic techniques are used to provide an optimal solution for a given problem in simulated atmosphere which later, can be implemented in real environment. In order to create a real time and acceptable solution, we use meta-heuristic approaches. Meta-heuristic is multi-objective technique among which swarm intelligence techniques are very efficient as containing decentralized and self-organizing approaches. The fish usually used to be nearer to each other in order to protect themselves from attacker and to avoid collision with in the swarm. They move in a colonial form to avoid male fish to prevent reproduction. An Artificial Fish Swarm Optimization (AFSO) is one of the swarm intelligent methods which refer to the colonial and individual behavior of an Artificial Fish (AF) as balanced and move in search of food.

In [2], authors FSA is combined with Particle Swarm Optimization (PSO) to improve its performance.

For this purpose, FSA communication behavior is added and PSO formulation is used to reconfigure FSA while reducing parameter complexity of FSA.

2 Related Work

Due to increasing population, energy crises are the major distress in every kind of field so in smart grid, results in blackouts and load shedding. To overcome this issue, certain mechanisms and techniques are proposed. These techniques are adopted by utility to facilitate user which indirectly benefits utility. Minimization in electricity cost while maximization in user comfort facilitates user truly, however there exists a trade-off between cost and waiting time. On the other side utility is benefited by normalization of load in the available time. We take a few references that describing HEMS and categories the references on the basis of their various characteristics (Fig. 1).

There are two types of techniques: State Alone (SA) Techniques and Hybrid Techniques (HB).

2.1 State Alone Techniques

In SA techniques, a single technique features and some other techniques that are used, discussed below. SA Technique is further divided in to two Integer Programming (IP) Based Techniques and Optimization techniques. Properties are organized to obtain the expected goals i.e. to reduce electricity cost using any heuristic algorithm. Different types of SAT are discussed below.

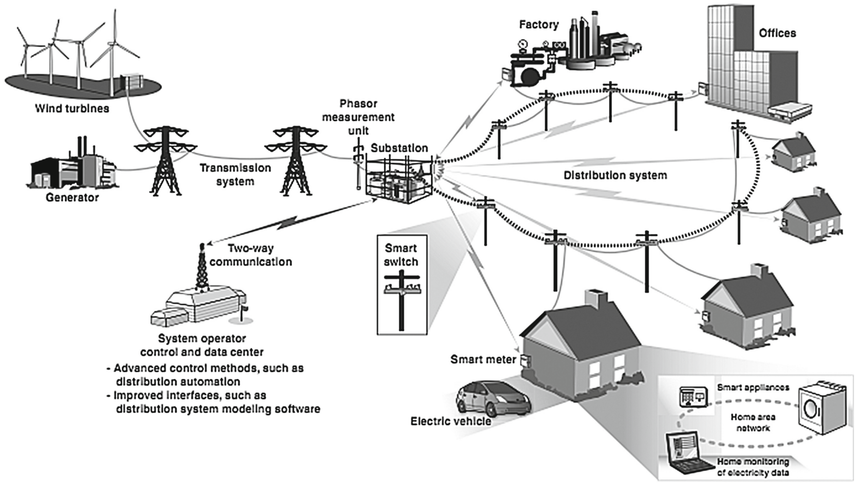


Fig. 1. Smart Grid

2.1.1 Integer Programming Based Technique

IP base techniques are mostly used to reduce the computational cost. Moreover, it can be used for optimal scheduling of appliances and improve demand response for user satisfaction. Below are the further discussions on IP. Authors in [3] take a Home Area. Network (HAN) and schedule the appliances using integer linear programming technique. PAR is a great issue now a day causes load shedding and energy power loss. The user preferences and operation of each appliance is trying to satisfy beside its major goal as reduction of peak load. The scheduler optimally schedules the appliances and reduces PAR. UC in sense of waiting time is compromised. The PAR value is reduced which automatically reduced the electricity bill of user. However waiting time of appliances are ignored.

Over the year, more people taking interest in DSM, as also discussed in paper [4]. Authors describe a smart energy management system in order to moderate Peak Load and electricity cost using Multi Objective Mixed Integer Linear Programming (MOMILP) technique. Pricing scheme used is Time of Use (TOU). The scheduler shift high load consumption appliances from On Peak-Hours (On-PHs) to Off-Peak-Hours (Off-PHS), by this PAR at an hour are in specified limit with the minimization of electricity cost. UC is sacrificed in terms of waiting time. User waits for the optimal time to run an appliance. The work in [5], discussed the balance between energy saving and UC using MOMILP with Mixed Objective Function. It presents joint scheduling and task or operation management of different appliances, UC in term of thermal and electrical zones and improved energy management optimization framework. UC is compromised as not core objective. Smart controllers improve mixed objective function value 25 % with respect to naive up to 55 and normal controller up to 63% in hot weather condition and 38% in cold weather condition. In [6], Authors describe the residential electrical scheduling problem. The appliances are classified in to five different classes: Elastic Appliances (EA) with memory less property, EA with full memory, EA with partial memory, EA with an interruption operation and EA with non-interruptible operation, on the basis of different operational properties and energy consumptions. Problem is formulated by Mixed Integer Nonlinear Programming (MINLP) and overcome by Generalized Bender's Decomposition (GBD) approach. TOU tariff is used as pricing scheme. The numerical result shows that each appliance is scheduled effectively according to its operation properties and energy consumptions considering preference of utility and cost.

2.1.2 Optimization Techniques

In Paper [7], focus is on the trading energy and scheduling the load in system. The RES is highly integrated to achieve corresponding goals. Dynamic programming is used for scheduling the appliances which are controllable. A game theory scenario that user can sell their extra energy back to utility or other neighboring user, is applied. Generalized Benders decomposition used to tackle trading problem. Main Goal of the paper is to reduce peak load per hour of user in order to get electricity cost reduction and PAR value minimization. As cost and PAR have a trade-off with UC, therefore UC is compromised in achieving the objective goals. The contribution work in [8], is based on SG appliances' optimal power scheduling. Two types of appliances time flexible and power flexible. Time flexible showing flexibility in starting time and power flexible are flexible in power to achieve minimum electricity cost. Day a head pricing scheme is used to get a desirable trade-off between UC and electricity cost. There are three operation modes. In first, only electricity cost minimization is the main goal, in second UC is main goal and in third both have equal priority. PAR is neglected. Also paper [9], focused on Demand Response (DR) to optimize operational time of appliance in HEMS. Gradient based Particle Swarm Optimization (GPSO) and Two-Point Estimate Method (2PEM) to produce an optimize solution and electricity cost minimization. The Monte Carlos Simulation (MCS) deals with uncertainty, however MCS convergence require larger number of simulations. Hence new technique referred as chance constraint programming (CCP) is used to deal with uncertainty. Major goal is to achieve the demand response in sense of computational cost. A lot of work is done to reduce power utilization. As in Paper [10], describe the efficient utilization of power. It explains that load of PHs can be utilized of in OPHs to reduce peak formation. GA in DSM (GA-DSM) benefits overall reduction of 21.91%. The result shows that after applying GA-DSM application power reduction is about 23.84% while in heuristic optimization power reduction is 14.2%. The power usage of electricity is reduced in PHs about 18.55% and 17.49% at different industrial areas, 23.81% in residential areas and 21.15% and 19.25% at various commercial areas. In SG huge amount of work is done for the implementation of energy management site in HEMS, same as in [11]. To produce the required results MILP problem that can be solved via step-wise approach. Dijkstra algorithm is used, which gives the same performance at very low with the use of much lower complexity.

Major Goal of the study is to reduce electricity cost under the maximum available energy. Efficient solution with lower complexity is obtained by dividing problem in to two independent sub problems.

2.2 Hybrid Techniques

In HB techniques two are more techniques combine together to cover each other deficiencies and provide more effective operations. The performance of HB is more optimal than other techniques, as it contains the properties of all techniques that merged. Below are some references that explain the concept in detail.

In [12], author's major goal is to obtained optimal minimum PAR value and smaller. HAN uses SG Energy Management System (EMS) to schedule appliances in better way. Real Time Pricing Scheme with Inclined Block Rate (IBR) scheme is used. GA scheduler optimally schedule load of the appliances to avoid peak creation. By combining IBR with RTP reduces mean PAR to 3.37 from individual value of mean RTP 5.51 it means that it reduces mean PAR up to 38.84%. With power scheduling electricity cost is reduced from 48.65 to 35.97 and PAR from 5.22 to 3.37, it means that electricity cost is reduced up to 36.07% and PAR values up to 35.44%. However UC is neglected in order to achieve the objective. As in paper [13], five heuristic optimization techniques are used as GA, BPSO, BFO, WDO and proposed hybrid genetic wind driven optimization (GWD). Pricing scheme used is RTP. Features are used in order to get its three major goals i-e Maximizing UC, minimizing PAR and electricity cost. Results shows that electricity bill of GWDO better than GA than WDO as 30% than 60% than 62 % respectively. GWD beats GA and WDO in PAR value having 40%, 60%, 75% PAR values respectively. WDO is worst at PAR due to its high pressure of particles. The maximum waiting time for an appliance is declared as 4 hours, utility will pay penalty in case of failure to meet the condition. UC of GA, WDO and GWD is 60%. About 40% of UC is sacrifices. To efficiently utilized energy management controllers with RES described in [14], Authors exploiting three different heuristic algorithms as GA, Binary Particle Swarm Optimization (BPSO) and Ant Colony Optimization (ACO). Goal Of the authors is the reduction of electricity bill, power utilization, PAR, maximize UC and efficient integration of RES. Overall result shows that these heuristic algorithms optimally reduce electricity cost by GA-EMC, BPSO-EMC and ACO-EMC 75.4787cent, 90.4918cent and 98.0409cent respectively. PAR is reduced by GA-EMC, BPSO-EMC and ACO-EMC from 244.6747 to 81.88, 95.2281 and 127.5380 respectively.

Also in [2], authors FSA is combine Particle Swarm Optimization (PSO) to improve its performance. For this purpose FSA communication behavior is added and uses PSO formulation to reconfigure FSA. Also reduce the parameter complexity of FSA.

3 Problem Statement

Increasing energy demand gives birth to control the excess usage of energy to take over blackouts and load shedding. Different DSM techniques are used to tackle different bidirectional communication parameters. Control parameters are adjusted to balance the performance parameters. These parameters in smart grids are the pricing schemes, number of appliances, operation time and length of operation. Previous control parameters affect the performance parameters as PAR, Cost, Load and UC. There are some trade-offs between the parameters that directly affect each other performance. Cost reduction causes decrement of UC, so user waits for their appliances to be scheduled. In [12–14], authors goal was to reduce cost and PAR. We cannot deliver a model that supports a single part i.e. utility or user. In order to agree both the parties must compensate. Main objective of our proposed scheme is to reduce PAR and cost while keeping UC maximum. The simulation result shows that our proposed solution beat GA and unscheduled in more efficient manner.

4 Proposed Methodology

Present power grid system is not sufficient to meet the state of art grid system requirements such as stability, flexibility, accountability, reliability and robustness. Therefore we propose a model to replace the present traditional power grid system with the power system called Smart Grid.

For this purpose, DSM techniques are designed provides variety of measures to reduce energy consumption. It provides the way of managing energy consumption to optimize available planned resources for generation of power. Peak clipping, conservation and load shifting are the forms of DSM. Any of the DSM technique implement may result in demand reduction. Some of the DSM techniques are given below:

DSM Techniques

- Direct Load Control
- Load Limiters
- Commercial or industrialized programs
- Regulation of frequency
- Pricing model (TOU)
- Demand bidding (DB) Programs

The techniques mentioned above, play an important role in balancing load profile of the customers in grid networks. Our focus is on Home Energy Management System (HEMS) in residential areas. Our goal is to reduce PAR, cost and keeping in mind trade-off with UC. We implement FSA with Harmony search algorithm (HSA) and compare its result with GA. Results shows that FSA performs pretty better than GA. We divide whole appliances in to three categories i.e. Interruptible, non-interruptible and fixed appliances.

- Interruptible Appliances:

As Elastic Appliances are those, which user does not recommends strictly in sense of operation time as well as its length of operation. They can be stopped at any time when needed, in order to prioritize user comfort.

- Non-Interruptible Appliances:

As non-interruptible Appliances are those, which user strictly recommends not interrupting its operation time but can be shifted in specific time interval. After starting its operation it will never stop until it complete its length of operation. FSA reduce waiting time of non-interruptible appliances as optimally scheduling, so user comfort is least compromised.

- Fixed Appliances:

As Fixed Appliances are those, which user strictly recommends not interrupting its operation time and it's time to execute. These appliances waiting time must be zero as no user comfort in this case will be interrupted. Scheduling so waiting time is 0. Un-interruptible waiting time is acceptable as pretty low. Interruptible waiting time is not exceeding its limit. Un-interruptible is 25% of the maximum limit available (Table 1).

Table 1. Appliances parameters

Group	Appliance type	Power rating (kW)	Length of operation (hour)
Fixed	Refrigerator	0.225	18
Fixed AC	1.5	1.5	15
Fixed	Oven	1.8	7
Interruptible	Vaccum cleaner	0.7	6
Interruptible	Water heater	5	10
Interruptible	Water pump	1	8
Interruptible	Dish washer	0.8	10
Non-Interruptible	Washing machine	0.7	5
Non-Interruptible	Cloth dryer	5	4

FSA:

The FSA is inspired of fish behavior in its colony and individually. Fish searches food in the form of swarm or individual. They search food by their capability of all points available in their site range up to an angle. Fish have three basic steps on basis of which it obtain an optimize solution in searching food.

Pray:

It first checks whether there is any food around it which is called Pray. If visual scope of is empty than it chooses a random direction. Also if visual scope is too crowded it chooses random direction (Table 2).

$$y = x(i,:) + Visual * rand(1, D);$$

$$x(i,:) = x(i,:) + ((y - x(i,:))/(abs(y - x(i,:)))) * step * rand(1, D);$$

$$J(i) = sum(100 * (x(i, d+1) - x(i, d)^2)^2 + (x(i, d)1)^2);$$

Where visual shows its range of sight.

Table 2. Algorithm parameters

Variables	Values	Variables	Values
try number	5	D	9
Nc	5	step	0.1
pop step	30	theta	0.5
Ns	2	visual	0.5
n	0.9		

Chasing:
If the minimum function value exists in its visual scope then it will execute chasing strategy. As given

$$x(i,:) = x(i,:) + ((y - x(i,:))/(abs(y - x(i,:)))) * step * rand(1, D);$$

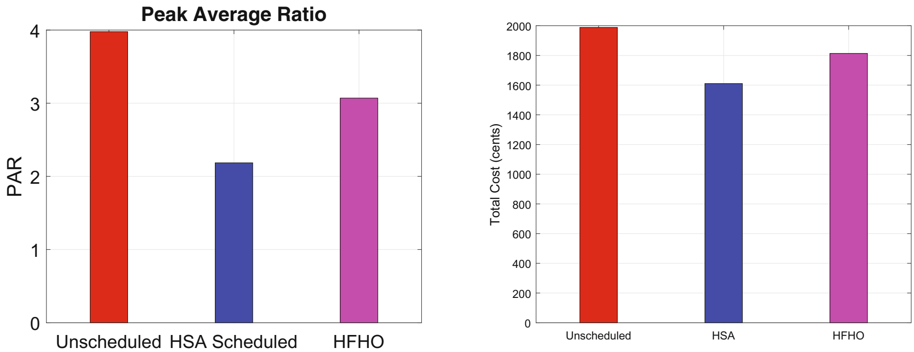
$$J1(i) = sum(100 * (x(i, d+1)x(i, d)^2)^2 + (x(i, d)1)^2).$$

Algorithm 1. Artificial Fish Swarm Optimization Alogrithm

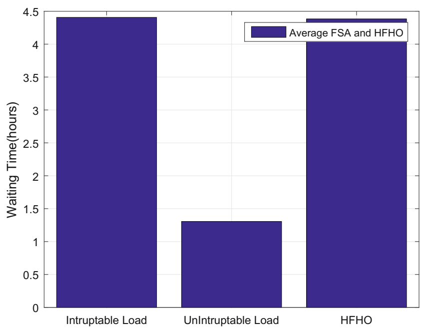
$InitializeParameters\ x, tryNum, step, pop_step, visual, sw_ln, FCS_fit, FSS_fit, FLS_fit$
$Initializexwithrandomformula$
for $k = 1 \rightarrow tryNum$ **do**
  for $i = 1 \rightarrow pop_step$ **do**
    if $ifFCS_fit(i) < FSS_fit(i)$ **then**
      if $FCS_fit(i) < FLS_fit(i)$ **then**
        $FLS_fit(i) = FCS_fit(i)$
      else
        $Behavior_Pray$
      end if
    else
      if $FSS_fit(i) < FLS_fit(i)$ **then**
        $FLS_fit(i) = FSS_fit(i)$
      else
        $Behavior_Pray$
      end if
    end if
  end for
end for

5 Simulations and Results

Our simulation results show that PAR in unscheduled case is 3.97, HSA when simulated alone in Matlab, gives result 2.1854, however our hybrid approach HSA with FSA also successfully brought the PAR value to 3.07. These values clearly show that FSO reduced the PAR value by almost 23% as compared to the unscheduled load, however more than HSA which is 45% reduced. The load consumption was optimally scheduled using HSA and FSO which reduced the PAR to 3.07 and 2.1854. The use of RTP is another reason behind the lower cost. The price of FSO optimized cost is lower than the unscheduled. Difference between FSO's maximum Electricity cost and unscheduled cost is 8.757% with RTP. FSO load's upper limit is lesser than that of unscheduled load. However, there's a trade off between waiting time and cost. We achieved efficient reduction in cost. In future we'll integrate renewable energy resources with the smart grid to maximize user comfort.



(a) Peak Average Ratio **(b) Total Cost**



(a) User Comfort

6 Conclusion

In this paper, the evaluation of load management problem is done in residential area's using meta heuristic approach. To achieve the defined goal, DSM techniques are used along with two heuristic techniques HSA with FSA. Using HSA individually and its hybrid approach with FSA yields Hybrid Fish Harmony Optimization (HFHO) algorithm. Their implementation on Matlab generate results in which PAR is minimized efficiently than Unscheduled while User Comfort is also maintained efficiently. Similarly, cost minimization is also done with the above performance parameters. Efficient use these heuristic techniques in practical environment can minimize the chance of load shedding and blackouts in any country.

7 Future Work

So far, energy optimization and load management techniques are discussed in this paper to prevent energy crises which results blackouts in smart grid systems. However, there's a major concern in smart grid in practical environment related to its security.

Smart meters, installed in open areas, can be a serious threat to costumer privacy and disturb its integrity. Except this, collusion b/w aggregator and utility or b/w SM's and aggregator even can expose user privacy. To fight with this fiercely, many privacy preserving and anonymous data aggregation schemes have been proposed which guarantees secrecy of the data, authentication of customer or gateway and preserve privacy of customers using different encryption and anonymization techniques [15–19]. In future, our goal is to achieve consumer's data privacy, integrity, authentication while keeping performance of the system efficient and reliable.

