

Xin-She Yang
Simon Sherratt
Nilanjan Dey
Amit Joshi *Editors*

Fourth International Congress on Information and Communication Technology

ICICT 2019, London, Volume 1

Advances in Intelligent Systems and Computing

Volume 1041

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing,
Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagras, School of Computer Science and Electronic Engineering,
University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University,
Gyor, Hungary


Vladik Kreinovich, Department of Computer Science, University of Texas
at El Paso, El Paso, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao
Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology,
University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute
of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro,
Rio de Janeiro, Brazil

Ngoc Thanh Nguyen , Faculty of Computer Science and Management,
Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at <http://www.springer.com/series/11156>

Xin-She Yang · Simon Sherratt ·
Nilanjan Dey · Amit Joshi
Editors

Fourth International Congress on Information and Communication Technology

ICICT 2019, London, Volume 1

 Springer

Editors

Xin-She Yang
School of Science and Technology
Middlesex University
London, UK

Simon Sherratt
University of Reading
Reading, UK

Nilanjan Dey
Department of Information Technology
Techno India College of Technology
Kolkata, West Bengal, India

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-981-15-0636-9

ISBN 978-981-15-0637-6 (eBook)

<https://doi.org/10.1007/978-981-15-0637-6>

© Springer Nature Singapore Pte Ltd. 2020, corrected publication 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This AISC volume contains the papers presented at ICICT 2019: Fourth International Congress on Information and Communication Technology in concurrent with ICT Excellence Awards. The conference was held during February 25–26, 2019, London, UK, and collaborated by Global Knowledge Research Foundation, City of Oxford College. The associated partners were Springer, InterYIT IFIP, and Activate Learning. The conference was held at Brunel University London. This conference was focused on e-business fields such as e-agriculture, e-education, and e-mining. The objective of this conference was to provide a common platform for researchers, academicians, industry persons, and students to create a conversational environment where in topics related to future innovation, obstacles to be resolved for new upcoming projects, and exchange of views and ideas. The conference attracted immense experts from more than 45 countries and deep discussions were held, and the issues were intended to be solved at international level. New technologies were proposed, experiences were shared, and future solutions for design infrastructure for ICT were also discussed. Research submissions in various advanced technology areas were received and then were reviewed by the committee members; 92 papers were accepted. The conference was overwhelmed by the presence of various members. Amit Joshi, Organizing Secretary, ICICT 2019, gave the welcome speech on behalf of conference committee and editors. Our special invitee guest—Sean Holmes, Vice Dean, International College of Business, Arts and Social Sciences, Brunel University London, UK—also addressed the conference by a speech. The conference was also addressed by our inaugural guest and speakers—Mike Hinchey, Chair, IEEE, UK and Ireland section, Director of Lero, and Professor of Software Engineering at the University of Limerick, Ireland; Aninda Bose, Sr. Publishing Editor, Springer Nature. Niko Phillips, Group Director, International Activate Learning, City of Oxford College, UK, addressed the Vote of Appreciation on behalf of the conference committee. There were 12 technical sessions in total, and talks on academic and industrial sector were focused on both the days. We are obliged to Global Knowledge Research Foundation for their immense support for making this conference a successful one. A total of 85 papers were presented in technical sessions

and 92 were accepted with strategizing on ICT and intelligent systems. At the closing ceremony, ten Best Paper Awards by Springer were announced among the best selected and presented paper. A 200 euro voucher to shop online at <http://www.springer.com> was given along with appreciation certificate by Springer and Editor of ICICT 2019. On behalf of the editors, we thank all sponsors, press, print, and electronic media for their excellent coverage of this conference.