References

1. Gelazanskas, L., Gamage, K.A.: Demand side management in smart grid: a review and proposals for future direction. Sustain. Cities Soc. **11**, 22–30 (2014)
2. Tsai, H.C., Lin, Y.H.: Modification of the fish swarm algorithm with particle swarm optimization formulation and communication behavior. Appl. Soft Comput. **11**(8), 5367–5374 (2011)
3. Zhu, Z., Tang, J., Lambotharan, S., Chin, W.H., Fan, Z.: An integer linear programming based optimization for home demand-side management in smart grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–5. IEEE, January 2012
4. Shakouri, H., Kazemi, A.: Multi-objective cost-load optimization for demand side management of a residential area in smart grids. Sustain. Cities Soc. **32**, 171–180 (2017)
5. Anvari-Moghaddam, A., Monsef, H., Rahimi-Kian, A.: Optimal smart home energy management considering energy saving and a comfortable lifestyle. IEEE Trans. Smart Grid **6**(1), 324–332 (2015)
6. Moon, S., Lee, J.W.: Multi-residential demand response scheduling with multi-class appliances in smart grid. IEEE Trans. Smart Grid (2016)
7. Samadi, P., Wong, V.W., Schober, R.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
8. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in smart grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
9. Huang, Y., Wang, L., Guo, W., Kang, Q., Wu, Q.: Chance constrained optimization in a home energy management system. IEEE Trans. Smart Grid (2017)

10. Bharathi, C., Rekha, D., Vijayakumar, V.: Genetic algorithm based demand side management for smart grid. Wireless Pers. Commun. **93**(2), 481–502 (2017)
11. Basit, A., Sidhu, G.A.S., Mahmood, A., Gao, F.: Efficient and autonomous energy management techniques for the future smart homes. IEEE Trans. Smart Grid (2015)
12. Zhao, Z., Lee, W.C., Shin, Y., Song, K.B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)
13. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
14. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
15. Wang, H., Wang, Z., Domingo-Ferrer, J.: Anonymous and secure aggregation scheme in fog-based public cloud computing. Future Gener. Comput. Syst. (2017)
16. Busom, N., Petrlic, R., SebÃl', F., Sorge, C., Valls, M.: Efficient smart metering based on homomorphic encryption. Comput. Commun. **82**, 95–101 (2016)
17. Jiang, R., Lu, R., Choo, K.K.R.: Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data. Future Gener. Comput. Syst. (2016)
18. Jo, H.J., Kim, I.S., Lee, D.H.: Efficient and privacy-preserving metering protocols for smart grid systems. IEEE Trans. Smart Grid **7**(3), 1732–1742 (2016)
19. Ni, J., Zhang, K., Lin, X., Shen, X.S.: EDAT: efficient data aggregation without TTP for privacy-assured smart metering. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, May 2016

Extraction and Tracking of Scientific Topics by LDA

Yongjun Zhang[1,2(✉)], Jialin Ma[1,2], Zijian Wang[2], and Bolun Chen[1]

[1] Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, China
13511543380@139.com
[2] College of Computer and Information, Hohai University, Nanjing, China
zhjwang@hhu.edu.cn

Abstract. Scientific papers play an important role in the scientific research. Scientific researchers understand the research trends of their interested scientific topics. But the huge amount of papers makes them difficult to get a global view of scientific topics. In this paper, we proposed an automatic *LDA*-based method to extract scientific topics and track their evolutions. The method firstly divides the papers into some subsets according to the paper's published year, then extracts scientific topics for each subset. To tract the evolutions of topics and their relations, a directed graph is constructed in terms of *KL* distances. The experimental results on Amind ACM-Citation-network dataset shows that our method is reasonable.

1 Introduction

Scientific papers play an important role in the development of science and technology, they may be the most important source for scientific researchers to find the scientific and technical knowledge. By reading scientific papers, scientific researchers get insight into the following problems: (1) How did their working scientific topics develop? (2) What are the hot research problems in these topics? (3) Are these topics they still active and attractive? Recently the amount of scientific papers grows explosively, which has greatly help researchers in their works, but also lead researchers be lost in the massive of papers and difficult to get a global view of their research topics. So an automated method is required to extract scientific topics from the existing scientific papers and explore how they evolved.

The topic model is an effective and promising method to mine and analyze scientific topics from scientific literatures in the semantic level. Latent Dirichlet Allocation (*LDA*) [1], may be the most well-known topic model, models the scientific papers through a three-level topology, i.e. document-term-topic, only by the textual contents of papers without any supervised data. The document exchangeable assumption of the original *LDA* ignores the time sequence of papers, which makes it can't track the scientific topics development when it is applied to scientific papers directly. The topic over time (*TOT*) topic model incorporates the time label of each document to associate time with the probability distribution of document topics [2], Griffiths [3] and Hall [4] fit a *LDA* model in the whole paper set to obtain the global scientific topics, then use the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_48

published time of each paper to evaluate the distribution of each global scientific topic. In this way they can find how scientific topics evolve. All the above methods assume the scientific topics is fixed but has different intensities during different time windows. The continuous dynamic topic model (CDTM) [5] and dynamic mixture model (DDM) [6] divide papers into different windows according to the published time, then the paper set on each time window are processed sequentially, finally the evolution of the scientific topic over time is formed, but their research works can't find the relations among topics. The remarkable dynamic topic model (DTM) [7] proposed by Blei uses state space models on the natural parameters of the multinomial distributions that represent the topics. Variational approximations based on Kalman filters and non-parametric wavelet regression are developed to carry out approximate posterior inference over the latent topics. But the research work must predefine the topic number K and can't model the birth of new topics and the death of old topics.

In this paper, we develop a novel *LDA*-based method to extract and track the scientific topics from scientific papers. It has the following advantages:

(1) Unlike previous methods, this method doesn't need to fix the topic number parameter and can determine it automatically.
(2) Our method can explore the birth and death of scientific topics.
(3) We use a topic directed graph to represent the relations among topics.

2 Use of *LDA* for Extracting and Tracking Scientific Topics

2.1 Review *LDA*

LDA, which was firstly proposed by Blei in 2003, can automatically extract hidden semantic topics from text documents. The easiest way to understand *LDA* is regarding it as a hierarchical Bayesian mixed membership model with implicit variables. In this view, each membership, also called component or topic, is represented as a multinomial distribution over terms to denote some special semantic, each text document then can be represented as a mixture of topics, and each term in the document is associated with a hidden topic indicator variable to indicate which topic it wants to convey. A live example is illustrated in Fig. 1.

In Fig. 1, the left figure illustrates a topic extracted from scientific papers. Since a topic is a multinomial distribution over terms, we just list the top 5 terms with highest probabilities. The listed top terms strongly indicate this topic may be corresponded to the cryptography scientific topic. The right illustrates how a document are represented in the topic level, it can be seen that the example document is a mixture of 5 topics, in which the topic #2 has the highest proportion to suggest the document is about it mainly. The next topic #1 has a proportion 0.183 suggests the document have a little content talking about it. The negligible proportions of remaining topics indicate they are irrelevant to the document.

Term	probability
signature	0.00813
cryptographic	0.00672
string	0.00144
encrypt	0.000981
rsa	0.000899

Topic	proportion
#2	0.744
#1	0.183
#3	0.051
#4	0.018
#5	0.004

Fig. 1. An example of *LDA*. The left is a topic with top 5 terms, the right is a document represented as a mixture of 5 topics

Another way to explain *LDA* is viewing it as a generative probabilistic model and Bayes network to generate each document and each term of documents. The probability generate process of *LDA* is shown in Fig. 2:

(1) For each topic $k \in \{1,2,\ldots,K\}$

(2) Generate a topic ϕ_k from the *Dirichlet* distribution $\phi_k \sim Dir(\beta)$

(3) For each document $d \in D$

(4) Generate the topic mixture θ_d of *d* from the *Dirichlet* distribution $\theta_d \sim Dir(\alpha)$

(5) For each word $w_{dn} \in d$

(6) Generate a topic index z_{dn} from the multinomial distribution $z_{dn} \sim Multi(\theta_d)$

(7) Generate w_{dn} from the topic $\beta_{z_{dn}}$ $w_{dn} \sim Multi(\beta_{z_{dn}})$

Fig. 2. The probability generate process of LDA

Figure 3 illustrates the graph model representation of Bayes network of *LDA*:

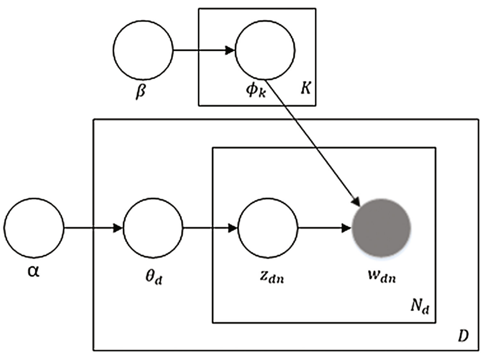


Fig. 3. Graph model representation of *LDA*

The notations listed in Figs. 2 and 3 are illustrated in Table 1.

Table 1. Notations of *LDA*

Notation	Meaning
K	The topic number
β	The parameter of *Dirichlet* prior distribution of topic-term distribution
ϕ_k	The topic-term distribution of topic k, where ϕ_{kv} is the probability of term v in topic k
D	The text document corpus
α	The parameter of *Dirichlet* prior distribution of topic-term distribution
θ_d	The document-topic distribution of document d, where θ_{dk} is the proportion of topic k in document d
N_d	The term count of document d
z_{dn}	The topic of the nth position in document d, it is drawn from the multinomial distribution $Multi(\theta_d)$, with a value range of from 1 to K
w_{dn}	The nth term in document d

Thomas L. Griffiths and Mark Steyvers [3] developed a *collapse Gibbs Sample* algorithm to inference the hidden variables θ_d, z_{dn} and ϕ_k as shown in the formulas (1)–(3):

$$P\left(z_{dn} = k \middle| \mathbf{z}_{-dn}, \boldsymbol{w}\right) \alpha \frac{n_{-dn,k}^{w_{dn}} + \beta}{n_{-dn,k}^{(\cdot)} + V\beta} \frac{n_{-dn,k}^{(d)} + \alpha}{n_{-dn,\cdot}^{(d)} + K\alpha} \tag{1}$$

$$\widehat{\phi}_{kv} = \frac{n_k^{(v)} + \beta}{n_k^{(\cdot)} + V\beta} \tag{2}$$

$$\widehat{\theta}_{dk} = \frac{n_k^{(d)} + \alpha}{n_{\cdot}^{(d)} + K\alpha} \tag{3}$$

Table 2 lists the notations in the formulas and their illustrations.

Table 2. Notations in formula (1), (2) and (3)

Notation	Illustration
$\mathbf{z}_{-dn}$	All the topic assignments for each term in the corpus, but not including the nth topic assignment in document d
$\boldsymbol{w}$	All the terms occurred in the corpus
$n_{-dn,k}^{w_{dn}}$	The count of term w_{dn} which has a topic assignment k in the corpus excluding the nth term of document d

(*continued*)

Table 2. (*continued*)

Notation	Illustration
$n_{-dn,k}^{(\cdot)}$	The count of all the terms having a topic assignment k in the corpus excluding the nth term of document d
$n_{-dn,k}^{(d)}$	The count of topic k in the document d excluding its nth position
$n_{-dn,\cdot}^{(d)}$	The term count of the document d excluding the nth position
$\widehat{\phi}_k^{(v)}$	The estimated value of ϕ_{kv}, which is the probability of term v for given topic k
$n_k^{(v)}$	The count of term v which has a topic assignment k
$n_k^{(\cdot)}$	The count of topic k in the corpus
V	The size of term table
$\widehat{\theta}_{dk}$	The estimated value of $\widehat{\theta}_{dk}$, which is the proportion of topic k in document d
$n_k^{(d)}$	The count of topic k in document d
$n_{\cdot}^{(d)}$	The term count of document d

2.2 Extraction and Tracking of Scientific Topics from Papers by *LDA*

2.2.1 Problem Formulation

We denote the papers used to extract and track scientific topics as D, for each paper $d \in D$, the published year of d is represented as y_d. We use the notation y_0 and y_M to denote the minimum and maximum of y_d, then y_k is the year $y_0 + k$. The paper set D then can be divided into M subsets $D_0, \ldots, D_M$, where $D_k = \{d | y_d = y_0 + k\}$ represents the papers published in the year $y_0 + k$. The problems we want to solve are as followings: (1) for a paper subset D_k, how many scientific topics are covered with these papers and what are they? (2) for each scientific topic ϕ_{kn} in D_k, where it which topic is its origin? (3) for two scientific topics ϕ_{kn} and ϕ_{lm}, is there some relation between them? (4) for a given topic ϕ_{kn} in the year $y_0 + k$, is it popular or inconspicuous?

2.2.2 Extract Scientific Topics from the Paper Subset

For each paper subset D_k, We use the textual content of each paper to learn a *LDA* model. The topic number parameter K_k used to fit the *LDA* for D_k are determined by the *KL* distance. The *KL* distance between two topics are defined as the formula (4):

$$J\left(T_i, T_j\right) = \frac{1}{2}\left(D(T_i||M) + D(T_i||M)\right) \tag{4}$$

Where $M = \frac{1}{2}\left(T_i + T_j\right)$, $D(T_i||M)$ is the *Kullback–Leibler divergence* from M to T_i, it is defined as:

$$D(T_i||M) = \sum\nolimits_{t=1}^{V} T_{i,t} log \frac{T_{i,t}}{M_t} \tag{5}$$

For a topic set $T = \{T_1, \ldots, T_n\}$, the minimal topic distance is defined as:

$$min_T = \min_{i,j \in \{1,\ldots,n\}} J\left(T_i, T_j\right) \tag{6}$$

The min_T is an indicator whether the discovered scientific topic granularity is appropriate or not. If it's low, the scientific topic granularity may be too fine and vice versa.

We initialize $K_k^{(0)} = K_{k-1}$, where K_{k-1} is the final topic number in the *LDA* model fitted by D_{k-1}, the algorithm 1 illustrates the process to determine the final *LDA* model of D_k:

Algorithm 1. The determining of the final *LDA* model of D_k

1: Initialize $K_k = K_k^{(0)}$
2: Fit a *LDA* model M_k with topic number K_k on the paper set D_k
3: Calculate the minimal topic distance topic distance min_{ϕ_k} of $\phi_k = \{\phi_{k1}, \ldots, \phi_{kK_k}\}$
4: if $min_{\phi_k} < \sigma_{min}$, set $K_k = K_k - 1$, goto step 2
5: else if $min_{\phi_k} > \sigma_{max}$, set $K_k = K_k + 1$, goto step 2
6: else output M_k

Where σ_{min} and σ_{max} are threshold parameters to control the topic granularity, it can be determined by experiment.

The problem (1) in the Sect. 2.2.1 can be solved by the Algorithm 1

2.2.3 Building of Topic Directed Graph

After we determine the *LDA* model M_k for each D_k, we can build a directed graph for all the topics. Considering the adjacent model M_{k-1} and M_k, for each topic ϕ_{kn} in M_k, there exists a nearest topic ϕ_{k-1m} in M_{k-1} according to the *KL* distance defined in the formula (4). A directed edge from ϕ_{k-1m} to ϕ_{kn} will be generated if $J(\phi_{k-1m}, \phi_{kn}) < \sigma$. The Fig. 4 illustrates an example topic directed graph:

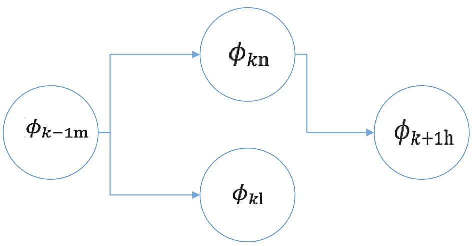


Fig. 4. An example of topic directed graph

In this example, the topic $\phi_{k-1\text{m}}$ evolves into ϕ_{kn} and ϕ_{kl} in the next year. The topic $\phi_{k-1\text{m}}$ originates from the topic $\phi_{k-1\text{m}}$ and has a latest source ϕ_{kn}. The topic ϕ_{kl} has no subsequent topics, so it can be regarded as dying in the year $y_0 + \text{k}$.

By means of building the topic directed graph for all topics of all paper subsets, we solve the problems (2) and (3) proposed in the Sect. 2.2.1.

2.2.4 Evaluation of the Interestingness of Scientific Topics

For a topic ϕ_{kn}, its interestingness refers to the degree of attention. The more papers it is focused on the higher its interestingness and vice versa. We define the interestingness of a topic as following:

$$F_{\phi_{kn}} = \frac{\sum_{d \in D_k} \theta_{d\phi_{kn}}}{|D_k|} \tag{7}$$

The formula (7) suggests that the interestingness of a topic is its average proportion of papers published in the year $y_0 + k$. It can solve the problem (4) proposed in the Sect. 2.2.1.

3 Experiments

3.1 DataSet

In our experiments, we investigated the Amind ACM-Citation-network V8 [8], which is a data set contains 2,381,688 papers published from 1958-2015.

We pruned the vocabulary by stemming each term to its root, removing function terms, and removing terms that occurred fewer than 10 times. A stop word list also is used to remove the stop words.

3.2 Experimental Results

Due to the limit of pages, we illustrate some topics about information security and their relations from the year 2008 to 2010 in the Table 3 and Fig. 5.

Table 3. The topics about information security

Topic	Top 5 terms
2008#2	Signature, cryptographic, string, encrypted, primitive
2009#14	Cryptography, pseudorandom, proven, stronger, convexity
2009#38	Security, attack, authentication, intrusion, trusted
2009#41	Password, system, payment, authentication, vulnerability
2010#4	Security, encryption, hash, cryptographic, robustly
2010#17	Key, protect, scheme, policies, credential

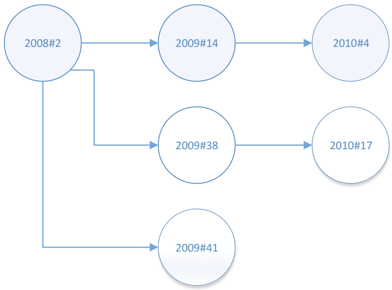


Fig. 5. The topic directed graph of topics about information security from 2008 to 2010

It seems that the topics 2008#2, 2009#14 and 2010#4 are about cryptology, 2009#38 and 2010#17 are about security protection, the topic #41 is about authentication. The topic evolutions and the relations illustrated in the Fig. 4 seems reasonable.

4 Conclusions

The method we proposed in this paper can extracts scientific topics from papers with the help of *LDA*. The topic directed graph building by *KL* distances of topics in adjacent time windows shows the evolutions of topics and their relations, it also illustrates the life circle of topics: when a topic is birth, when it reaches peak and when it is died. Some further improvements include: represent each topic as a multinomial distribution over both words and phrases rather than only words, take the correlation between topics of the same time window into account.

Acknowledgments. This work was supported by the Chinese National Natural Science Foundation (Grant No.: 61602202), the Natural Science Foundation of Jiangsu Province, China (Grant No.: BK20160428), the Social Key Research and Development Project of Huaian, Jiangsu, China (Grant No.: HAS2015020) and the Graduate Student Scientific Research and Innovation Project of Jiangsu Province, China (Grant No.: 2015B38314).