London, UK
Reading, UK
Kolkata, India
Gandhinagar, India
February 2019

Xin-She Yang
Simon Sherratt
Nilanjan Dey
Amit Joshi

Contents

Wearable Device Technology in Healthcare—Exploring Constraining and Enabling Factors	1
Mike Krey	
Proposed System for Effective Adoption of E-government to Obtain Construction Permit in Egypt	15
Heba Fawzy and Dalia A. Magdi	
Potential Use of Bitcoin in B2C E-commerce	33
Ralf-Christian Härting and Christopher Reichstein	
Smart Cabin: A Semantic-Based Framework for Indoor Comfort Customization Inside a Cruise Cabin	41
Atieh Mahroo, Daniele Spoladore, Massimiliano Nolich, Raol Buqi, Sara Carciotti and Marco Sacco	
Formal Modeling and Analysis of Probabilistic Real-Time Systems	55
Christian Nigro, Libero Nigro and Paolo F. Sciammarella	
Regional Agricultural Land Classification Based on Random Forest (RF), Decision Tree, and SVMs Techniques	73
Nassr Azeez, Wafa Yahya, Inas Al-Taie, Arwa Basbrain and Adrian Clark	
On Hierarchical Classification Implicative and Cohesive M_{GK}-Based: Application on Analysis of the Computing Curricula and Students Abilities According the Anglo-Saxon Model	83
Hery Frédéric Rakotomalala and André Totohasina	
A Modeling Environment for Dynamic and Adaptive Network Models Implemented in MATLAB	91
S. Sahand Mohammadi Ziabari and Jan Treur	
Face Authentication Using Image Signature Generated from Hyperspectral Inner Images	113
Guy Leshem and Menachem Domb	

Unsupervised Learning of Image Data Using Generative Adversarial Network	127
Rayner Alfred and Chew Ye Lun	
Adaptive Message Embedding in Raw Images	137
Tamer Shanableh	
Megapolis Tourism Development Strategic Planning with Cognitive Modelling Support	147
Alexander Raikov	
Design and Implementation of Ancient and Modern Cryptography Program for Documents Security	157
Samuel Sangkon Lee	
Methodology for Selecting Cameras and Its Positions for Surround Camera System in Large Vehicles	171
S. Makarov Aleksei and V. Bolsunovskaya Marina	
Fuzzy Models and System Technical Condition Estimation Criteria ...	179
Gennady Korshunov, Vladimir Smirnov, Elena Frolova and Stanislav Nazarevich	
Software Tools and Techniques for the Expert Systems Building	191
Arslan I. Enikeev, Rustam A. Burnashev and Galim Z. Vakhitov	
A Component-Based Method for Developing Cross-Platform User Interfaces for Mobile Applications	201
Marek Beranek and Vladimir Kovar	
Improvement of Vehicles Production by Means of Creating Intelligent Information System for the Verification of Manufacturability of Design Documentation	219
Irina Makarova, Ksenia Shubenkova, Timur Nikolaev and Krzysztof Żabiński	
Group Delay Function Followed by Dynamic Programming Versus Multiscale-Product for Glottal Closure Instant Detection	229
Ghaya Smidi and Aicha Bouzid	
Activity Logging in a Bring Your Own Application Environment for Digital Forensics	241
Bernard Shibwabo Kasamani and Duncan Litunya	
Game Theory for Wireless Sensor Network Security	259
Hanane Saidi, Driss Gretete and Adnane Addaim	

DSP Implementation of OQPSK Baseband Demodulator for Geosynchronous Multichannel TDMA Satellite Receiver 271
 A. Chinna Veeresh, S. Venkata Siva Prasad, S. V. Hari Prasad, M. Soundarakumar and Vipin Tyagi

Textile Sensor-Based Exoskeleton Suits for the Disabled 285
 P. R. Sriram, M. Muthu Manikandan, Nithin Ayyappaa and Rahul Murali

A Comparison of Indoor Positioning Systems for Access Control Using Virtual Perimeters 293
 Brian Greaves, Marijke Coetzee and Wai Sze Leung

q-LMF: Quantum Calculus-Based Least Mean Fourth Algorithm 303
 Alishba Sadiq, Muhammad Usman, Shujaat Khan, Imran Naseem, Muhammad Moinuddin and Ubaid M. Al-Saggaf

Process Driven Access Control and Authorization Approach 313
 John Paul Kasse, Lai Xu, Paul de Vrieze and Yuewei Bai

Comprehensive Exploration of Game Reviews Extraction and Opinion Mining Using NLP Techniques 323
 Stefan Ruseti, Maria-Dorinela Sirbu, Mihnea Andrei Calin, Mihai Dascalu, Stefan Trausan-Matu and Gheorghe Militaru

MHAD: Multi-Human Action Dataset 333
 Omar Elharrouss, Noor Almaadeed and Somaya Al-Maadeed

Machine Learning-Based Approaches for Location Based Dengue Prediction: Review 343
 Chamalka Seneviratne Kalansuriya, Achala Chathuranga Aponso and Artie Basukoski

Critical Evaluation of Different Biomarkers and Machine-Learning-Based Approaches to Identify Dementia Disease in Early Stages 353
 Gayakshika Gimhani, Achala Chathuranga Aponso and Naomi Krishnarajah