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
2. Wang, X., Mccallum, A.: Topics over time: a non-Markov continuous-time model of topical trends. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 424–433. ACM (2006)
3. Griffiths, T.L., Steyvers, M.: Finding scientific topics. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 101(Suppl 1.1), p. 5228 (2004)
4. Hall, D., Jurafsky, D., Manning, C.D.: Studying the history of ideas using topic models. In: Conference on Empirical Methods in Natural Language Processing, pp. 363–371 (2008)
5. Wang, C., Blei, D., Heckerman, D.: Continuous time dynamic topic models. In: Mcallester, D., Nicholson, A. (eds.) Uncertainty in Artificial Intelligence, pp. 579–586 (2009)

6. Wei, X., Sun, J., Wang, X.: Dynamic mixture models for multiple time series. In: International Joint Conference on Artifical Intelligence, pp. 2909–2914. Morgan Kaufmann Publishers Inc. (2007)
7. Blei, D.M., Lafferty, J.D.: Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning. ACM (2006)
8. Tang, J., et al.: ArnetMiner: extraction and mining of academic social networks. In: ACM SIG KDD International Conference on Knowledge Discovery and Data Mining, pp. 990–998 (2008)

Mining Unknown Network Protocol's Stealth Attack Behavior

Yan-Jing Hu[1,2(✉)]

[1] Key Laboratory of Cryptology and Information Security, Chinese PLA, Engineering University of the Armed Police Force, Xi'an 710086, China
huyanjing2007@126.com

[2] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China

Abstract. Unknown network protocol's Stealth Attack behavior is becoming a new vehicle to conduct invisible cyber attacks. This kind of attack is latent for a long time, and the execution is triggered only under special conditions. As the application layer protocols flourish, there is an exponential increase in the number and diversity of the protocol stealth attack behaviors. Even though numerous efforts have been directed towards protocol reverse analysis, the unknown protocol's stealth attack behavior mining has rare studied. This paper proposes a method of mining stealth attack behavior by instruction clustering. First, all protocol samples are divided into functional instruction sequences. Then clustering analysis of all the functional instruction sequences using the instruction clustering algorithm. The stealth attack behavior instruction sequences can be mined quickly and accurately according to the calculation of the behavior distance. Experimental results show that our solution is ideal for mining protocol's stealth attack behavior in terms of efficiency and accuracy.

Keywords: Protocol reverse · Clustering analysis · Stealth attack behavior

1 Introduction

Network protocol's stealth attack behavior, especially with harmful intents to target computers and networks is rapidly becoming one of the major threats to network security [1]. Over the past years, the protocol's stealth attack behavior has evolved dramatically from simple activities to a lot of complex and concealed malicious behaviors [2]. The protocol developers always sell their products and tools, and the users can convenient and rapidly access tens of thousands of infected hosts for nefarious purposes [3]. A comprehensive study showed that a large number of Internet-connected computers is infected with stealth attack behaviors [4–6], and on average, each detected host has at least 15 different stealth attack behavior programs installed [7].

Different from the normal protocol behavior, such as send or receive messages, the goal of stealth attack behavior is usually not to cause obvious damage or to spread itself to other systems [8].

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems,*
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_49

Instead, protocol's stealth attack behavior is always waiting specific conditions, monitor the network behaviors and steal sensitive or private information, such as important files and original data [2, 9, 19]. Such information is then sent back to the protocol designer or remote attackers, and used as a basis for attack targeted hosts or networks, or analyzing in-depth. Protocol's stealth attack behavior can also lead "hijack" a user's browser, and direct the unsuspecting user to a specific web site that the protocol designer or attacker choosing. In addition to the violation of users' sensitive information and privacy, protocol's stealth attack behaviors are also responsible for the degradation of system performance, because they are always running secretly [11, 12].

Protocol's instruction sequences are able to provide the most intimate details of a protocol's dynamic behavior. They can be used for protocol behavior analysis, protocol failure diagnosis, collecting protocol program metrics like test prioritization and coverage, etc., especially for protocol's stealth attack behavior analysis and mining [12]. There exist two major obstacles in protocol's stealth attack behavior analysis and mining when they are based on the execution of instruction sequences. First, running a protocol binary or executing a protocol program can hardly discover stealth attack behaviors, because they are always concealed or disguised in an invisible state. Second, how to express a protocol's behavior, and how to collect and analysis of protocol's behaviors is also a difficult issue. In this paper, we use instruction sequences to express protocol's behaviors. Obviously, there are a large number of instruction sequences in a protocol binary. The applicability of process such large number of instruction sequences is also a difficult question.

In this paper, we develop a dynamic binary analysis combined with instruction clustering analysis techniques to mine the unknown protocol's stealth attack behavior and security-relevant functionality from the protocol binaries. We demonstrate that our techniques enable previously unsolved security applications, such as protocol behavior representation, encrypted instructions discovering and decrypting, and enable more accurate solutions for other important security applications such as division and labeling of gene instructions, vulnerability mining, detecting and mining attacks on web applications, and protocol's stealth attack behaviors mining and analyzing.

Our framework consists of three main components: instruction sequences extracting, genetic instruction labeling and instruction clustering analysis. The instruction sequences extracting engine uses a combination of dynamic binary interpretation and analysis to emulate the instruction level of a protocol's behavior. The instruction sequences extracting engine extract the binary code of behavior instruction sequences from the protocol program binaries, so that all the extracted instruction sequences can be reused by external source code. Instruction sequences reuse is necessary when the behavior instruction sequences to be reused is only available in binary form. It is useful when the behavior instruction sequences are complex but the application does not require a low level understanding of how the instruction sequences works. We only reusing their functionalities. In this paper we use our proposed instruction sequences reuse techniques for extracting the cryptographic functions and keys used by protocols to protect their stealth attack behaviors. We can deploy the extracted behavior instruction sequences in an intrusion detection system (NIDS), enabling the NIDS to decrypt the encrypted traffic.

The genetic instruction labeling component label the extracted instructions with three genetic instructions, they are F (function call related instructions), C (conditional jump related instructions) and D (data process related instructions). The genetic instruction labeling is important because such a step enables automatic reasoning about the security properties of the extracted behavior instruction sequences. In this paper we use the labeled instructions to automatically find stealth attacks on unknown protocols, to identify deviations between different protocols of the same functionality, and to generate stealth attack signatures for protocols. The genetic instruction labeling techniques are widely applicable, they can be applied to protocol samples when the source code is not available. It is important because closed-source protocols are prevalent in network and only distributed in binary form. In addition, the techniques do not require cooperation from the protocol designers. It is important because the protocol designers may not support the security analysis of their protocols.

The protocol's stealth attack behavior is a large class of closed-source protocols where it is important not to rely on the protocol designers. Our method enables users to analyze the closed-source protocols they are deploying for security issues. The instruction clustering analysis have high fidelity because they analyze the binary, no matter the instruction sequences are executed or not. The protocol binary is a lower abstraction representation of the protocol, it enables the analysis of security issues that may be hidden in the protocol binary and security issues related to the memory layout of the unknown protocol. The extracted instructions include attributes of the instructions being executed, such as the address of the instructions, events like thread creation, exceptions, module loads, etc. If there are memory-accessing instructions, the address of the read or written value of a memory location also included. The instruction clustering analysis does not require source code or additional symbolic information for stealth behavior discovery. After instruction clustering, it can execute dynamically generated instruction sequences for regression test and analysis of the mined stealth attack behavior.

Researchers and commercial security companies constantly improve their techniques to detect and prevent protocol's stealth attack behaviors, while the protocol designers routinely improve their techniques to evade detection. The use of binary packing and encryption to prevent straight forward, signature-based detection is an example. We provide evidence that our tool effectively locates protocol's stealth attack behavior instructions for various protocol samples. We also present a case study of the ability of our tool to mine the stealth attack behavior in both network communication and protocol binary, it is a real bot (C&C) protocol sample. We argue that our approach can be used as a tool for revealing the intent of attackers that try to masquerade their activities using protocol's stealth attack behaviors.

2 Related Work

In this section, we discuss related work that gives background on the general problem of protocol behavior analysis and the specific problem addressed in this paper. Generally, protocol's behavior analysis encompasses two main techniques: static and dynamic analysis. The static analysis is a white-box approach in which the target protocol sample is disassembled to enable an analyst to understand its whole functionalities and the structure of instruction sequences [13]. Early work in static malware analysis has tended to focus on approaches to malware detection [14]. In this analysis, the main problem is how to disassemble the executables because most of the malicious protocol binaries are obfuscated by great variety of packers [15]. The semantic behavior of codes is analyzed by Christodorescu et al. as a way to thwart some binary obfuscation methods [9]. Moser et al. considered that the limitations of static analysis can be evaded by advanced semantics-based malware detectors [16]. This means that static analysis alone is no longer sufficient for malware identification. On the other hand, the dynamic analysis is a black-box approach, in which the protocol binary samples are executed in an environment that is designed to closely observe the protocol's internal activities in detail. For example, the loaded DLLs, created processes, the file and registry access, the API call sequence, and external activities, such as scan activities or information collections are observed. Recent research around automatic protocol reverse analysis have been biased towards dynamic analysis. The main drawback of dynamic analysis is that it only observes a single execution path. However, it still can extract a lot of information about the behavior of the protocols.

Dynamic analysis needs to execute the protocol program in a protected virtual environment [17]. During the execution process, the protocol's dynamic behavior characteristics, such as network characteristics, file system and the access to the registry are analyzed to get a favorable basis for the detection and analysis of protocol's stealth attack behaviors. In dynamic analysis, CwSandbox [18] system apply API Hooking technology to monitor dynamic behavior of protocol program [19]. As API Hooking works on the user level, if stealth attack behavior requests system kernel for service calls, it can avoid API Hooking analysis. Norman Sandbox [20] adopts simulated operating system compatible with Windows for executing protocol binary. In order to capture unknown protocol's behavior, Norman Sandbox also creates a virtual P2P network. Unfortunately, the emulator is less stable than the virtual machine, and the incomplete emulation of information usually leads to information loss and system crash. The protocol behavior analysis mainly includes behavior-based analysis and content-based analysis [21]. Behavior-based analysis methods can get comprehensive information of the protocol's behavior and are semantically strong. But they usually need to keep a large number of state information and it always spend more time. The content-based methods are able to provide signature for network detection, but they could not get network behavior profile.

The above mentioned dynamic analysis tools or systems focus exclusively on the monitor protocol's internal behavior in the infected host or network rather than on their external behavior. This is appropriate for host-based security because the internal behavior that is observed can be used for detecting malicious behavior or for making detection tools. However, protocol's stealth attack behavior is different from the normal malicious behavior. Our approach tries to be more generic and not assume the use of any particular method to protect protocol's stealth attack behavior.

3 Mining System of Network Protocol's Stealth Attack Behavior

3.1 Extracting of Protocol's Instruction Sequences

In recent years, we have captured a large variety of protocol samples, but most of the protocol's behaviors are unknown. As most protocols adopt encryption, obfuscation and other techniques to adopt re-reverse analysis, traditional static analysis and dynamic analysis are very difficult to mine protocol's stealth attack. With the help of our own self-developed virtual analysis platform HiddenDisc, all behavior instruction sequences of all protocol samples are extracted and analyzed. HiddenDisc is designed for automatic reverse analysis of protocol's unknown behavior. It is based on the virtual platform TEMU for second development, the extension adds 4 major functional components to mine the protocol's stealth attack behaviors. The framework is shown in Fig. 1.

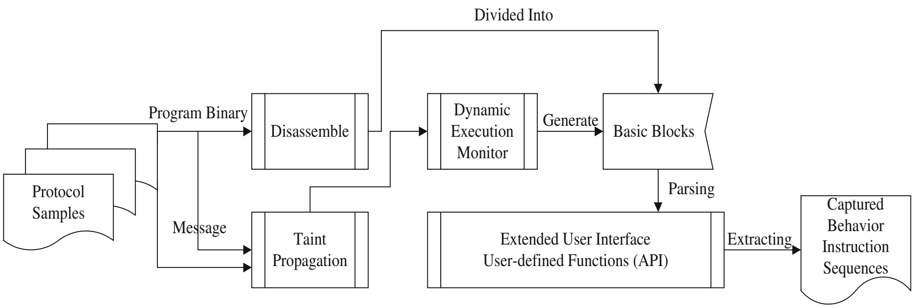


Fig. 1. Framework of capturing protocol's behavior instruction sequences

The analysis of protocol samples can be carried out through 2 technical lines, one is to disassemble the binary protocol program, and further divided into basic blocks; the other is to mark each byte of the protocol message for taint propagation analysis. If the protocol samples are protected by encryption or obfuscation method, the disassembly static analysis is hard to work, then the protocol samples should be executed and analyzed in dynamic analysis environment. In the procedure of dynamic analysis, the dynamic execution monitor component looks into how the program binary parses the protocol message, and how every bit of the tainted data is propagated. Under the help of dynamic execution monitor component, all the basic blocks that related to the message parsing processes are generated. As some system calls and meaningless blocks of code are included, the number of original basic blocks may be very large. So the core instruction sequences should be extracted from the basic blocks to improve the efficiency of security analysis.

Protocol analysts have to put more efforts into reexamining the protocol samples which may have been investigated before. 1297 protocol samples' 3200 behavior instruction sequences are mined and implement a compiler-level prototype to automate analysis of the instruction sequences. The protocol samples we collected contains instruction sequences dependency graphs generated for 1297 protocol samples and covers a large variety of stealth attack behaviors, such as bot (C&C) protocol, backdoor, Trojan, worm and virus binaries.

3.2 The Instruction Clustering Analysis Method

The stealth attack protocol sample set includes behavior instruction sequences extracted from 1638 malicious protocols, and more than 78% samples are the variants of various bot C&C protocols. We calculate the most popular instruction sequences dependencies, OS operations and dependencies. The most instruction sequences are mainly related to the operations on Windows file system, such as access to registry and memory. Other instruction sequences stand for various OS objects, such as registry, file operation, process, and memory-mapped file, etc. We believe as long as we diversify these popular dependencies and behavior features, the similarity among protocol binaries and their variants can drop significantly.

To extract protocol's binary information, the protocol binary sample files are disassembled first, using disassembling tools. After assembly instructions are extracted using an instruction capture engine, the instruction sequence of assembly instructions is divided into basic blocks according to some instructions that are used as delimiters. The opcodes of each instruction included in individual basic blocks is used as the protocol's binary information.

Protocol's behaviors can be seen as the behavior instruction sequences. The extracted behavior instruction sequences may be only a part of the protocol's behaviors. Protocol's stealth attack behaviors can be hidden in other instruction sequences which have not been discovered. The instruction clustering analysis is able to identify and discover a lot of varied abnormal behaviors in unknown protocols. Instruction Clustering is the process that partitioning all behavior instruction sequences no matter discovered or not into groups of similar behaviors. Each group, called a behavior cluster, is consists of some behavior instruction sequences that are similar with respect to a certain similarity measure, and which are dissimilar is belong to behavior instruction sequences of other groups.

The instruction clustering analysis component is trained by numerous protocol samples with their extracted instruction sequences. The component is used to predict and mine new behaviors of protocol instances. We call a particular clustering algorithm with a specific view of the data a cluster. Each cluster outputs a clustering or labeling, comprising the group labels for some or all objects. In this paper, we propose using combination of static analysis, dynamic and instruction clustering analysis to detect and mine protocol's stealth attack behaviors before the behavior executed locally.

In order to prevent any obfuscation technique poses obstacles to the scanning and detection module, an automating process for identifying and extracting stealth attack behavior instruction sequences is proposed. An algorithm for identifying if a protocol binary is applied any obfuscation mechanism is designed for the analysis platform. No matter what packing methods or how many hidden layers are applied, the original protocol binary code and data should eventually be present in memory to be executed. Taking advantage of this inevitable nature of stealth attack behaviors, we propose a module to extract the stealth attack behavior instruction sequences from the protocol binary samples. Stealth attack behavior mining function is built on top of instruction clustering components and it is developed to analyzing the protocol programs, and determines if the protocol binary hidden with any stealth attack behaviors. If the stealth attack behaviors are discovered, the corresponding instruction sequences are analyzed, especially monitoring and analyzing for common potential malicious behaviors such as replication, file overwrites, and attempts to hide the existence of the suspicious instruction sequences. If one or more stealth attack-like behaviors are mined, the suspicious protocol is flagged as a potential stealth attack protocol.

4 Experimental Results and Validation

The purpose of this empirical study is to investigate the impact of considering the dynamic analysis and instruction clustering analysis on the mining capability of the system. Of the 32 days of recorded audit data, the first 19 days were used for training the protocol's behavior models. Thresholds were computed using the following 2 days of audit data, and detection was performed on the final 11 days. The behavior models are trained for each protocol samples, but each behavior instruction sequence is evaluated not only on the model for its native context, but on all non-native models as well. If normal behavior models capture the context-specific features, we would expect the protocol's behaviors to be clustering as normal in their native context, and as anomalous in all other contexts.

Before the experiment, we have already prepared 25 captured behavior instruction sequences, which representing 25 different protocol behaviors. In order to effectively achieve cluster analysis, the 25 typical behavior instruction sequences are used as initial behavior models, such as http, ftp, DNS and SMB etc. These known behaviors are the foundation of protocol instruction clustering analysis. Using HTTP as an example, we capture and record the instruction sequences about how the Apache server processes a http get request for the file "index.jsp", and the reply generated by the server. In order to understand the content of the FTP protocol, we have analyzed and extracted the instruction sequences of the messages which sent by the FileZilla server in response to a connection, as well as the sent messages when the username and password are received. An example for SMB, we have analyzed and extracted the behavior about a Negotiate Protocol Request received by the Sambad open source server. In addition, we have also analyzed some common and typical stealth attack behavior instruction sequences, such as keyboard sniffing, password sniffing, backdoor accessing, rootkiting and spying etc. With 172 s, the 1297 unknown protocol samples are executed on HiddenDisc one after another automatically. All between the two behavior distance is computed according to the instruction clustering algorithm. Finally, by contrasting with the 25 original basic behaviors, 193 different behavior clusters are automatically generated.

Under the help of instruction clustering, 193 potential stealth attack behavior instruction sequences are extracted. Finally, we mined 187 stealth attack behaviors from the 193 instruction sequences by regression test.

5 Conclusions

In this paper, we presented a novel approach to the mining of protocol's stealth attack behaviors. Different from previous approaches, our solution combines dynamic binary context analysis with the characterization of instruction clustering analysis. The instruction clustering analysis is effective against protocol's stealth attacks, which do not modify the initial behavior instruction sequences that may be executed in certain situations. The dynamic binary context analysis is used to extract protocol's public behavior and instruction clustering is used to mine the protocol's stealth attack behavior which has not executed. Under the help of instruction clustering analysis, the protocol's stealth attack behavior instruction sequences are mined automatically. The use of instruction clustering analysis is orthogonal with respect to analysis techniques that characterize protocol's behavior instruction sequences. In future work, we will explore how the two approaches can be composed to achieve even more precise mining and better resilience to stealth attacks.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant no. 61103178, 61373170, 61402530, 61309022 and 61309008.