Interactive Visualization of Ontology-Based Conceptual Domain Models in Learning and Scientific Research 365
 Dmitry Litovkin, Anton Anikin and Marina Kultsova

A Secure Lightweight Mutual Authentication and Message Exchange Protocol for IoT Environments Based on the Existence of Active Server 375
 Omar Abdulkader, Alwi M. Bamhdi, Vijey Thayananthan, Kamal Jambi, Bandar Al Ghamdi and Ahmed Patel

Monetary Transaction Fraud Detection System Based on Machine Learning Strategies	385
Lakshika Sammani Chandradeva, Thushara Madushanka Amarasinghe, Minoli De Silva, Achala Chathuranga Aponso and Naomi Krishnarajah	
The Impact of Service-Level Agreement (SLA) on a Cloudlet Deployed in a Coffee Shop Scenario	397
Mandisa N. Nxumalo, Matthew O. Adigun and Ijeoma N. Mba	
A New Use of Doppler Spectrum for Action Recognition with the Help of Optical Flow	407
Meropi Pavlidou and George Zioutas	
Meeting Challenges in IoT: Sensing, Energy Efficiency, and the Implementation	419
Toni Perković, Slaven Damjanović, Petar Šolić, Luigi Patrono and Joel J. P. C. Rodrigues	
Small Cells Handover Performance in Centralized Heterogeneous Network	431
Raed Saadon, Raid Sakat, Maysam Abbod and Hasanein Hasan	
CRISP-DM/SMEs: A Data Analytics Methodology for Non-profit SMEs	449
Jhon Montalvo-Garcia, Juan Bernardo Quintero and Bell Manrique-Losada	
Bridges Strengthening by Conversion to Tied-Arch Using Monarch Butterfly Optimization	459
Orlando Gardella, Broderick Crawford, Ricardo Soto, José Lemus-Romani, Gino Astorga and Agustín Salas-Fernández	
Deep Learning Approach for IDS	471
Zhiqiang Liu, Mohi-Ud-Din Ghulam, Ye Zhu, Xuanlin Yan, Lifang Wang, Zejun Jiang and Jianchao Luo	
Outlier Detection Method-Based KPCA for Water Pipeline in Wireless Sensor Networks	481
Mohammed Aseeri, Oussama Ghorbel, Hamoud Alshammari, Ahmed Alabdullah and Mohamed Abid	
Data Extraction and Exploration Tools for Business Intelligence	489
Mário Cardoso, Tiago Guimarães, Carlos Filipe Portela and Manuel Filipe Santos	
Systems and Methods for Implementing Deterministic Finite Automata (DFA) via a Blockchain	499
Craig S. Wright	

**Closest Fit Approach Through Linear Interpolation to Recover
Missing Values in Data Mining** 513
Sanjay Gaur, Darshanaben D. Pandya and Deepika Soni

**Correction to: Process Driven Access Control and Authorization
Approach** C1
John Paul Kasse, Lai Xu, Paul de Vrieze and Yuewei Bai

Author Index 523

About the Editors

Xin-She Yang completed his D.Phil. in Applied Mathematics at the University of Oxford. He then worked at Cambridge University and the National Physical Laboratory (UK) as a Senior Research Scientist. He is currently a Reader in Modelling and Optimization at the Middlesex University London and an Adjunct Professor at Reykjavik University (Iceland). He is also an elected Bye Fellow at Cambridge University and the IEEE CIS Chair for the Task Force on Business Intelligence and Knowledge Management. He was included in the “2016 Thomson Reuters Highly Cited Researchers” list.

Simon Sherratt is a Professor of Consumer Electronics at the University of Reading. His research has primarily been on OFDM, particularly for digital TV and wireless USB. He specializes in DSP system architecture and hardware implementation, and is leading the development of wireless communications and on-body sensors for Reading’s part in Sphere. Prof. Sherratt is the current Editor-in-Chief of IEEE Transactions on Consumer Electronics and an elected member of the IEEE Consumer Electronics Society Board of Governors. In addition, he is a Fellow of the IET and IEEE.

Nilanjan Dey is an Assistant Professor at the Department of Information Technology, Techno India College of Technology, Kolkata, India. He holds an honorary position of Visiting Scientist at Global Biomedical Technologies Inc., CA, USA, and is a Research Scientist at the Laboratory of Applied Mathematical Modeling in Human Physiology, Territorial Organization of Scientific and Engineering Unions, Bulgaria. He has published 20 books and more than 300 international conference and journal papers. He is Editor-in-Chief of the International Journal of Ambient Computing and Intelligence, International Journal of Rough Sets and Data Analysis, International Journal of Synthetic Emotions (IJSE), and International Journal of Natural Computing Research.