References

1. Harale, S.T.A.: Detection and analysis of network & application layer attacks using honey pot with system security features. Int. J. Adv. Res. Ideas Innovations Technol. **3**, 1–4 (2017)
2. Singh, S., Sharma, P.K., Moon, S.Y., et al.: A comprehensive study on APT attacks and countermeasures for future networks and communications: challenges and solutions. J. Supercomput., 1–32 (2016)
3. Albanese, M., Battista, E., Jajodia, S.: Deceiving attackers by creating a virtual attack surface. Cyber Deception. Building the Scientific Foundation, 167–199 (2016)
4. Damopoulos, D., Kambourakis, G., Gritzalis, S.: iSAM: an iPhone stealth airborne malware. In: Camenisch, J., Fischer-Hübner, S., Murayama, Y., Portmann, A., Rieder, C. (eds.) SEC 2011. IAICT, vol. 354, pp. 17–28. Springer, Heidelberg (2011). doi: 10.1007/978-3-642-21424-0_2
5. Filiol. E.: Viruses and malware. In: Handbook of Information and Communication Security, pp. 747–769 (2010)
6. Jung, S., Kim, J., Cagalaban, G., Lim, J.-h., Kim, S.: Design of cyber attack precursor symptom detection algorithm through system base behavior analysis and memory monitoring. In: Kim, T.-h., Vasilakos, T., Sakurai, K., Xiao, Y., Zhao, G., Ślęzak, D. (eds.) FGCN 2010. CCIS, vol. 120, pp. 276–283. Springer, Heidelberg (2010). doi:10.1007/978-3-642-17604-3_33
7. Lee-Urban, S., Whitaker, E., Riley, M., et al.: Two complementary network modeling and simulation approaches to aid in understanding advanced cyber threats. In: Advances in Human Factors in Cybersecurity: Proceedings of the AHFE 2016 International Conference on Human Factors in Cybersecurity, pp. 401–414, Walt Disney World®, Florida, USA, 27–31 July 2016
8. Ming, J., Xin, Z., Lan, P., et al.: Impeding behavior-based malware analysis via replacement attacks to malware specifications. J. Comput. Virol. Hacking Tech., 1–15 (2016)

9. Pawlowski, A., Contag, M., Holz, T.: Probfuscation: an obfuscation approach using probabilistic control flows. In: Caballero, J., Zurutuza, U., Rodríguez, Ricardo J. (eds.) DIMVA 2016. LNCS, vol. 9721, pp. 165–185. Springer, Cham (2016). doi: 10.1007/978-3-319-40667-1_9
10. Payer, M.: HexPADS: a platform to detect "stealth" attacks. In: Caballero, J., Bodden, E., Athanasopoulos, E. (eds.) ESSoS 2016. LNCS, vol. 9639, pp. 138–154. Springer, Cham (2016). doi:10.1007/978-3-319-30806-7_9
11. Teixeira d'Aguiar Norton Brandão, L., Neves Bessani, A.: On the reliability and availability of replicated and rejuvenating systems under stealth attacks and intrusions. J. Braz. Compu. Soc. **18**(1), 61–80 (2012)
12. Xiangjian He, T.C., Nanda, P., Tan, Z.: Improving cloud network security using the tree-rule firewall. Future Gener. Comput. Syst. **30**, 116–126 (2014)
13. Han, K., Kang, B., Im, E.G.: Malware analysis using visualized image matrices. Sci. World J. **2014 (**2014). 132713
14. Santos, I., Brezo, F., Nieves, J., et al.: Idea: opcode-sequence-based malware detection. Eng. Secure Softw. Syst., 35–43 (2010)
15. Saxe, J., Mentis, D., Greamo, C.: Visualization of shared system call sequence relationships in large malware corpora. In: Proceedings of the Ninth International Symposium on Visualization for Cyber Security, pp. 33–40 (2012)
16. Canfora, G., Iannaccone, A., Visaggio, C.: Static analysis for the detection of metamorphic computer viruses using repeated-instructions counting heuristics. J. Comput. Virol. Hacking Tech. **10**(1), 11–27 (2014)
17. Egele, M., Scholte, T., Kirda, E., et al.: A survey on automated dynamic malware-analysis techniques and tools. ACM Comput. Surv. **44**(2), 1–42 (2012)
18. Anderson, B., Storlie, C., Lane, T.: Improving malware classification: bridging the static/dynamic gap. In: Proceedings of the 5th ACM workshop on Security and Artificial Intelligence, pp. 3–14 (2012)
19. Hu, X., Shin, K.G.: DUET: integration of dynamic and static analyses for malware clustering with cluster ensembles. In: Proceedings of the 29th Annual Computer Security Applications Conference, pp. 79–88 (2013)
20. Kang, B., Kim, T., Kwon, H., et al.: Malware classification method via binary content comparison. In: Proceedings of the 2012 ACM Research in Applied Computation Symposium, pp. 316–321 (2012)
21. Kolter, J.Z., Maloof, M.A.: Learning to detect and classify malicious executables in the wild. J. Mach. Learn. Res. **7**, 2721–2744 (2006)

The 3rd International Workshop on Theory, Algorithms and Applications of Big Data Science (BDS-2017)

Research on Behavior Modeling and Property Verification for Cyber-Physical Systems

Mingfu Tuo[1(✉)], Cheng Zhou[2(✉)], Zhonghai Yin[2], Xin Zhao[2], and Lei Wang[2]

[1] School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China
mftuo@163.com

[2] Air Force Engineering University, Xi'an, China
zhou417461659@163.com

Abstract. Cyber-Physical System has been characterized by deep integration of computing and physical process. So it is difficult to model Cyber-Physical System and to analyze its properties. Thus, events are usually as a natural way to connect computing and physical process. In order to formally describe the constitution of composite events, an algebra of events is introduced for complex events process. And an approach of extend automata is proposed for behavior modeling Cyber-Physical System. At last, a simulation of lunar rover is introduced to verify its performance and safety of lunar rover.

1 Introduction

CPS are new type of hybrid systems which characterized by deep integrations of computation with physical processes [1]. CPS has attracted huge interest for widely fields, such as intelligent system, wireless sensor networks, the smart-grid, aeronautics and cloud computing, and so on [2–6].

As a result of integrate computation, communication and control, CPS requires rigorous interactions between the cyber and physical worlds both in time and space [7]. Hence, events are treated as a significant way to CPS. Recently, complex event process (CEP) plays an important role for the system based on event-driven. Thus CEP are widely used in fields of internet of thing, database, and so on [8]. However, as events in CPS has properties of temporal and spatial, some new technologies of CEP need to be obtained to suit for more complex and intelligent CPS.

Also, safety is another key property for CPS to be applied in critical application fields [9, 10]. Simulations could help us to examine the behavior of all system parts under normal as well as under exceptional and failure conditions, to investigate cascading effects and to test various scenarios [11]. Thus, new modeling is needed to describe the interaction process between computation and physical world for CPS.

In this paper, we propose an algebra of events (*AOE*) to describe the process of composite events in CEP. Also, we present an extend automata model to better understand the transition among several states through actuator in CPS based on event-driven. At last, a simulation of lunar rover is introduced to verify the

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_50

performance. Result shows that the automata model can help us understand the execution of simulation.

2 The Algebra of Event

In CPS, events are regarded as a significant means to combine computational and physical process. An arbitrary composite event can be described by the synchronization of several atomic events that it involves [12]. Here are some notions.

Definition 1 Event. An event is the occurrence of interest for system, denoting $E_{Phy}[attr, tim, loc]$. *attr*, *tim* and *loc* are the attribute, occurrence of time and location, respectively.

Definition 2 Atomic Event. The atomic event is the primary event that can not be divided any more. A set of atomic events can be denoted $E = \bigcup_{i=1}^{n} e_i = \{e_1, e_2, \cdots, e_n\}$.

This section formally describes the algebra of events for composite events. The algebra of events has two operators of "+" and "·", called union and synchronization respectively. In the following, we specify the algebra.

Let E be a set of atomic events. The syntax definition of the Algebra of Events (*AOE*) is described as follows

$$x ::= 0|1|e|x+x|x \cdot x|(x) \tag{1}$$

where special elements $0, 1 \notin E$. Additionally, "·" has a stronger constraint than "+". It should be note that "·" could be omitted without confused, i.e., $e_1 \cdot e_2$ simplifies as $e_1 e_2$.

It is clear that the composite operations "+" and "·" satisfy the following axioms. (1) " $+$ " is idempotent, associative, and commutative. Also, 0 is identity element for it. (2) "·" is idempotent, associative, and commutative. Also, 1 and 0 are identity and absorbing element respectively; synchronization "·" distributes over union.

The semantics of *AOE* is obtained by the function $|| \cdot ||$, defined by:

(1) $||0|| = \phi$, $||1|| = \{\phi\}$;
(2) $\forall e \in E$, $||e|| = \{\{e\}\}$;
(3) $\forall x_1, x_2 \in AOE$, $\|x_1 + x_2\| = \|x_1\| \cup \|x_2\|$;
(4) $\forall x_1, x_2 \in AOE$, $\|x_1 \cdot x_2\| = \{\alpha \cup \beta | \alpha \in \|x_1\| \wedge \beta \in \|x_2\|\}$;
(5) $\forall x \in AOE$, $\|(x)\| = \|x\|$.

3 Extended Automata for CPS

In this section, an extended automata is formally described based on the basis of *AOE*, which provides a valid method to understand the transitions generated by composite events in CPS.

Definition 3 Component. A component can be described as a tuple

$$COM = [E, Q, T, \{C_\tau\}_{\tau \in T}, \{A_\tau\}_{\tau \in T}], \tag{2}$$

where E is a set of atomic event. Q is the collection of states. And $T \subseteq Q \times E \times Q$ is the transitions in COM.$\{C_\tau\}_{\tau \in T}$ is the transition conditions.$\{A_\tau\}_{\tau \in T}$ is the activities for transition.

Definition 4 Composite Event. A composite event f is the combination of atomic events through the Algebra of Events (AOE) in CEP engine, denoting by (E, γ).

- E is the set of atomic events in component COM_i;
- $\gamma = \{+, \cdot\}$ is the operators of union and synchronization in AOE;

Definition 5 System. A system is denoted by $(\Gamma(\{COM_i\}), Cond_i, Act_i, Init), i \in [1, n] = I$. $\Gamma(\{COM_i\})$ are components in systems. $Cond_i$ is transition conditions. And Act_i is a set of activities for each state. $Init$ is set of initial states $q = (q_1, \cdots, q_n)$. The transition can be described as $(Q, E_{In}, \rightarrow)$,

- $Q = \prod_{i=1}^{n} COM_i.Q$ represents all states in systems;
- $E_f = \cup f_i$ represents composite events of all components;
- $\rightarrow$ represents the state transition. It means that a transition will be generated only if the relevant composite event is obtained; otherwise, the state would not change.

4 Case Study

In this section, a lunar rover will be described and specified with the model of automata. In the scenario of lunar rover, sensors will detect surroundings to find obstacles. Some events will be processed by CEP and will send to computation unit to compute appropriate acceleration and decisions. Nest, actions are executed by actuators. The work process is depicted by Fig. 1.

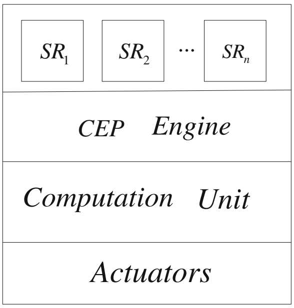


Fig. 1. The work process for lunar rover

Furthermore, we use extended automata to describe the interaction between computation and physical unit for lunar rover. Through the analysis about the scenario of lunar rover, the automata model can be constructed by Fig. 2.

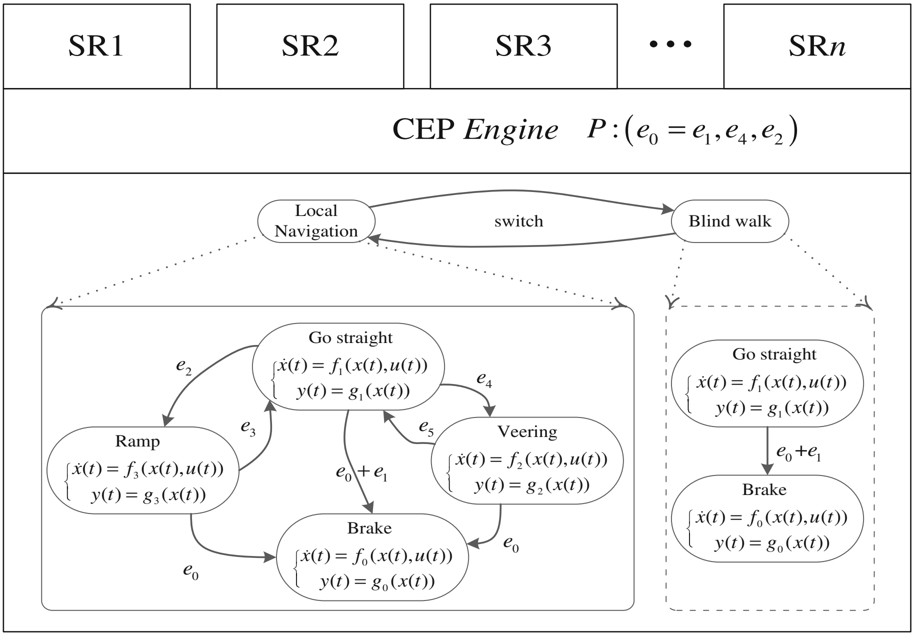


Fig. 2. The automata model for lunar rover

From Fig. 2, sensors are utilized to detected environment. All kinds of information will be processed by CEP engine to generate those events that are significant for systems. Also, there are two kinds of modes which are switched by control rods, called Local Navigation and Blind Walk. The mode of Local Navigation is utilized to control lunar rover to complete the precise behavior, such as ramp, obstacle avoidance. However, Blind Walk just be utilized to move lunar rover on a flat and straight road. It is clear that Blind Walk can be regarded as a particular case for Local Navigation. Thus, we just specify the mode of Local Navigation to avoid repetition.

Our experiment is implemented under the MATLAB. There are 5 obstacles, and the object point is (80, 80). The results are shown in Fig. 3 where the mobile lunar rover can reach to the object point (80, 80) with the obstacle avoidance. Thus the simulation result shows that the performance of this method is available.

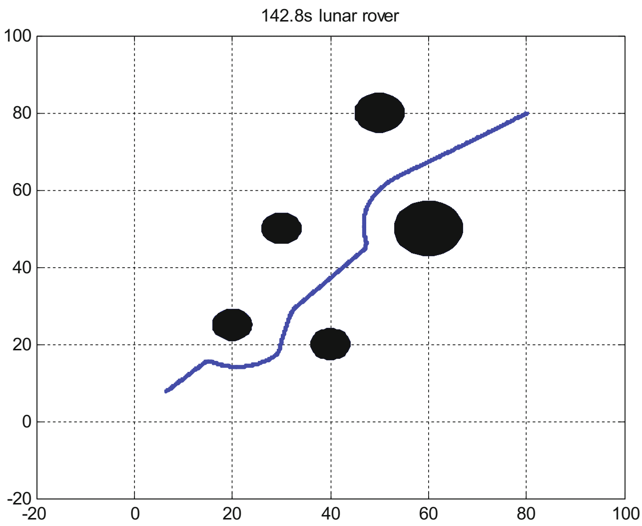


Fig. 3. Performance of lunar rover in MATLAB

5 Conclusions

In this paper, we proposed an algebra of events (AOE) to describe the constitution of composite events through two operators of "+" and "·", called *union* and synchronization respectively. Moreover, an extend automata model also is proposed to specify the transitions based on composite event in CPS. At last, a sample of lunar rover is introduced to verify the performance of the automata model. The simulation result shows that the performance of this method is available.

References

1. Klein, R., Rilling, S., Usov, A., Xie, J.: Using complex event processing for modelling and simulation of cyber-physical systems. Infrastructures **9**(1/2), 148–172 (2013)
2. Mola, L., Rossignoli, C., Carugati, A., Giangreco, A.: Business intelligence system design and its consequences for knowledge sharing. Collab. Decis. Making Explor. Study **11**(4), 1–25 (2015)
3. Mezghani, K., Ayadi, F.: Factors explaining IS managers attitudes toward cloud computing adoption. Int. J. Technol. Hum. Interact. **12**(1), 1–20 (2016)
4. Yuriyama, M., Kushida, T.: Integrated cloud computing environment with IT resources and sensor devices. Int. J. Space-Based Situat. Comput. (IJSSC) **1**(2–3), 174–182 (2011). doi:10.1504/IJSSC.2011.040342
5. Boyinbode, O., Le, H., Takizawa, M.: A survey on clustering algorithms for wireless sensor networks. Int. J. Space-Based Situat. Comput. (IJSSC) **1**(2–3), 137–150 (2011). doi:10.1504/IJSSC.2011.040339
6. Di Stefano, A., Morana, G., Zito, D.: QoS-aware services composition in P2PGrid environments, pp. 148–155. doi:http://dx.doi.org/10.1504/IJGUC.2011.04060

7. Lee. E.A.: Cyber physical systems: design challenges. In: The 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing, ISORC-2008, 5–7 May 2008, Orlando, Florida, USA, pp. 363–369 (2008)
8. Lee, E.A.: The past, present and future of cyber-physical systems: a focus on models. Sensors **15**, 4837–4869 (2015)
9. Guyeux, C., Makhoul, A., Atoui, I., Tawbi, S., Bahi, J.M.: A complete security framework for wireless sensor networks: theory and practice. Int. J. Inf. Technol. Web. Eng. **10**(1), 47–74 (2015)
10. Saripalle, R.K., Demurjian, S.A., Blechner, M., Agresta, T.: HOD2MLC: Hybrid Ontology Design and Development Model with Lifecycle. Int. J. Inf. Technol. Web. Eng. **10**(2), 16–42 (2015)
11. Xingshe, Z., Yalei, Y., Gang, Y.: A method for cyber-physical system dynamic behavior modeling. Chin. J. Comput. **37**(6), 1411–1423 (2014)
12. Zhou, C., Yin, Z., Feng, Y.: Events algebra of triggers. Comput. J. **60**(5), 1–10 (2017)

Pigeon Inspired Optimization and Enhanced Differential Evolution Using Time of Use Tariff in Smart Grid

Hafsa Arshad, Saadia Batool, Zunaira Amjad, Mudabbir Ali, Syeda Aimal, and Nadeem Javaid(✉)

COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
nadeemjavaidqau@gmail.com
http://www.njavaid.com

Abstract. In this paper, a scheduler for Home Energy Management (HEM) is proposed using Pigeon Inspired Optimization (PIO) and Enhanced Differential Evolution (EDE). Performance of these two optimization algorithms is evaluated in this study. Performance is determined by the amount of energy consumed by the appliances in on-peak hours and off-peak hours. Time Of Use (TOU) tariff is used for bill calculation of the consumed energy. Evaluation is performed in terms of Peak to Average Ratio (PAR) and electricity cost. Simulation results show that PIO outperforms EDE in terms of cost, PAR reduction and waiting time.

Keywords: Smart grid · Home energy management · Time of use tariff · User comfort · Pigeon inspired optimization · Enhanced differential evolution

1 Introduction

Electricity demand is increasing day by day as it has became the essential part of our daily life. Smart grids are introduced to utilize and deliver electricity efficiently and reliably to the consumers as traditional grids are not very effective. Smart grid shown in Fig. 1 is a system that includes physical power system and information system that links a variety of equipments and assets together to form a customer service platform [1].

Smart homes involve incorporating smartness into homes for comfort, healthcare, safety, security, and energy conservation [2,3]. Smart homes are equipped with smart meters. There is a burden on utility in peak hours as so many people are using electricity in that particular time period. Hence, utility increases cost in that time slot. Consumers can manage their energy consumption and cost by optimally scheduling the appliances. Utility also has a great benefit from the modernized grid as peak loads are reduced, security in enhanced with lower operational costs. Our objective is to reduce electricity price, to maximize user comfort and to manage load, i.e., shifting on-peak hours to off-peak hours.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_51

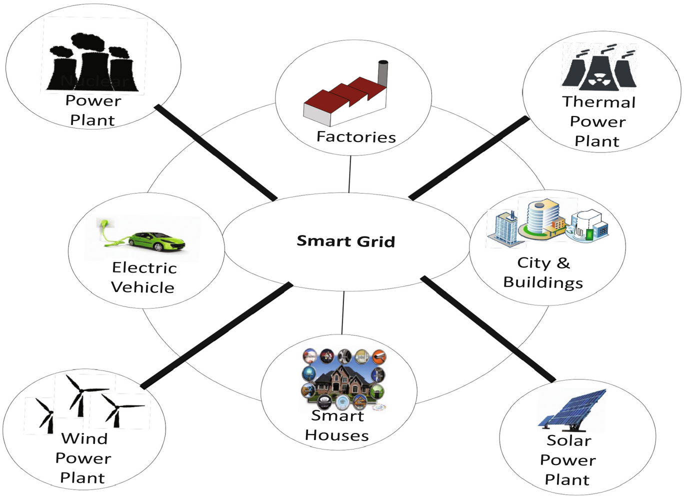


Fig. 1. Proposed system model in a Smart Grid

DSM is used by the electric utilities to enhance customer service. DSM refers to as controlling the amount of energy used at specific times to reduce system peak demand, utilize energy efficiently and balance the DR of the system. Along with balancing the supply and demand, minimizing peak power requirements is also important in DSM which include energy efficiency programs and smart metering, i.e., real-time pricing. RTP is one of the most important DR strategies where the prices announced by retailer change hourly over time to reflect variations in the market prices. Customers are notified of RTP prices before delivery time [4]. The prices in TOU are set well before the period and do not adjust to reflect actual conditions. The customers already know that how many costs they will pay for electricity during pre-set time periods. This allows customers to adjust their usage in response to the price signals and manage the overall energy costs by altering their usage at lower cost or by reducing complete usage.