Amit Joshi is Director of the Global Knowledge Research Foundation, Ahmedabad, India. He holds a BTech in Information Technology and an MTech in Computer Science and Engineering, as well as a doctorate in Computer Science & Engineering with a specialization in the areas of cloud computing and cryptography. He is an active member of the ACM, CSI, AMIE, IACSIT-Singapore, IDIES, ACEEE, NPA, and many other professional societies. He has more than 50 papers and 25 edited books to his credit.

Wearable Device Technology in Healthcare—Exploring Constraining and Enabling Factors



Mike Krey

Abstract The aim of this literature review is to investigate enabling and constraining factors of wearable devices in healthcare. While offering patients a better quality of life as they may spend less time in hospitals, wearables can also play a key role in solving the current crises in the health sector. 1'195 articles were screened, and 41 papers in total were analyzed for the review. Most studies focused on product design, specifically user acceptance and user adaption. Some studies investigated how machine learning can improve the accuracy and reliability of wearables or focused on the quality of treatment and how wearables can improve a patient's quality of life. However, one important aspect, how to handle big data issues like security and privacy for wearables is mostly neglected. Further research is required, dealing with the questions, how devices can become secure for patients and how the data of patients will not become accessible.

Keywords Wearable · Healthcare · Literature review · Factors

1 Introduction

The Swiss healthcare system is one of the best in the world, but also one of the most expensive ones. This is the conclusion of a recent study by the Organization for Economic Cooperation and Development (OECD). The price for this above-average health service is high since in 2016 12.4% of the Swiss gross domestic product (GDP) flowed into healthcare; the OECD average was 9.0% [1]. Therefore, it is becoming increasingly necessary for Swiss hospitals to develop concepts and reforms in order to work cost-efficiently and optimize processes sustainably.

Considering technologies that have the potential to contribute positively to rising costs for healthcare—wearable devices have already proven a positive effect on process costs, e.g., by replacing conventional health monitoring systems [2].

M. Krey (✉)

Zurich University of Applied Sciences, 8401 Winterthur, Switzerland
e-mail: Mike.Krey@zhaw.ch

The term wearable is used for devices that measure a user's data in real time using built-in sensors and then send it to a connected device.

Although they were initially designed for military applications, wearables are now enjoying increasing popularity in the mass market. Among other things, users can measure running distance, calories burned or heart rate. This first class of wearables, which includes fitness trackers and smart watches such as the Samsung Galaxy Gear is characterized by its ease of use.

The second class of wearables is used for medical purposes and, unlike fitness trackers, not only collects data on a person's physical condition but also gives medical suggestions [3]. They are either devices worn directly by the patient (wearable portable medical device) or implanted in the patient's body (embedded device), thus enabling early detection of risks such as an imminent heart attack [4].

These devices are not only be used as a fitness tracker or calorie counters but have the potential to revolutionize healthcare through their capability of collecting large amounts of data and openness of communicating with other devices. In fact, these devices show a great opportunity for monitoring patients with cardiac and circulatory troubles, diabetes or low blood sugar.

By constantly monitoring the patient's state of health, the device can issue a warning as soon as the patient's state of health is critical, thus preventing certain events such as strokes. Being not bigger than an ordinary watch, wearables give users the ability to go through daily life without restrictions. Studies even confirm that non-medical devices like the Apple Watch can already detect abnormal heart rhythm with 97% accuracy [5].

Nonetheless, there are still reasons why wearables have not yet become standard equipment in healthcare. One reason for this is that currently, the reliability of those devices is still less than of a standard medical device [6]. The danger of giving false treatment to patients due to a wrong signal is a crucial factor why doctors would not rely on such a device.

According to the Health Insurance Portability and Accountability Act (HIPAA), manufacturers of wearables can share information about their users if the data is not personally identifiable. What the manufacturer is doing with that data does not have to be stated in the terms and conditions of the company. Therefore, the user has no influence on what happens with his data [7].

Current studies already deal with the question of the acceptance of wearables [3] and their sensors [8–10].

In addition to acceptance and sensor technology, the areas of data protection and security risks raise open questions that have so far only been insufficiently discussed in the literature [11]. Lee et al. conducted the first large-scale study on security and privacy issues related to portable devices [12]. They investigated possible risks from the most common portable devices of the time.

It should be noted that existing research has currently only addressed one aspect related to wearables in healthcare. A broad and multi-layered consideration of different aspects and their implications for processes in the healthcare system is still lacking.

Therefore, this study explores key factors that motivate stakeholders to make use of wearable devices and secondly investigates constraining factors representing the greatest barriers of wearable technology in healthcare.

With this holistic approach, new findings are to be gained for science and practice so that wearables can be brought to market maturity.

The structure of this paper is as follows: In the next section, the methodology is presented. Section three gives an overview of the different categories that have a positive or restrictive impact on wearable technology. Finally, the contribution closes with the presentation of the results, limitations of the research work and an outlook for further research.

2 Methodology

In order to answer the research question, a systematic literature review was carried out in several electronic reference databases including ProQuest, PubMed and Web of Science. The search was limited to peer-reviewed articles written in English between 2000 and 2018. All research focused on factors that limit or enable wearables, especially in healthcare. In addition, the concept of portable technologies and their challenges were evaluated.

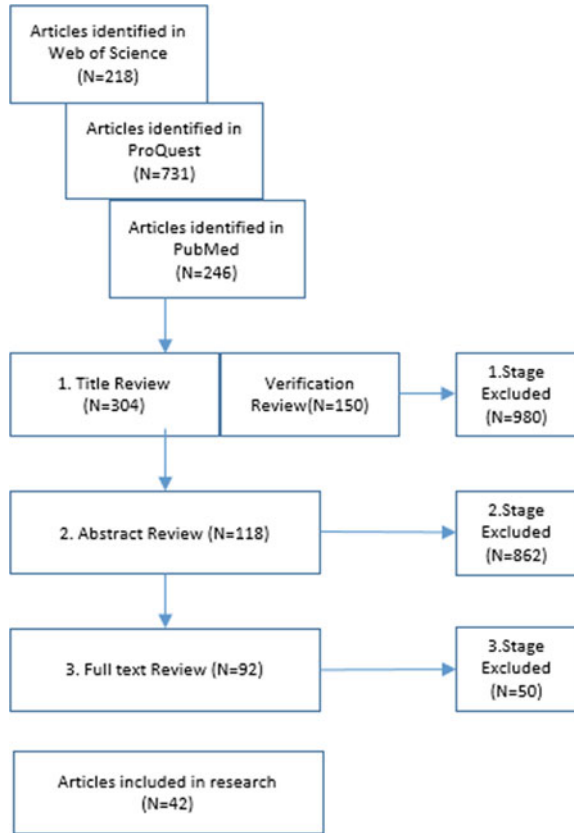
The search for suitable keywords was carried out in Google Scholar and grouped into four main topics: wearable, healthcare, risk, benefit. The search strategy was designed with a Boolean full-text search term. Different terms for keywords were used:

- wearable AND
- (healthcare or health) AND
- [(constrain OR fear OR risk OR inhibit OR barrier) OR (benefit OR acceptance OR favor OR contribute)].

A total of 731 titles were found in ProQuest, 218 titles in the Web of Knowledge and 279 titles in PubMed.

In total, 1'195 research papers were suitable for the review. The screening process was conducted in three stages. Starting with a review on the title and abstract of each contribution several duplicates could be excluded for further review. Only contributions were addressed dealing with wearables in healthcare incl. constraints and benefits of this technology. Articles which were not considered as original research, letters to the editor, comments or reviews were also excluded. Studies that focus on sports, fitness activity, health and well-being were also excluded. In the first phase, the evaluation of the titles led to a selection of 304 articles. In the second phase, the abstracts were reviewed and 118 articles were considered further. The process of the literature review is shown in Fig. 1.

Next, the contributions were categorized. The list of categories was developed by the research team using the full-text review process (cf. Sect. 3). Each paper was validated against constraining and enabling factors of wearable devices.

Fig. 1 Literature review

3 Results

A total of 118 abstracts and 92 full-text research papers were analyzed. In total, 42 of the 92 articles were included in the systematic review. Based on the review four different areas have been identified, in which wearables are already used in healthcare: quality of life, quality of treatment, product design and big data. Each area has been further divided into categories to describe the related aspects of each area (see Table 1). Each of these areas contains constraining and enabling factors that are summarized in the following sections.