In this paper, two optimization techniques; PIO and EDE, are used. These two techniques are used to find out the best optimal solution with maximum user comfort, minimum PAR and cost reduction. In PIO homing behavior of pigeons is considered whereas, in EDE trial vector strategy is followed to find out the best optimal solution. List of abbreviations is given in Table 1.

The rest of the document is organised as follows. Section 2 contains brief description of related work. In Sect. 3, the problem is stated. Section 4 describes the proposed model. Section 5 explains the simulations and results of the proposed system. In Sect. 6, the document is concluded.

Table 1. List of abbreviations

Abbreviations	Definition
PIO	Pigeon Inspired Optimization
EDE	Enhanced Differential Evolution
EMS	Energy Management Systems
HAN	Home Area Network
RTP	Real-Time Pricing
IBR	Inclined Block Rate
PAR	Peak to Average Ratio
DSM	Demand Side Management
ILP	Integer Linear Programming
GA	Genetic Algorithm
BPSO	Binary Particle Swarm Optimization Algorithm
BFOA	Bacterial Foraging Optimization Algorithm
WDO	Wind-Driven Optimization
GWD	Genetic Wind-Driven
RERs	Renewable Energy Resources
DERs	Distributed Energy Resources
ACO	Ant Colony Optimization
HEM	Home Energy Management
TOU	Time Of Use
ICTs	Information and Communication Infrastructures
OPEX	Operational Expenditure
GDSM	Generic Demand Side Management
WTA	Waiting Time of Appliances
DR	Demand Response

2 Related Work

Many of the scholars worked on the scheduling of the smart appliances. We categorize these contributions according to the issues faced in scheduling the appliances. The contributions of some of the researchers are as follows:

The authors in [1], introduced EMS in a HAN. The problem addressed by the authors is optimal power scheduling for home power usage. They use RTP tariff model along with IBR to avoid peak formation. The simulation results show that using RTP combined with IBR pricing model reduces cost and PAR more efficiently as compared to RTP alone. In order to schedule power consumption, authors propose a dynamic pricing approach along with game theoretic DSM framework in [2]. Proposed system is better in terms of cost, peak demand and convergence time however, total bill is increased a little. In article [11], the authors have highlighted the power scheduling issues for residential users in smart grid. Optimal scheduling strategies are obtained under three operational modes using day ahead pricing signal. Desired tradeoff between electricity payments and discomfort is achieved. However, PAR is not determined. Hence, by solving the problem of power scheduling, cost and PAR can be reduced as mentioned in [1,2], and convergence time can also be decreased.

In [5], the authors have used the algorithms for scheduling the residential load. They have designed an energy management controller based on heuristic algorithm for residential area in a smart grid. Five of the heuristic algorithms are used, along with RTP pricing signal, for scheduling the domestic load which namely are GA, BPSO, BFOA, WDO and hybrid GWD algorithm. Samadi et al. in [6], have proposed an algorithm for controlling load for DSM and adopted the approximate dynamic programming for scheduling of appliances. Simulations show that this algorithm lowers the energy cost. An ILP technique is proposed in [10], which is based on optimization mechanism for the home area load management. The aim of this proposed scheduling technique is to reduce the peak hourly load in order to achieve an optimal daily load schedule and to minimize combined power consumption. Simulation results show that a more balanced hourly load is achieved, when multiple households participate in scheduling. The authors in [13], introduce a concept of residential load scheduling framework on the basis of cost efficiency to enhance the economical efficiency of the residential electricity consumption. In the framework, the service fee and DERs are also considered and their influence on the cost efficiency is examined. The proposed algorithm results in better utilizing and saving the power. The cost efficiency criterion can be used in consumption with variable pricing scheme. The aim of [17], is to inaugurate a decentralized framework to organise DR of residential users in a smart grid. This framework is used to transform system load profile in order to minimize the payments of customers, and to keep their comfort and privacy. The results tell that the used approach provides great benefits without bothering for the cost and comfort of customers. In this paper, it is presumed that customers are supported by time changing prices and their response for energy minimization changes.

Liu et al. in [7], proposed a demand queuing-based energy management scheme on residential area. The goal of authors was to minimize cost by managing demand. They propose adaptive dynamic programming in order to solve the optimization problem. Achievements of proposed scheme are that it is able to manage trade-off between operational delay and energy consumption and minimization of energy cost. In [8], authors focus on the problem of excessive use of power. They took a survey and found out that the ICTs consume more power and emits greenhouse gases. The authors took survey on smart grid in order to find ways to reduce cost, efficient energy usage and to reduce emission of gasses. In [9], author presented a game theoretic demand side management in order to reduce PAR, energy cost and WTA. This is applied for multiple users with different power consumption. GA and RTP is used. Load shifting strategy is used instead of load reduction. The proposed model is bi-directional. Rahim et al. have introduced an architecture for DSM in [11]. GA, BPSO and ACO are proposed to calculate the performance of HEM controller. The problem with which the authors are dealing is the multiple knapsack problem. Combined model of TOU tariff and IBR is used for energy pricing. This results in cost effective solution to increase maintainability of smart grid. The main achievements of this study are electricity bill reduction, PAR minimization and user comfort maximization. Appliances are classified according to their features and users preferences. Hence, PAR minimization and user comfort maximization are the major goals along with energy cost minimization as mentioned in [7–10].

3 Problem Statement

The major objectives of this work are to maximize user comfort and reduce PAR along with cost minimization by optimizing power consumption patterns of the end users. The problem is stated as an optimization problem with time shiftable, power shiftable and fixed appliances. As in [5–9], user comfort is ignored. Hence, along with cost we are taking user comfort in consideration which is being tackled in this paper. In [17], comfort level of users is achieved on the cost of high electricity price as there is always a tradeoff between user comfort and cost.

4 Proposed Solution

Non-schedulable appliances have fixed power and time while schedulable appliances include power shiftable and time shiftable appliances. By assigning priorities to the appliances the scheduling problem can be solved. In scheduling problem, fixed appliances are given high priority and they are omitted from scheduling strategies. In terms of user comfort, the user has to wait for the appliances to turn on when the maximum load is shifted from on peak hours to off peak hours in order to reduce the electricity bills. Hence, appliances are assigned priorities to reduce the waiting time and maximize the user comfort. Along with waiting time, PAR reduction is also considered. PAR reduction is important in order to maintain balance between demand and supply. It can be defined as ratio of peak load to average load. Internal structure of one of the home including several appliances is shown in Fig. 2. The system model includes a smart grid which supplies energy to the smart homes. Home is equipped with smart meter which records the energy consumption of electricity in specific intervals of time and communicates that information on daily basis back to the utility for monitoring and billing.

Table 2 shows the power consumption, length of operational time and working hours of the appliances. Pigeon inspired optimization and EDE is used along with TOU pricing signal to reduce the PAR, cost, and WTA.

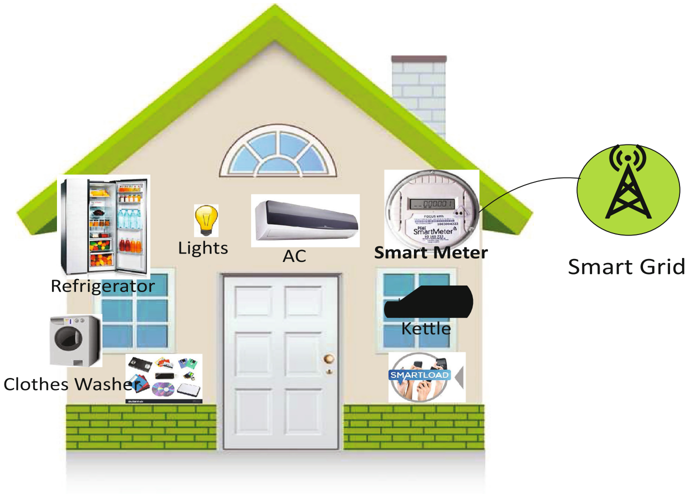


Fig. 2. An under consideration home's system model

Table 2. Proposed system model's parameters

Appliances	Working hours	Power (kWh)	LOT (hours)
Clothes washer	6p.m.–7a.m	0.7	2
Lights	6a.m.–11p.m	0.5	14
A.C	8a.m.–8a.m	1.4	15
Toaster	7a.m.–10a.m	1.146	3
Kettle	5a.m.–9a.m., 5p.m.–7p.m., 7p.m.–8p.m	1.2	3
Refrigerator	6a.m.–6a.m	0.2	17

4.1 PIO

PIO algorithm is a new bio-inspired swarm intelligence algorithm proposed by Duan et al. in [14] inspired by the homing behaviours of pigeons. It is an optimization technique which uses two operators:
(1) In map and compass operator, pigeons can sense the earth magnetic field by using magnetoreception to shape the map in their brains. To adjust the direction, pigeons use the altitude of the sun as compass. As pigeons fly to their destination, they rely less on sun and magnetic particles.
(2) In landmark operator, the pigeons will rely on neighbouring landmarks when they fly close to their destination. If they are familiar to the landmarks, they will fly straight to the destination. On the other hand, they will follow the pigeons who are familiar to the landmarks if they are far from the destination and unfamiliar to the landmarks. The working of PIO is shown in the Algorithm 1. Some of the steps of this algorithm are taken from [15].

Algorithm 1: PIO

```
Parameters initialization
Set initial path Xi for each pigeon
Set Xp=Xi, Nc=1
Calculate fitness values of different pigeon individuals
Xg:=arg min [ f(Xp)]
for Nc=1 to Nc1maxdo do
    for i=1 to Np do while Xi is beyond the search range do do
        Calculate Xi
    end
end
Evaluate Xi, and update Xp and Xg
for Nc= Nc1max+1 to Nc2max do do
    while Xp is beyond the search range do do
        Rank all the available pigeon individuals according to their fitness values
        NP = NP/2
        Keep half of the individuals with better fitness value, and abandon the
        other half
        Xc=average value of the paths of the remaining pigeon individuals
        Calculate Xi
    end
    Evaluate Xi, and update Xp and Xg
end
Xg is output as the global optima of the fitness function f
```

4.2 EDE

In EDE trial vector strategy is followed to improve the accuracy. Crossover and mutation takes place for finding the best optimal solution. Five trial vectors are formed. First three trial vectors are obtained by taking three different crossover rates. Fourth trial vector increases convergence speed while fifth increases diversity of search space. Mutant vector and target vector are created for finding the best optimal solution. Some of the algorithmic steps are taken from [16] in order to map the working of EDE shown in Algorithm 2.

5 Simulations and Results

The simulations are performed in matlab in order to compare the electricity cost, PAR and power consumption. For simulation results, a home with 6 appliances is considered along with 24-hour time slot. The appliances are divided into two categories; schedulable appliances and non-schedulable appliances. This categorization is done for effective scheduling of appliances. Schedulable appliances are further divided into two categories: (i) Time shiftable appliances and (ii) Power shiftable appliances. EDE and PIO along with TOU pricing signal is used to reduce the PAR and electricity bills of the appliances.

Algorithm 2: EDE

```
Initialize all parameters
Evaluate fitness of initial memory
Generate new solution using crossover and mutation
Initialize particle position to P_best
for i = 1:T do
    Evaluate fitness of initial memory
    Generate new solution using crossover and mutation
    Perform mutation
    for itr = 1:Max. iterations do
        Evaluate fitness
        Perform crossover
        Generate 1st trial vector using crossover rate 0.3
        if rand() ≤ 0.3 then
            u_j = v_j
        end
        if rand() > 0.3 then
            u_j = x_j
        end
        Generate 2nd trial vector using crossover rate 0.6
        if rand() ≤ 0.6 then
            u_j = v_j
        end
        if rand() > 0.6 then
            u_j = x_j
        end
        Generate 3rd trial vector using crossover rate 0.9
        if rand() ≤ 0.9 then
            u_j = v_j
        end
        if rand() > 0.9 then
            u_j = x_j
        end
        Generate 4rth trial vector
        Generate 5th trial vector
        Find the best vector among 5 trial vectors
        x_new ← besttrialvector
    end
    Perform selection using HSA selection procedure
    if f(x_new) < f(x_worst) then
        x_worst = x_new
    end
end
```

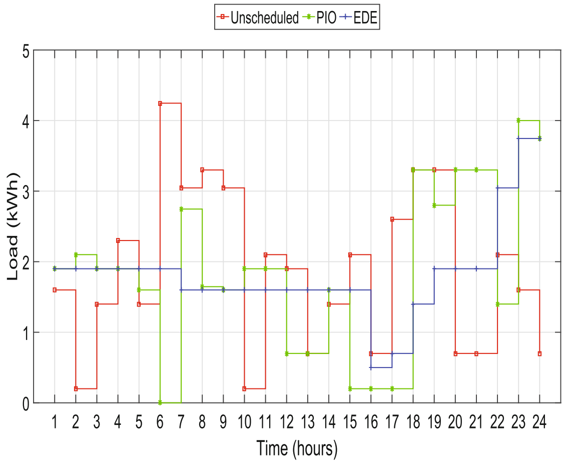


Fig. 3. Hourly load in kiloWatt hours

Figure 3 represents the load of the appliances. Load is reduced during on-peak hours and shifted to the off-peak hours. Load for unscheduled appliances is 4.2 kWh while that of PIO and EDE is 4 kWh and 3.8 kWh respectively. Hence, the load is decreased upto 10% in schedulable appliances as compared to non-schedulable appliances.

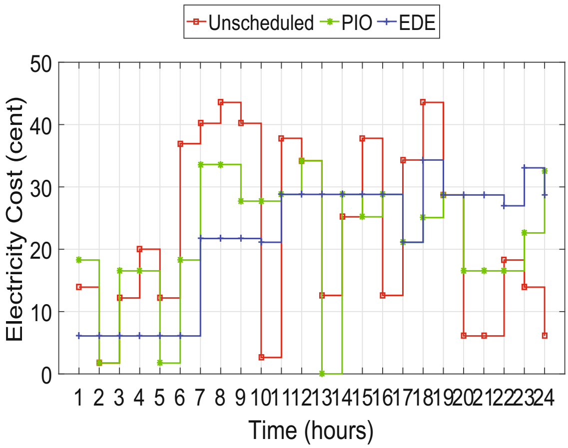


Fig. 4. Hourly electricity cost

The cost of scheduled and non-scheduled appliances in shown in Figs. 4 and 5. From Fig. 4, it can be understood that hourly electricity cost of scheduled appliances is decreased per cent as compared to the non-scheduled appliances. This

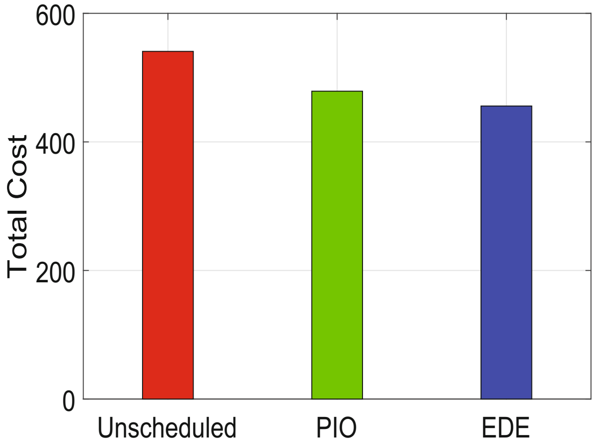


Fig. 5. Total cost in cents

decrease in the cost is due to the selection of operational time of the appliances by the consumers themselves. The user may choose to adjust the operational time of the appliance to the time slots with less electricity price to lower the payments. However, holding up the operation of the appliance will result in uneasiness. The maximum electricity cost of non-scheduled appliances is 44 cents while that of scheduled appliances is 34 cents in case of PIO while 33 cents in case of EDE. Hence, the cost of EDE and PIO is approximately 10% less than that of the non-scheduled appliances. This is because the cost of scheduled appliances is always less than that of non-scheduled appliances. Figure 5, demonstrates that the cost of non-scheduled appliances is 570 cents approximately while that of scheduled appliances is 490 cents and 450 cents in PIO and EDE, respectively. Hence, the customer has to pay approximately 10% more for achieving comfort. This is because there always exists a trade-off between the cost and the discomfort.

Figure 6, demonstrates that the PAR of scheduled appliances is less than that of the non-scheduled appliances. PAR reduces the formation of peaks and maintains balance between on-peak hours and off-peak hours. Hence, along with peak load, peak to average ratio is also reduced. In case of non-scheduled appliances, the PAR is 1.4 while in case of scheduled appliances it is reduced upto 0.75 approximately in PIO and EDE. Hence, PAR of scheduled appliances is 40% less than that of non-scheduled appliances. By shifting load in schedulable appliances peak formation can be avoided. However, in non-schedulable appliances peaks are formed as load is not balanced. Hence, PAR is reduced to a large extend in case of scheduled appliances. Thus, PAR of scheduled appliances is always less than that of non-scheduled appliances.

Figure 7, shows the user comfort which is calculated in terms of waiting time. There is a trade-off between discomfort and electricity cost. To reduce electricity bills, load should be shifted because of this user has to wait for an appliance to

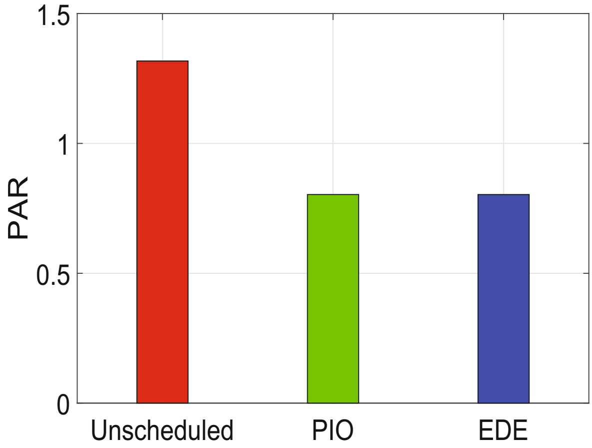


Fig. 6. Peak to average ratio

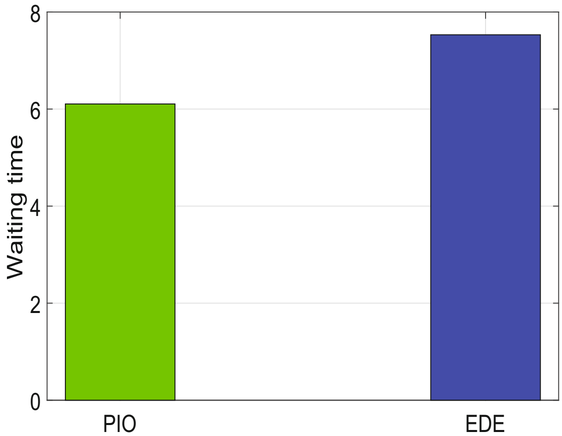


Fig. 7. User comfort

turn ON. If a user does not want to compromise their comfort, then it will result in high electricity cost and also peak down the on-peak hours which will burden the user. Waiting time in case of PIO is 6.1 min and that of EDE is 7.8 min which clearly shows that waiting time of PIO is 20% less than that of EDE. This employs that users can achieve more comfort using PIO rather than EDE. Hence, the desired trade-off is achieved.

Figure 8, shows the TOU price signal. This pricing signal is used for the cost estimation in on peak and off peak hours of the day. Under TOU model, electricity prices are set at different prices during different day times. The electricity prices are lower when used in off-peak hours and most expensive when many customers are using electricity simultaneously. Utilities may even develop the

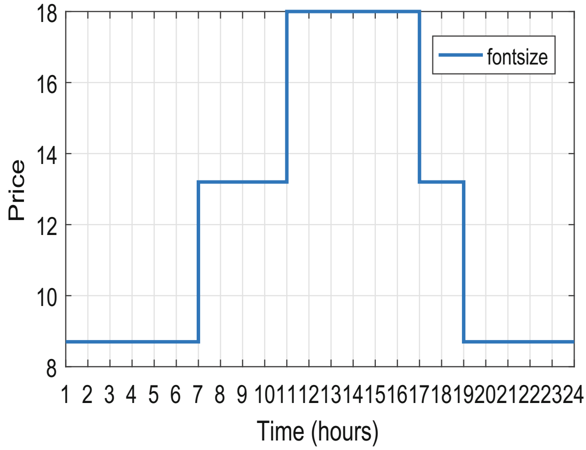


Fig. 8. TOU pricing signal

rates between on-peak and off-peak times of the day. The customers already know that how many costs they will pay for electricity during pre-set time periods. This pre-determination of the electricity cost allows customers to adjust their usage in response to the price signals and manage the overall energy costs by altering their usage at lower-cost or by reducing complete usage. Simulation results show that PIO outperforms EDE in terms of user comfort maximization among all the selected performance parameters.

6 Conclusion

In this paper, a scheduler for HEM is proposed using PIO and EDE. The performance of both of these optimization techniques is evaluated by determining the amount of energy consumed. TOU pricing signal is used for bill calculation of the consumed electricity. Evaluation is performed in terms of PAR, electricity cost and waiting time. From the simulation results, we conclude that waiting time of PIO is 20% less than that of EDE which clearly shows that users can achieve more comfort using PIO rather than EDE. PAR of PIO and EDE are 40% less than that of unscheduled appliances. Electricity cost of EDE and PIO is determined to be 10% less than unscheduled appliances. This is because of the fact that there is always a tradeoff between user comfort and cost. Load of PIO is 4 kWh and that of EDE is 3.8 kWh. Hence, it is clear that PIO outperforms EDE in user comfort maximization.

References

1. Zhao, Z., Lee, W.C., Shin, Y., Song, K.-B.: An optimal power scheduling method for demand response in home energy management system. IEEE Trans. Smart Grid **4**(3), 1391–1400 (2013)

2. Rottondi, C., Barbato, A., Chen, L., Verticale, G.: Enabling privacy in a distributed game-theoretical scheduling system for domestic appliances. IEEE Trans. Smart Grid **8**(3), 1220–1230 (2017)
3. Alam, M.R., Reaz, M.B.I., Ali, M.A.M.: A review of smart homes past, present, and future. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(6), 1190–1203 (2012)
4. Meng, F.-L., Zeng, X.-J.: An optimal real-time pricing algorithm for the smart grid: a bi-level programming approach. In: OASIcs-OpenAccess Series in Informatics, vol. 35. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2013)
5. Javaid, N., Javaid, S., Abdul, W., Ahmed, I., Almogren, A., Alamri, A., Niaz, I.A.: A hybrid genetic wind driven heuristic optimization algorithm for demand side management in smart grid. Energies **10**(3), 319 (2017)
6. Samadi, P., Wong, V.W.S.: Load scheduling and power trading in systems with high penetration of renewable energy resources. IEEE Trans. Smart Grid **7**(4), 1802–1812 (2016)
7. Liu, Y., Yuen, C., Rong, Y., Zhang, Y., Xie, S.: Queuing-based energy consumption management for heterogeneous residential demands in Smart Grid. IEEE Trans. Smart Grid **7**(3), 1650–1659 (2016)
8. Erol-Kantarci, M., Mouftah, H.T.: Energy-efficient information and communication infrastructures in the smart grid: a survey on interactions and open issues. IEEE Commun. Surv. Tutorials **17**(1), 179–197 (2015)
9. Khan, M.A., Javaid, N., Mahmood, A., Khan, Z.A., Alrajeh, N.: A generic demand side management model for Smart Grid. Int. J. Energy Res. **39**(7), 954–964 (2015)
10. Zhu, Z., et al.: An integer linear programming based optimization for home demand-side management in Smart Grid. In: 2012 IEEE PES Innovative Smart Grid Technologies (ISGT). IEEE (2012)
11. Rahim, S., Javaid, N., Ahmad, A., Khan, S.A., Khan, Z.A., Alrajeh, N., Qasim, U.: Exploiting heuristic algorithms to efficiently utilize energy management controllers with renewable energy sources. Energy Build. **129**, 452–470 (2016)
12. Ma, K., Yao, T., Yang, J., Guan, X.: Residential power scheduling for demand response in Smart Grid. Int. J. Electr. Power Energy Syst. **78**, 320–325 (2016)
13. Ma, J., Chen, H.H., Song, L., Li, Y.: Residential load scheduling in Smart Grid: a cost efficiency perspective. IEEE Trans. Smart Grid **7**(2), 771–784 (2016)
14. Duan, H., Qiao, P.: Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning. Int. J. Intell. Comput. Cybern. **7**(1), 24–37 (2014)
15. Zhang, B., Duan, H.: Predator-prey pigeon-inspired optimization for UAV three-dimensional path planning. In: Tan, Y., Shi, Y., Coello, C.A.C. (eds.) ICSI 2014. LNCS, vol. 8795, pp. 96–105. Springer, Cham (2014). doi:10.1007/978-3-319-11897-0_12
16. Zafar, A., Shah, S., Khalid, R., Hussain, S.M., Rahim, H., Javaid, N.: A meta-heuristic home energy management system, January 2017
17. Safdarian, A., Fotuhi-Firuzabad, M., Lehtonen, M.: Optimal residential load management in smart grids: a decentralized framework. IEEE Trans. Smart Grid **7**(4), 1836–1845 (2016)

The "Gut-Feeling" in the Decision-Making Process: A Computationally Efficient Approach to Influence Relation Assessment

Olayinka Johnny[1], Marcello Trovati[2(✉)], and Jeffrey Ray[2]

[1] Department of Computing and Mathematics, University of Derby, Derby, UK
fabyinka@yahoo.com

[2] Department of Computer Science, Edge Hill University, Ormskirk, UK
{trovatim,rayj}@edgehill.ac.uk

Abstract. Deep learning is a relatively new research area, motivated by the need to obtain more accurate, flexible and applicable methods for knowledge discovery. However, deep learning is a much wider concept, which entails a deep understanding of a scenario and its corresponding parameters, aiming to fully describe the interconnections between them. In this paper we will argue that a "gut-feeling" approach can be potentially utilised to obtain accurate results, and we will consider an initial approach to evaluate specific information captured by dependency networks.

1 Introduction

Over the last decades, deep learning has become the focus of significant research in machine learning [1], which has demonstrated to have a wide impact on many multidisciplinary fields and research areas [2,14]. Loosely speaking, deep learning sits between numerous research topics, including neural networks, artificial intelligence, optimisation, and pattern recognition, to name but a few. There are several similar, yet slightly different definitions of deep learning, based on its scope and context. In this paper, we refer to deep learning as a set of models with multiple layers of information processing, based on supervised or unsupervised learning of feature representation, where all the parameters are fully investigated to provide an exhaustive modelling approach [1].

Complex networks with different uncertainty layers have been extensively used to describe and model a variety of scenarios within the decision-making domain [3]. Examples of such networks, include Dependency Networks and Influence Diagrams, which are characterised by the mutual relationships between nodes representing specific concepts. However, the identification of such concepts and their relationships is usually a complex task, as a variety of probabilistic and topological constraints need to be addressed, especially when extracted from textual sources [4]. Furthermore, depending on the scenario, there are potentially a

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_52

large number of parameters, which increase the overall computational complexity [3]. As a consequence, when deep learning has been applied to this context, the extraction of actionable information from such a scenario may raise a variety of issues [3,9].

However, when a decision process is manually modelled by expert only a partial amount of information is explicitly processed as they often rely on their "gut feeling". There is compelling evidence from neuroscience that emotions and gut feelings play an important role in the rational process of intelligence extraction, where knowledge based on historical data, personal experience and expertise drives the decision-making process, rather than deep rational decision making [10]. According to [5], gut feeling refers to the skill of focusing on those potentially important but sometimes faint signals that fuel imagination, creativity and innovation and feed corporate success in globally competitive business environments.

In [6], the author identified a number of ways in which executives use gut feeling in making decisions. These include sensing a problem, performing pre-programmed behaviour patterns, understanding the bigger picture, which allow the by-passing of deep analysis. Therefore, a more effective decision making process would result from a full integration of gut feelings [7].

In modelling a complex scenario, one of the first tasks is to assess the existence of relationships linking its different components. There are various types of relations, defined by strict semantic constraints, such as temporality, type of action and direction. *Influence* is defined by fewer constraints compared to other relations. In fact, two concepts are influenced by each other if there is a more general semantic link [9]. Loosely speaking, influence relations capture general semantic relationships suggesting a link between the corresponding concepts. In other words, the strict mathematical and semantic constraints of other types of relations, such as causal ones, are relaxed due to the vagueness inherent in influence relations.

The aim of this paper is to provide an initial discussion and implementation, which models a "gut-feeling" approach in the assessment of an influence relation between pairs of concepts extracted from textual sources. The proposed model is based on the following assumptions:

- Some initial knowledge regarding the influence relationships between the corresponding concepts is present
- Additional information is then obtained iteratively via appropriate text analysis of small sets of texts.
- If the overall knowledge is consistent with the initial assessment, then we assume it is accurate and no further analysis is suggested. This suggests that the gut-feeling associated with the original "shallow knowledge" (as opposed to a deep analysis of the necessary data) is an appropriate modelling representation of the corresponding scenario.

The paper is structured as follows: in Sect. 2 an overview of the current research in the field, and Sect. 3 describes the method introduced. Section 4 discusses the evaluation results and finally Sect. 5 concludes our work.

2 Related Work

There are a variety of approaches to extract and evaluate relevant information to assess influence relations between concepts [15]. All these methods are based on *rational thinking*, which virtually assumes unlimited knowledge, time, and information-processing power. In other words, all possible scenarios are considered and their outcomes are assessed via a logical and systematic manner to identify the best possible choice. The ability to perform such evaluation based on large quantities of parameters, which define a specific scenario is at the core of deep learning. However, there is compelling evidence from neuroscience research that emotions play an important role in the rational process of intelligence and the decision-making process [10].

Emotions are changes in both body and brain states in response to different stimuli [11]. When physiological changes occur in the body, they are relayed to the brain where they are transformed into an emotion. In fact, over time, these emotions and their corresponding states of the body become associated with particular situations and their past outcomes. Consequently, when making decision, these physiological changes and their evoked emotion are consciously or unconsciously associated with their past outcomes, which therefore influence the overall decision-making process. This reasoning and decision-making is said to be associative because it compares similar situations that have been encountered in the past with current situations and then make decision accordingly, On the other hand, some have argued that rational thinking and decision-making does not leave much room for emotions [12]. In [13], the authors cautioned that we are only rational within the limits of our cognitive capacities and that decision making itself is often an emotional process, and without emotional involvement, decision making might not even be possible or might be far from optimal [11]. In fact, the decision making process depends on emotional processing and the resulting feelings, which involve images that relate to the state of the body. This study agrees with the views of [11,13], captured by the following statement: "*reason without emotion is inadequate for making the decisions that guide our lives, and in fact make up our lives*" [8]. This provides motivation and a new perspective on incorporating emotions in decision-making systems. Furthermore, an automated method based on a "gut-feeling" assessment would enable a more computationally efficient approach in this context.

3 Description of the Method

As discussed above, the main intuition behind the method proposed in this paper is that when we "feel" that a specific decision or information assessment is to be preferred over other options, this is based on some background knowledge. Usually, this type of knowledge has a variety of sources, from personal experience, to common knowledge and beliefs. An exhaustive modelling of all the possible types of knowledge, which play a central role in this aspect is beyond the scope of this paper as it would require a full investigation of several psychological and sociological issues. As a consequence, we assume the existence of some *a priori* knowledge in the form of structured data where concepts are linked by mutual influence relations, which vary in strength. In particular, this can be defined by a network $G = G(V, E)$, where V is the *node-set* containing the different concepts, and E is the *edge-set* so that if $e_{v_i,v_j} \in E$ then v_i and $v_j \in V$ are assumed to the linked by an influence relation [3]. Note that *influence* can be viewed as a many-to-one map.

As depicted in Fig. 1, the dynamical properties captured by the above network are investigated to provide an assessment of the reliability of the initial knowledge on a specific scenario. In particular, we say that such knowledge is *stable* if its properties remain (relatively) constant. This allows us to measure the level of reliability of some existing knowledge when it used to infer an outcome. In other words, let $G_t = G_t(V_t, E_t)$ be the knowledge at a time t. If the dynamics of G_t is stable over a specific time iteration, then we assume it can be reliably used.

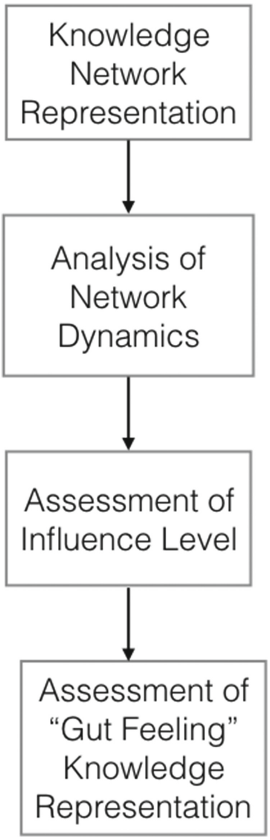


Fig. 1. The main components of the method discussed in Sect. 3.

For $e_{v_i,v_j} \in E$, we define its influence level as $i(e_{v_i,v_j}) \in (0,1]$, so that if $i(e_{v_i,v_j}) = 1$, then v_i and v_j are strongly influenced by one another. Note that influence relations might not have a clear direction, as they are assumed to be much more general than directed relationships such as causality [3]. Loosely speaking, if two concepts influence each other, then we know they are linked without necessarily knowing which of them is directly affected by the other. In order to assess the dynamics of G_t for $t \geq 1$, we need to evaluate the differences between G_t and G_T for $T > t$. We assume that no nodes can be removed, or in other words, $V_t \subset V_T$. However, new nodes can be added.

In assessing the validity of our current knowledge regarding the influence between two concepts A and B, we will consider the following properties:

- The shortest path connecting A and B with the biggest average influence value over its edges. If we have more than one, any of them can be chosen
- The average change of the influence between concepts in the path between A and B, and
- How widely they change over a certain amount of time iterations.

The former can be modelled via the following equation

$$\frac{1}{m}\left|\sum_{k=2}^{m}\left(\sum_{e_{v_i,v_j}\in p(A,B)} \frac{1}{\#E_{p(A,B)}}\left|i_1(e_{v_i,v_j}) - i_k(e_{v_i,v_j})\right|\right)\right|, \tag{1}$$

where $p(A,B)$ is the path connecting A and B, $\#E_{p(A,B)}$ is the number of edges in the path and $i_l(e_{v_i,v_j})$ is the influence level at time l.

The latter is modelled via the following equation

$$\log\left(\frac{n}{W}(e-1)+1\right), \tag{2}$$

where $W = \#\{w : |i_1(e_{v_i,v_j}) - i_k(e_{v_i,v_j})| > th, k = 2,\ldots,m\}$ and th is the outlier threshold of the influence, which can be either set manually, or inferred from a training dataset. In other words, W measures the number of influence values far from the initial value at $t = 1$, relatively to the threshold th.

We therefore, define the *change of influence level* $\tilde{i}(A,B) \in [0,1]$ by combining Eqs. 1 and 2

$$\tilde{i}(A,B) = \frac{1}{2}\left(\frac{1}{m}\left|\sum_{k=2}^{m}\left(\sum_{e_{v_i,v_j}\in p(A,B)} \frac{1}{\#E_{p(A,B)}}\left|i_1(e_{v_i,v_j}) - i_k(e_{v_i,v_j})\right|\right)\right| + \log\left(\frac{n}{W}(e-1)+1\right)\right) \tag{3}$$

Note that if $\tilde{i}(A,B)$ is close to 0, then the initial evaluation of the influence relation level is reliable. On the other hand, if it is close to 1, then it should be discarded suggesting that a deeper analysis should be carried out.

4 Evaluation

In this section, we discuss a preliminary validation of the method discussed in Sect. 3, based on an annotated textual dataset extracted from PubMed [18], which includes over 24 million citations from the biomedical research field.

Similarly to [9], we identified over 250 publicly accessible articles based on "breast cancer", which were analysed using the Python NLTK toolkit [19] to assess the syntactic structure, via specific text patterns. For more details, refer to [3]. This process created a semantic network, which allows the assessment of the influence between concepts represented by its nodes.

We selected three keywords pairs, which correspond to well known and widely documented concepts that are influenced by each other [9]. These are:

- Breast cancer – Age
- Breast cancer – Gender
- Breast cancer – Menopause

Table 1. Values of the initial influence level.

Concept 1	Concept 1	Initial influence level
Breast cancer	Age	0.61
Breast cancer	Gender	0.82
Breast cancer	Menopause	0.73

The initial influence level was determined by analysing 30 texts, which produced the values shown in Table 1. Subsequently, each iteration was defined by adding information extracted from 10 additional texts, for a total of 20 iterations.

Figure 2 depicts the different influence level for the above pairs over the different iterations. Note the values of the initial influence levels are shown as a blue line.

The threshold values th were determined by analysing 20 texts and measuring their average variation with respect to their initial influence levels. Equation 3 was then used to assess the values of $\tilde{\imath}(A, B)$ for the above pairs. As shown in Table 2, we obtained the following results

- $\tilde{\imath}(\text{Breast cancer, Age}) = 0.1353$
- $\tilde{\imath}(\text{Breast cancer, Gender}) = 0.1664$
- $\tilde{\imath}(\text{Breast cancer, Menopause}) = 0.14616$.

Considering that all the influence level values are relatively close to 0, this suggests that the initial influence levels reliably capture their corresponding influence relations, which is consistent with the initial assumption of the evaluation process.

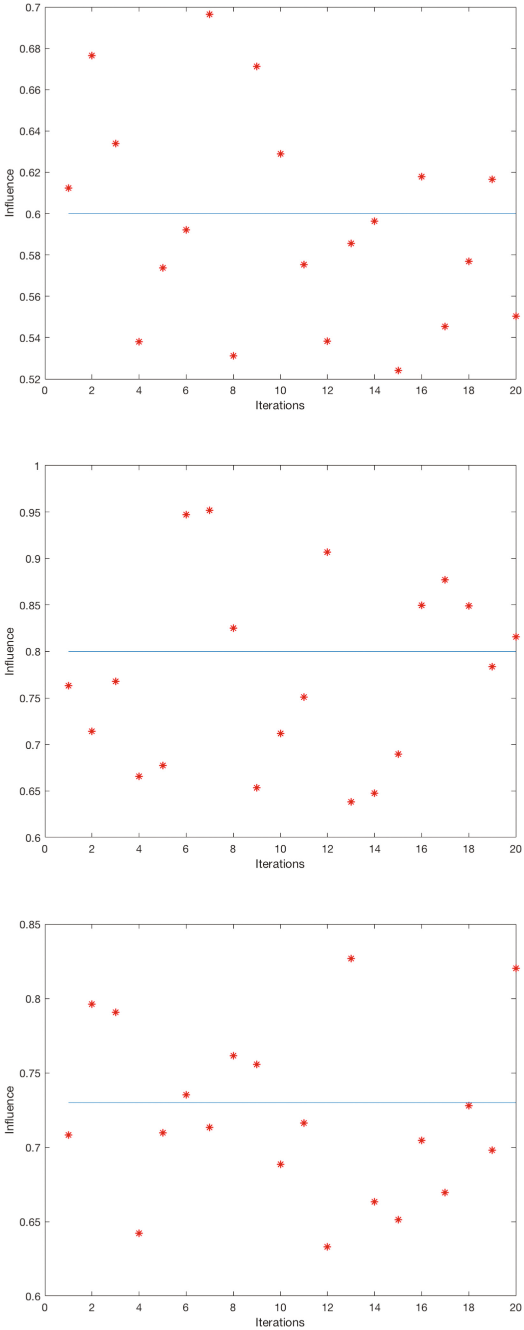


Fig. 2. Values of the influence levels between the keywords described in Sect. 3.

Table 2. Evaluation details

Concept 1	Concept 1	Threshold	Influence level	Number of outliers
Breast cancer	Age	0.074	0.1353	3
Breast cancer	Gender	0.152	0.1664	2
Breast cancer	Menopause	0.094	0.14616	2

5 Conclusion

In this paper, we have introduced an initial approach to assess influence relationships between concepts based on an interpretation of the "gut feeling", which provides an accurate and computationally efficient approach to the decision-making process. This is part of a wider line of inquiry, which aims to define complex decisional networks by bypassing complex calculations, and by providing an innovative approach to their analysis. In future research, we are aiming to provide a more rigorous evaluation of the threshold values, as well as a better predictive assessment of the influence levels to be embedded in the automated creation of suitable networks, such as Dependency and Bayesian Networks.

References

1. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning (2009)
2. Arel, I., Rose, C., Karnowski, T.: Deep machine learning - a new frontier in artificial intelligence. IEEE Comput. Intell. Mag. **5**, 13–18 (2010)
3. Trovati, M.: An influence assessment method based on co-occurrence for topologically reduced big datasets. Soft. Comput. **20**, 2021–2030 (2015). Springer, Heidelberg
4. Blanco, E., Castell, N., Moldovan, D.I.: Acquiring Bayesian networks from text. In: LREC (2008)
5. Sadler-Smith, E., Shefy, E.: The intuitive executive: understanding and applying 'Gut Feel' in decision-making. Acad. Manage. Executive **8**(4), 76–91 (2004)
6. Isenberg, D.: How senior managers think? In: Harvard Business Review, pp. 81–90 (1984)
7. Mousavi, S., Gerd, G.: Risk, uncertainty, and heuristics. J. Bus. Res. **67**(8), 1671–1678 (2014)
8. Binali, H., Chen, W., Vidyasagar, P.: Computational approaches for emotion detection in text. In: 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST) (2010)
9. Trovati, M., Castiglione, A., Bessis, N., Hill, R.: A kuramoto model based approach to extract and assess influence relations. In: Li, K., Li, J., Liu, Y., Castiglione, A. (eds.) ISICA 2015. CCIS, vol. 575, pp. 464–473. Springer, Singapore (2016). doi:10.1007/978-981-10-0356-1_49
10. Loewenstein, G., Lerner, J.S.: The role of affect in decision making. In: Handbook of Affective Science, pp. 619–642 (2003)
11. Damasio, A.R.: Descartes' Error: Emotion, Rationality and the Human Brain. New York, Putnam, 352 (1994)

12. Livet, P.: Rational choice, neuroeconomy and mixed emotions. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**(1538), 259–269 (2010)
13. Zeelenberg, M., Nelissen, R.M., Breugelmans, S.M., Pieters, R.: On emotion specificity in decision making: why feeling is for doing. Judgment Decis. Making **3**(1), 18 (2008)
14. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **38**, 1798–1828 (2013)
15. Shachter, R., Bhattacharjya, D.: Dynamic programming in influence diagrams with decision circuits. In: Proceedings of 25th UAI, Catalina Island, CA, USA (2010)
16. De Marneffe, M.F., MacCartney, B., Manning, C.D.: Generating typed dependency parses from phrase structure parses. In: LREC (2006)
17. Manning, C.D.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)
18. PubMed Website. http://www.ncbi.nlm.nih.gov/pubmed/. Accessed Apr 2017
19. Natural Language Toolkit Website. Natural Language Toolkit Website. http://www.nltk.org/. Accessed Apr 2017

A Preliminary Automated Approach to Assess Hospital Patient Feedback

Jeffrey Ray[1], Marcello Trovati[1](✉), and Simon Minford[2]

[1] Department of Computer Science, Edge Hill University, Ormskirk, UK
{rayj,trovatim}@edgehill.ac.uk
[2] Alder Hey Children's Hospital, Liverpool, UK
minfords@alderhey.nhs.uk

Abstract. Hospitals gather huge amounts of valuable data every day, from clinical studies to patient satisfaction surveys, whose assessment is often carried out either manually or via limited automated approaches. However, improving analysis of hospital data can contribute to a more informed decision-making process with a positive impact on patients. In particular, analysing patients' feedback can identify potential events, procedures and infrastructure issues, which might potentially affect hospital users' well-being and medical treatments. This paper discusses an initial implementation of some components of a data analytics tool, which combines a variety of techniques based on text mining and mathematical modelling. It actively assesses real-time data extracted from patients' feedback, social media and other suitable data created within a hospital setting, allowing immediate action to be taken. This will provide a system, which actively learns positive actions and outcomes whilst discovering actionable information to facilitate the decision-making process, enhance the patient's journey, and potentially optimise treatment outcomes.

1 Introduction

Hospital data size is likely to increase dramatically in the coming years [1]. It is therefore vital for healthcare establishments to use the existing devices, structure and methods to control big data efficiently to potentially avoid large revenue losses [3]. Previous and current research on data analytical methods had been found to be effective on huge amount of un-analysed health and medical patient data. During discussions with members of the Innovation Hub at the Alder Hey Children's Hospital, it was confirmed that the NHS currently identifies most of its feedback based on a patient experience survey along with the friends and family test (FFT). However, current literature suggests that the use of the FFT is limited due to the small range of patients that willingly complete the form, who are usually negatively biased [4]. Other methods include patient panels, focus groups and the mystery shopper data collection style often used within retail. All of these methods require lengthy processing time [5], which increases

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_53

the action time between collation of feedback and action being taken. Any delay between feedback and action in a hospital setting can result in patients feeling being ignored [6].

From a review of published sources, it is apparent that much of the available literature is of a non-experimental, descriptive nature. Such a trend is not uncommon for research questions that address the effectiveness of organisational processes, and the information gathering questions greatly vary if they are posed to managers, clinicians, or patients. This issue raised by many reports, creates a divide between perceived good health care from a patient's point of view, poor service provided by a clinician and vice-versa [7].

Only a small amount of literature exists that directly addresses the action feedback loop in a hospital context, i.e. the feedback process by which incident data in any form are transformed into beneficial improvements in operational safety or actionable events that produce positive improvements in patients' experiences. Academic review of patient satisfaction surveys reveals that measuring patient satisfaction has numerous problems, as most surveys oversimplify complex issues that each patient faces during his or her hospital stay. Furthermore, the notion of *satisfaction* is highly debated, which is defined as a judgement formed by individuals over time as they reflect on their experience. This experience can differ from quality health care as perceived by a health care professional [8]. Most surveys or feedback gathering exercises have a long processing and administration overhead, leading to lengthy delays between comments, suggestions and experience reviews from a patient transforming into visible action, usually long after the patient has left the hospital.

In [9], the authors discuss a qualitative interview study concerning the effectiveness of data feedback in supporting performance improvement efforts in eight US hospitals. Data quality, timeliness and credibility were identified as important factors for effective improvement, along with leadership and persistence in data feedback processes "*data feedback must persist to sustain improved performance. Embedded in several themes was the view that the effectiveness of data feedback depends, not only on the quality and timeliness of the data, but also on the organisational context in which such efforts are implemented*" [9].

In [10], the authors highlight two critical determinants for success: timely, effective feedback and demonstrable utility. The former assures reporters that their reports are acted upon and are not trapped into an administrative "black hole". Demonstrating the local usefulness of incident data, in addition to the development of external reports, influences user adoption and compliance, and can improve reporting rates. These factors highlight the need for immediate analysis and processing of patient feedback, clinical and hospital data. The ability to follow action in a timely manner is further highlighted by [11], who discuss the importance within US health care of follow-up actions after a safety or error report. The authors suggest that more emphasis should be placed on event follow-up, prioritising opportunities and actions, assigning responsibility and accountability, whilst producing an action plan to meet the needs of the reported issue [11].

Current literature treats feedback from patients as a separate category from that produced by clinical staff [12], with differing procedures and protocols in place for dealing with feedback. NHS staff training literature contains no guidance for feedback. Each hospital manages and addresses its feedback, comments and complaints in a unique manner, with some publishing reports after an unspecified period. In some cases this can take upwards of four months. Reports also indicate that formal language and taxonomy varies between hospital setting and background of each patient who provides feedback. A standardised taxonomy to focus on specific areas is recommended to narrow feedback into actionable groups, creating manageable action groups with patient satisfaction metrics applied to each area [13].

The literature relating to hospital data collection indicates that current methods are largely ineffective at providing immediate feedback to patients, allowing the patient to feel involved, and that any reports they submit will have an impact upon the conditions experienced throughout the hospital. With real-time data collection and feedback opportunities, the use of free form text entries, as well as emotional and personal wellbeing index scores, would ensure the data being collected will directly relate to the current experience.

In this paper, we discuss an initial implementation of a method to assess patient satisfaction based on text mining techniques, to populate suitable decision networks. In particular, we consider the method introduced in [23,25], where fragments of Bayesian Networks (BNs) are automatically extracted from textual sources. Furthermore, sentiment analysis of tweets can facilitate the assessment of the influencing factors that affect patient satisfaction.

The paper is structured as follows: in Sect. 2, the proposed method is introduced, and Sect. 3 describes the dataset used in this work. Section 4 focuses on the implementation and validation results, and finally Sect. 5 concludes the paper.

2 Description of the Method

In this paper, a grammar-based information extraction is utilised, which is defined by a set of *text patterns* aiming to identify textual fragments with a specific syntactic structure [20]. This determines the different parts of sentences, which refer to nouns, verbs and attributes, as well as suitably defined keywords, which may indicate a specific state [18].

Sentiment analysis is carried out to assess the mood, or *polarity* embedded in textual information [19]. In particular, the output consists of the following couples

`(nouns, pol)`

where

- `nouns` refers to the concepts identified in the sentences.

- `pol` refers to the sentiment polarity identified in the sentences, and it is a numerical value between -1 and 1.

The value of the sentiment polarity has been evaluated using the VADER sentiment analysis algorithm [21]. This is a rule-based model for specifically designed to carry out sentiment analysis from social media content, whilst generalisable to various domains.

The output is then used to populate a network, whose nodes include the nouns extracted from `nouns` connected by an edge with a weight in the range $[-1, 1]$ corresponding to the sentiment polarity, where negative and positive values signify an overall negative or positive mood, respectively. In general, we might have more than an edge between pairs of node as these might have been extracted from different textual data.

Using the method introduced in [23], specific (semantic) networks can be defined by identifying the different relations captured by textual fragments. In particular, the following text pattern was utilised

```
(NP1, keyword, NP1),
```

where

- `NP1` and `NP2` are the noun phrases, which contain nouns corresponding to specific concepts
- `keyword` refers to a set of keywords related to semantic dependency. See [24] for more details.

The above text pattern allows the identification of probabilistic relationships between concepts, which can be of *dependence* or *independence* types, depending on whether a dependency or independence relation is present between the corresponding concepts, respectively. The topology of the resulting network is subsequently investigated to generate fragments of BNs [25]. Bayesian Networks (BNs) are acyclic networks which model the probabilistic relationships between variables, as well as their historical information. In particular, their nodes are associated with concepts, which are conditionally dependent if they are joined by an edge [26]. BNs have been used in a variety of modelling scenarios defined by uncertain or partially-known information, and they are particularly useful in understanding the influencing factors in a decision system.

3 Description of the Dataset

Our initial validation is based on the a textual dataset focusing on the patients' satisfaction at UK hospitals. More specifically, the main components included approximately 3000 tweets and over 370 abstracts freely available from PubMed, which were identified by the following keywords: *patients' satisfaction, hospital,* and *NHS*.

The dataset was subsequently pre-processed to remove any format inconsistencies, such as extra lines, strange characters and information not relevant to the context, including authors' names and affiliations, etc.

4 Evaluation

The evaluation was carried out in Python with the NLTK, vaderSentiment and NetworkX libraries. NLTK and vaderSentiment allow to access more than 50 corpora and lexical resources such as WordNet [27], which integrates text processing libraries to classify and analyse textual data and assess their sentiment polarity [28, 29]. In particular, it integrates the Stanford Parser [20], which was used in the analysis in this article. The NetworkX library is designed to define, analyse and investigate the properties of complex networks [30]. This was used to define the networks extracted from the textual sources, and subsequently analyse them via the numerous methods available in this library. In fact, it contains several algorithms to fully assess the topology of networks.

A network was created based on the PubMed articles, which identified dependence relations using the method introduced in [25]. The aim was to assess whether the above concepts are strongly linked as well as how they are discussed in the tweets extracted. We identified the following concepts:

- Quality of life
- Patient's satisfaction
- Treatment satisfaction
- Diagnosis
- Readmission rate
- Medication Management
- Cost-effectiveness
- Age
- Care Management
- Experience of care
- Social requirements
- Vulnerable patient
- Improvement of services
- NHS
- Staff satisfaction
- Wellbeing.

Figure 1 depicts the complete network associated with the above concepts, were the edges represent a dependency relation. Subsequently, using the method introduced in [25], we identified the Bayesian network depicted in Fig. 2.

The assessment of the conditional probabilities between the concept above as in [25], which identified the following pairs of concepts, which exhibit a strong mutual influence:

- *Patient's satisfaction – Age*
- *Patient's satisfaction – Quality of life*, and
- *Patient's satisfaction – Readmission rate*

More specifically, the conditional probabilities are as follows:

- $P(\text{Patient's satisfaction}|\text{Age}) = 0.73$

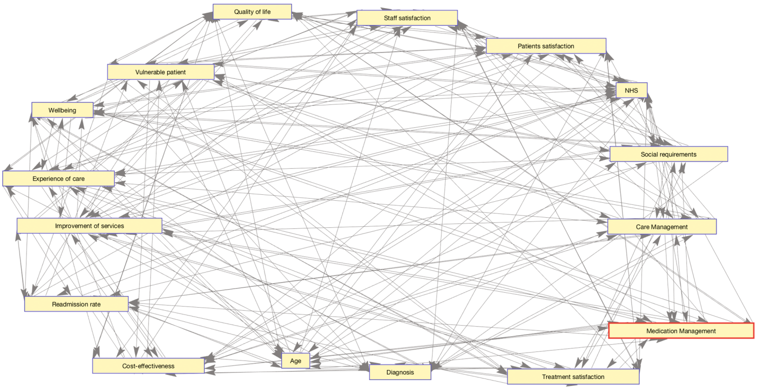


Fig. 1. The relational network as discussed in Sect. 4.

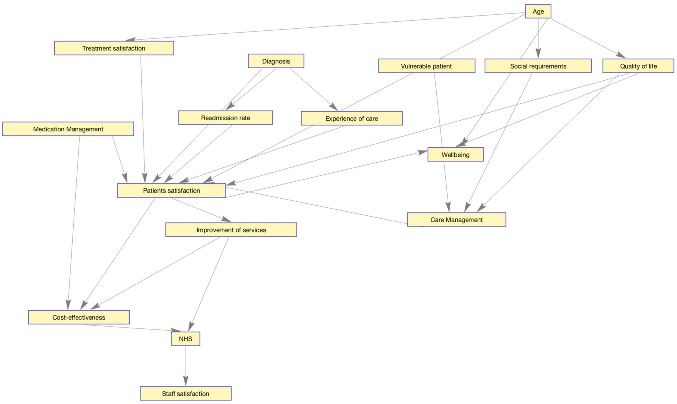


Fig. 2. The BN extracted from the network depicted in Fig. 1.

- $P(\text{Patient's satisfaction}|\text{Quality of life}) = 0.59$
- $P(\text{Patient's satisfaction}|\text{Readmission rate}) = 0.61$

As discussed above, approximately 3000 tweets were analysed, which were selected based on the following keywords:

- Patient's satisfaction
- Hospital
- NHS

Figure 3 depicts the polarity based on positive, negative, neutral keywords, as well as their compound (overall) value for each tweet.

More specifically, we found the following values:

- Positive polarity: mean = 0.076 standard deviation = 0.12
- Negative polarity: mean = 0.13 standard deviation = 0.15

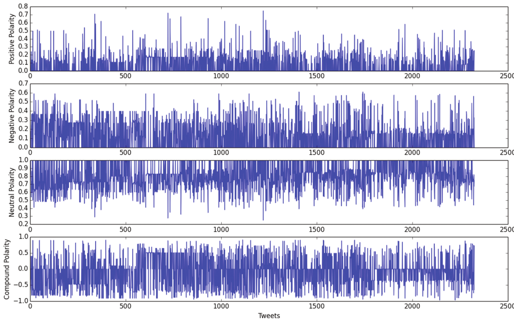


Fig. 3. Polarity values of the tweets extracted as described in Sect. 4.

- Neutral polarity: mean = 0.79 standard deviation = 0.16
- Compound (overall) polarity: mean = −0.1 standard deviation = 0.48

Furthermore, we calculated the correlation coefficient between positive polarity and overall polarity, and negative polarity and overall polarity, which gave 0.72 and −0.81, respectively. This indicates a strong positive and negative correlation between these pairs of concepts, respectively.

The above suggests that the overall polarity is close to being neutral with a significant fluctuation measured by the corresponding standard deviation. However, such tweets were identified just after the terrorist attack in Manchester, and as consequence many tweets contained negative sentiment related to such event.

Combining the above two analyses, *Age, Quality of life,* and *Readmission rate* play an important role on patient satisfaction, and it appears that different age ranges are particularly significant. Even though the overall polarity is close to neutral and closely linked to positive polarity values, this can only be applied to social media users, which usually does not include older patients. Furthermore, considering the recent terrorist attack in Manchester, this suggest that negative polarity values may have been affected. Overall, the closeness of the overall polarity to neutral level (represented in this case by 0) and statistical measures discussed above, suggest that the mood is fluctuating and it is directly linked with positive polarity values.

5 Conclusion

In this paper we have discussed a preliminary approach to determine and assess the contributory factors in patient satisfaction, based on an automated knowledge extraction method from articles from PubMed, as well from Twitter feeds. As discussed above, this is part of a much wider research project, and future investigations will focus on the integration of a variety of tools to combine the extraction of actionable intelligence from a variety of sources. This will also include an enhanced sentiment analysis of social platforms to provide a fully automated approach to patient wellbeing and satisfaction.

References

1. Cottle, M., Kanwal, S., Kohn, M., Strine, T., Treister, N.: Transforming health care through big data: strategies for leveraging big data in the health care industry. Report (2013). https://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.ssl.cf2.rackcdn.com/iHT2_BigData_2013.pdf. Accessed 27 Mar 2017
2. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**(1), 1–10 (2014)
3. Lavalle, S., Lesser, E., Shockley, R., Kruschwitz, N.: Big data, analytics and the path from insights to value researchgate. https://www.researchgate.net/publication/284611187_Big_Data_Analytics_and_the_Path_From_Insights_to_Value. Accessed 27 Mar 2017
4. Peacock, H.A.: Family and friends tests may not give us the answers we are looking for. Br. Med. J. **346** (2013)
5. Picker Institute: Transforming health care through big data: using patient feedback: a practical guide to improving patient experience. Report (2009). http://www.nhssurveys.org/Filestore/documents/QIFull.pdf. Accessed 28 Mar 2017
6. Richards, T.: Listen to patients first. Br. Med. J. **349**, 5765–5765 (2014)
7. Lomas, J.: Using research to inform healthcare managers 'And Policy Makers' questions. Healthc. Policy **31**(1), 55–71 (2005)
8. Draper, M., Cohen, P., Buchan, H.: Seeking consumer views: what use are results of hospital patient satisfaction surveys? Int. J. Qual. Health Care **13**(6), 463–468 (2001)
9. Bradley, E.H., Holmboe, E.S., Mattera, J.A., Roumanis, S.A., Radford, M.J., Krumholtz, H.M.: Data feedback efforts in quality improvement: lessons learned from US hospitals. Qual. Saf. Health Care **13**(1), 26–31 (2004)
10. Kaplan, H.S., Fastman, B.R.: Organization of event reporting data for sense making and system improvement. Qual. Saf. Health Care **12**(suppl 2), ii68–ii72 (2003)
11. Gandhi, T.K., Graydon-Baker, E., Huber, C.N., Whittemore, A.D., Gustafson, M.: Closing the loop: follow-up and feedback in a patient safety program. Joint Comm. J. Qual. Patient Saf. **31**(11), 614–621 (2005)
12. LaVela, S., Gallan, A.: Evaluation and measurement of patient experience. Patient Exper. J. **1**(1), 28–36 (2014)
13. Harrison, R., Walton, M., Healy, J., Smith-Merry, J., Hobbs, C.: Patient complaints about hospital services: applying a complaint taxonomy to analyse and respond to complaints: Table 1. Int. J. Qual. Health Care **28**(2), 240–245 (2016)
14. Lees, C.: Measuring the patient experience. Nurse Res. **19**(1), 25–28 (2011). (through 2013), London
15. Ohlhorst, F.J.: Big Data Analytics: Turning Big Data into Big Money. Wiley, New York (2012)
16. Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge (2010)
17. Duda, R., Hart, M.: Pattern Classification and Science Analysis. Wiley, New York (1973)
18. De Marneffe, M.F., MacCartney, B., Manning, C.D.: Generating typed dependency parses from phrase structure parses. In: LREC (2006)
19. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers, San Francisco (2012)
20. Manning, C.D.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)

21. Hutto, C.J., Gilbert, E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International AAAI Conference on Weblogs and Social Media (2014)
22. PubMed. http://www.ncbi.nlm.nih.gov/pubmed/. Accessed 1 May 2017
23. Trovati, M., Bessis, N.: An influence assessment method based on co-occurrence for topologically reduced big data sets. Soft Comput. **20**, 2021–2030 (2015). Springer, Berlin, Heidelberg
24. Trovati, M., Castiglione, A., Bessis, N., Hill, R.: A kuramoto model based approach to extract and assess influence relations. In: Li, K., Li, J., Liu, Y., Castiglione, A. (eds.) ISICA 2015. CCIS, vol. 575, pp. 464–473. Springer, Singapore (2016). doi:10.1007/978-981-10-0356-1_49
25. Trovati, M., Hayes, J., Palmieri, F., Bessis, N.: Automated extraction of fragments of bayesian networks from textual sources. Applied Soft Computing (2017, Submitted)
26. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco (1998)
27. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
28. Bird, S., Loper, E., Klein, E.: Natural Language Processing with Python. O'Reilly Media Inc., Sebastopol (2009)
29. vaderSentiment 2.5. https://pypi.python.org/pypi/vaderSentiment. Accessed 1 May 2017
30. Hagberg, A., Schult, D., Swart, P.: Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux, G., Vaught, T., Millman, J. (eds,) Proceedings of the 7th Python in Science Conference (SciPy2008), (Pasadena, CA USA), pp. 11–15 (2008)

A Survey of Topological Data Analysis (TDA) Methods Implemented in Python

Jeffrey Ray and Marcello Trovati(✉)

Department of Computer Science, Edge Hill University, Ormskirk, UK
{rayj,trovatim}@edgehill.ac.uk

Abstract. The aim of this paper is to provide an overview of the current Python implementation to address a variety of topics in Topological Data Analysis, which include Persistent Homology, Manifold Learning and Mapper. We will discuss the effectiveness of each process based upon existing literature where TDA has been investigated. The purpose of this work is to inform future research efforts focusing on the implementation of TDA methods for managing and discovering, patterns, and trends for Big Data.

1 Introduction

The extraction of meaningful and actionable information from Big Data is at the core of much of the research in this field [1]. In fact, organisations and companies are likely to store enormous quantities of data, which are continuously produced by their external and internal processes. Understanding and assessing the value and relevance of actionable information contained within this wealth of data, is a key challenge.

Topological Data Analysis (TDA) is a new research area, which utilises topological concepts to classify and analyse data [2]. Broadly speaking, the focus of TDA is on the structure and *shape* of data, in terms of the interconnections between its components. One aspect of topology is the ability to classify objects based on their common properties, or in other words, those exhibiting invariant features. A example in topology, is that by stretching and deforming a doughnut, it can be turned into a coffee mug, as depicted in Fig. 1. In fact, there is a "hole" in both the handle of the mug and at the centre of the doughnut.

In many clustering and classification algorithms, the geometrical properties of the data investigated play an important role. In fact, many of such approaches focus on the concept of distance, which has to be defined according to the mathematical spaces where the data is embedded [3]. On the other hand, TDA focuses on the notion of shape and connectivity that such data exhibit. As in the example above, although a doughnut is clearly different from a coffee mug, they share the common property of having a single hole in them, which can be used to classify them in terms of their connected components [3]. In fact, no matter how much they are stretched, they will still be characterised by the existence of such a hole.

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*,
Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6_54

Fig. 1. Transforming a doughnut into a coffee mug

Topology is usually specified by the metric of its corresponding space. Therefore, there is no strict requirement to limit to a specific coordinate system, thus allowing a more flexible approach [3].

A crucial aspect of TDA is the concept of *persistent homology*, which focuses on the identification of the topological properties which remain invariant [4]. Figure 2 depicts three separate datasets, which have in common the distinctive (topological) hole. Despite being different at a micro level, they could be regarded to be sufficiently similar. Persistent topology allows a classification of datasets based on such properties [5]. In particular, it is characterised by homology groups, so that two shapes are considered similar if they are isomorphic equivalent. In other words, datasets which can be topologically deformed and stretched into the same shape would fall into the same category. The three datasets in Fig. 2 could be intuitively interpreted as referring to a similar scenario, with some added noise. In fact, TDA allows noisy data to be addressed in a more efficient manner [2]. On the other hand, the two objects depicted in Fig. 3 are not topologically equivalent, as they have different connected components in terms of their number of holes.

One of the fundamental concepts of persistent topology, and homology in general, is simplicial complexes. These are space triangulations defined by combined, non-overlapping polyhedra, covering a topological space. A trivial, yet informative example of a space triangulation is image pixellation, where a real image is covered with pixels, to provide an accurate representation. One of the most important aspects of simplicial complexes is that fact that they provide an "approximation" of the object they are covering. Examples of triangulations include Voronoi diagrams, Delaunay triangulations, Vietoris and Čech complexes. Broadly speaking, they are defined by either a specific distance, or in terms of ball intersections whose centres are the data points. For more details, refer to [4]. Furthermore, the adjacency graphs generated by these triangulations can provide a variety of information on their invariant topological properties, and therefore relevant to TDA investigation [2].

The aim of this paper is to provide a survey of the Python implementations which can be used in TDA. The paper follows the following structure: in Sect. 2

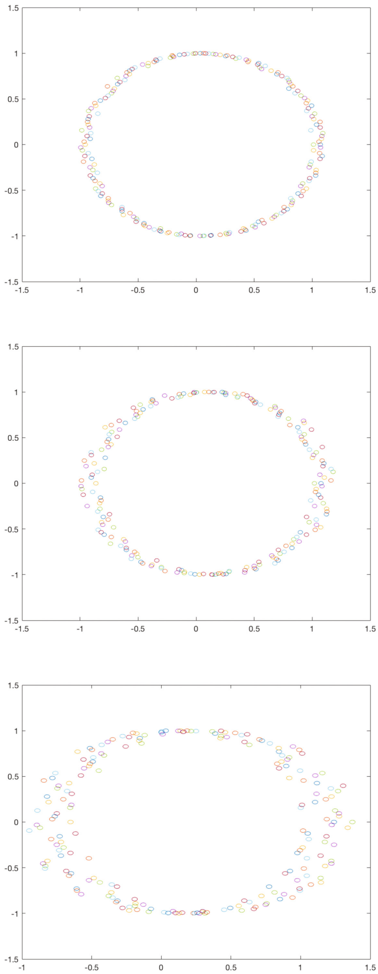


Fig. 2. Three datasets with similar topological properties.

Fig. 3. Two sets with different topological properties.

the existing technology is discussed, and Sect. 3 will discuss specific features of some Python methods and algorithms currently used in this research field. Section 4 will present the final remarks and future research.

2 State-of-the-Art TDA Methods

In [6], the authors discuss the application of the mapper TDA to noisy computer network data collected by a "darknet", with the goal of identifying activities such as port scanning and DDos attacks. In particular, they consider a dataset comprising over 3 million data points, which would prove difficult to visualise using traditional methods. Previously, software packages such as Suricata [8] would be utilised. However, it would not be sufficiently robust to successfully address noise whilst identifying all the attack patterns. The default Mapper code has been recently extended with KeplerMapper [9] for Python 3 (instead of Scikit-Learn [10]), and utilises C for the DBScan clustering to improve efficiency.

Motivated by the fact that large unstructured data sets can be visualised to provide valuable information, in [11] a commercial implementation based on TDA is described. Its objective is to provide a reliable diagnosis system to analyse large biomedical datasets, which has been shown to have high accuracy. In order to achieve this, TDA is used to define specific artificial neural networks coupled with Kolmogorov-Smirnov test [12]. The software package Ayasdi [7] is another commercial package to provide topological data analysis.

Biomedical data are often incomplete. Many analysis methods suffer when incomplete rows must be omitted, as valuable data might be subsequently removed. In comparison, the use of TDA does not require incomplete rows to be removed and can therefore produce information using all available information.

The use of TDA as a first stage in data analysis enables a wider range of initial data to be sampled, creating a more detailed grouping of data, regardless of any missing or incomplete fields. In fact, TDA allows for generality, as any notion of similarity can be successfully exploited in this context. On the other hand, classic machine learning algorithms typically require a comparably high level of similarity to produce any meaningful output.

2.1 Topological Algorithms

The methods discussed above fall into three main categories: persistent homology and Mapper and Manifold Learning, which will be discussed in this section, with particular focus on their Python implementation.

As described in Sect. 1, persistent homology is an algebraic method to assess topological features of shapes, via suitable triangulations, simplicial complex are defined in terms of a distance function of the underlying space [13]. This algorithm is relatively robust and well tested in the academic field [14]. An implementation of the persistent homology algorithm is found in the Python library Dionysus [15], which contains a variety of algorithms (Lower-Star, Vietoris-Rips, etc.) to cater for a large number of data shapes and types. Furthermore, Dionysus is a Python interpreter, which has been shown to be computationally efficient (but slower than a pure C++ implementation), whilst producing accurate data representation [15].

Manifold Learning, allows data of high dimensionality to be visualised by extracting key underlying parameters which form a low dimensional manifold.

Manifold Learning has various approaches, each implementing the extraction of data in a slightly different manner. The ability to utilise many of these methods can be achieved through the widely used and well document Python library Scikit-Learn [16]. This Python library allows Locally Linear Embedding, Modified Locally Linear embedding, Hessian Eigen mapping, Spectral Embedding, Local Tangent Space alignment, Multi-dimensional Scaling and $t-$distributed Stochastic Neighbour Embedding. With such a wide and carefully implemented set of Manifold Learning algorithms, it is possible to carry out TDA allowing for rapid and accurate data analysis of almost any data type.

Mapper is a widely utilised algorithm [17], which has also been successfully implemented in Python as Python Mapper [18]. The Mapper algorithm uses multiple filter functions to allow a coordinate system to represent a set of data, thus reducing complexity of the data set via dimensionality reduction [19]. Data points are connected via non-empty intersections between identified clusters, resulting in a topological summary of the data [20]. The Mapper algorithm can be successfully applied to cloud point data analysis, and it offers an alternative solution to the Dionysus implementation. However, the benefit of Mapper compared with Dionysus, is that its implementation is in pure Python as opposed to a port from C++. This creates a more efficient and more manageable code when used in a live environment.

3 Discussion

As discussed above, the Dionysus and Mapper algorithms focus on point cloud data sets, which implies that the data must be embedded onto a specific coordinate system. Mapper also supports two-dimensional and one-dimensional data sets defined as two-dimensional vector input data and one dimensional pairwise distances, respectively. This approach enables a more flexible and efficient method of analysing data.

Even though the Dionysus library also allows two-dimensional inputs, it introduces a limit in the construction of an alpha shape filtration [21]. As a consequence, from a data analysis point of view, the ease and flexibility of the Mapper algorithm and Python Mapper solution is preferable to the Dionysius library.

The Manifold Learning algorithms contained within the Scikit-Learn package require the data to be embedded onto a low dimensional sub-manifold, as opposed to the Mapper algorithm, which allows higher dimensionality. In particular, they need the dataset to be locally uniform and smooth often with restriction on sampling uniformity. In contrast, the Mapper algorithm output is not intended to faithfully reconstruct the data or reform the data to suit a data model, as it provides a representation of the data structure.

The Manifold learning and Mapper solutions provide a useful set of data analysis tools, which enable a suitable representation of data structures. Furthermore, they provide different methods based on specific assumptions on data inputs and output. Therefore, a choice between them can only be made once the corresponding data set has been created and its structure is known. The ease of implementation due to both libraries being native to the Python programming language, allows an integration with other popular data science Python packages. This clearly facilitates the creation of an efficient and computationally feasible data analysis platform.

The Mapper algorithm has been extensively used for commercial data analysis tools, due to its ability to process complex data sets containing over 500,000 features. This also enables complex Big Data to be handled without the requirements and complexity of deploying Hadoop, map reduce and SQL data base, prove the flexibility and reliability of the algorithm. Figure 4 shows details of the performance of the above algorithms.

Method	Properties	Limitations	Strengths
Python Mapper	Native Python Combines Filter functions, the Mapper algorithm and visualisation results	A pure python implementation, slow than a traditional programming language implementation such as C	Robust algorithm capable of relative equal output regardless of sample size. Easy integration with Other Native python libraries
Dionysus (Persistent Homology)	C++, with Python bindings Persistent homology computation, Vineyards Persistent cohomology computation, Zigzag persistent homology.	Difficult installation, requires a large list of dependencies Not all functionality is completely available in Python	Highly resistant to data noise.
SciKit Learn (Manifold Learning)	Native Python. Non-linear dimensionality reduction. Wide variety of possible approaches.	Same scale must be used over all features - Manifold learning uses nearest-neighbour search. Certain input configurations can lead to singular weight matrices (more than two points identical)	Able to adapt Linear Frameworks to be sensitive to non-linear structures in data. Comprehensive community support Integrates seamlessly with widely used Machine learning libraries already available in Python

Fig. 4. Performance data for networkX and Graph tools.

4 Conclusion

The manifold learning and Mapper methods for TDA allow greater flexibility and data input sources compared to the Dionysus implementation. Furthermore, highly specialised data scientists have been contributing to the Scikit-Learn and Mapper projects, creating a peer reviewed and well maintained implementation, which is not the case for Dionysus. Industry Standard analysis tools, such as Ayasdi, have been built with the Mapper algorithm at their core, demonstrating their industry appeal and credibility. This suggests that data analysis platforms aiming to obtain efficiently interpreted results with applications to a decision-making process should prioritise the implementation of the Mapper algorithm.

References

1. Wang, L., Wang, G., Alexander, C.A.: Big data and visualization: methods, challenges and technology progress. Digit. Technol. Digit. Technol. **1**(1), 33–38 (2015)
2. Carlsson, G., Harer, J.: Topology and data. Bull. Math. Soc. **46**(2), 255–308 (2009)
3. Janich, K.: Topology. Springer, New York (1984)
4. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society, New York (2010)
5. Weinberger, S.: What is persistent homology? Not. AMS **58**(1), 36–39 (2011)
6. Coudriau, M., Lahmadi, A., Francois, J.: Topological analysis and visualisation of network monitoring data: darknet case study. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6 (2016)
7. Ayasdi (2017). https://www.ayasdi.com. Accessed 24 May 2017
8. Suricata (2017). https://suricata-ids.org/. Accessed 24 May 2017
9. MLWave: kepler-mapper: KeplerMapper is a Python Class for Visualization of High-Dimensional Data and 3-D Point Cloud Data. kepler-mapper (2017)
10. Scikit-learn: machine learning in Python scikit-learn 0.18.1 documentation (2017)
11. Rucco, M., Falsetti, L., Herman, D., Petrossian, T., Merelli, E., Nitti, C., Salvi, A.: Using topological data analysis for diagnosis pulmonary embolism. Phys. Q-Bio (2014). arXiv14095020
12. Upton, G., Cook, I.: Kolmogorov-Smirnov test - Oxford Reference. http://www.oxfordreference.com.edgehill.idm.oclc.org/view/10.1093/acref/9780199679188.001.0001/acref-9780199679188-e-868. Accessed 24 May 2017
13. Goodman, J.E.: Surveys on Discrete and Computational Geometry: Twenty Years Later: AMS-IMS-SIAM Joint Summer Research Conference. American Mathematical Society (2008)
14. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: Proceedings 41st Annual Symposium on Foundations of Computer Science, pp. 454–463 (2000)
15. Morozov, D.: Welcome to Dionysus documentation. http://www.mrzv.org/software/dionysus/. Accessed 29 May 2017
16. Scikit-Learn: Manifold learning scikit-learn 0.18.1 documentation. http://scikit-learn.org/stable/modules/manifold.html. Accessed 29 May 2017
17. Singh, G., Memoli, F., Carlsson, G.: Mapper: a topological mapping tool for point cloud data. In: Eurographics Symposium on Point-Based Graphics (1991)
18. Müllner, D., Babu, A.: Python Mapper: an open-source toolchain for data exploration, analysis, and visualisation. Stanf, Edumuellnermapper (2013)
19. Carlsson, G.: Topology and data. Bull. Am. Math. Soc. **46**(2), 255–308 (2009)
20. Chow, Y.Y.: Application of Data Analytics to Cyber Forensic Data A Major Qualifying Project Report. MITRE Corporation, McLean (2016)
21. Giesen, J., Cazals, F., Pauly, M., Zomorodian, A.: The conformal alpha shape filtration. Vis. Comput. **22**(8), 531–540 (2006)

Author Index

L. Barolli et al. (eds.), *Advances in Intelligent Networking and Collaborative Systems*, Lecture Notes on Data Engineering and Communications Technologies 8,
DOI 10.1007/978-3-319-65636-6

X

Y

Z