3.1 *Quality of Life*

Wearables have proven to be an effective tool for the prevention, early detection and treatment of chronic diseases [8, 15]. Over the past decade, research has focused

Table 1 Areas of wearable device usage in healthcare

Area	Category	Total of articles	References
Quality of life	Acceptance	4	[3, 8, 13, 14]
	Disease related	3	[4, 15, 16]
	Usability	2	[15, 17]
	User behavior	1	[18]
Quality of treatment	Sensory	1	[19]
	Disease related	7	[14, 15, 19–23]
	Interoperability	1	[24]
Product design	Size of device	1	[25]
	Energy	5	[23, 26–29]
	Usability	3	[3, 30, 31]
	Accuracy	9	[3, 8, 25, 30, 32–36]
	Disease related	6	[3, 28, 30, 34, 35, 37]
	Acceptance	5	[3, 17, 25, 30, 38]
	User behavior	2	[3, 39]
	Clothing	2	[40, 41]
	Safety and privacy	1	[25]
Big data	Benefits	4	[42–45]
	Privacy	4	[3, 44–46]
	Ethical issues	1	[46]
	Machine learning	1	[42]
	Acceptance	1	[47]
	Security	4	[26, 44, 48, 49]
	Efficiency	1	[26]

on portable sensor technologies. Today, the focus is increasingly on improving the quality of life through portable applications [3]. Studies have shown that wearing devices leads to increased physical activity [16]. Especially people with physical inactivity, such as obese people, can sustainably improve their health and physical performance [15]. In the fitness and sports sector, the devices help to track and monitor daily fitness conditions such as steps, distance, calorie consumption, sleep and nutrition. In addition, modern devices also offer position-based tracking using the Global Positioning System (GPS) or motion tracking with acceleration sensors or gyroscopes. With visualization tools and real-time monitoring, users can track their actions, e.g., climbing stairs [50]. In everyday decisions, wearables can be an encouraging factor and strengthen self-confidence [13, 18].

3.2 *Quality of Treatment*

Two papers published in 2007 focused on remote or wireless health monitoring [20, 21]. According to [21], remote health monitoring would allow to monitor patients at any location at any time. This could be combined with situations where additional services are needed like sending an alert message to the authorities. These services can reduce the time between the event (e.g., stroke) and the arrival time of the medical team. Combined with pervasive access to medical records from the patient through the medical response team would result in a reduced number of medical errors and medical treatments [20]. The sharing of data (especially the large data populations—without personal information) not only to the medical response team but also to researchers around the world could help to produce new evidence about unknown symptoms and personal treatments. This would lead to better treatment for certain diseases [22].

Data from wearable devices can provide relevant and significant information about the patient [15]. Constant monitoring of physiological data only used to be possible at hospitals. Technological advances helped that a patient can now be monitored in real time at home. The data helps doctors to determine the symptoms more accurately [15, 23]. The wearables even can be used for remote rehabilitation treatment enabling a better quality of treatment. The feedback from the wearables can be used for tracking the rehabilitation process and adjusting it if necessary [15]. Doctors can even terminate the treatment after a short amount of time if the treatment works sufficiently or if side effects can be identified [23]. For diabetes, a painless monitoring system like a wearable is more comfortable for long-term monitoring of blood sugar than constant blood testing. This brings an alternative treatment method which increases the quality of treatment [14]. Remote healthcare monitoring also enables contacting the patient individually and directly about his medical condition. If any abnormality can be detected, then a doctor can be sent out to the patient [24].

3.3 *Product Design*

The acceptance and adoption of wearable technology in health are depending on several factors [17]. Only 1.3% of researches had investigated the perception of end users [38]. According to [3], technology acceptance, health behavior and privacy context are perspectives, which should be considered in case of consumer decision to buy a wearable device. These multiple perspectives were validated by Gao et al. [3] with an integrative model through a survey. This model explains individual's adoption of healthcare wearable from each perspective. The model was built with the unified theory of acceptance and use of technology 2 (UTAUT2). According to UTAUT2, seven direct factors are affecting consumer's intention to use new technology. In addition to technology acceptance of wearable devices, the protection motivation theory PMT was used to analyze factors related to health behavior [39]. Their finding

was that all factors from these perspectives would influence consumer's decision to buy a wearable device. Certainly, the adaptation distinguishes between fitness wearable devices and medical wearable devices. Users who consider buying fitness wearable expect more enjoyment, comfort and pricing reasonability. They give attention to the following factors: hedonic motivation, functional congruence, social influence, perceived privacy risk and perceived vulnerability. However, on a medical wearable device, they stress the importance of factors perceived expectancy, effort expectancy, self-efficacy and perceived severity [3].

For the successful development of wearable devices, consumer preferences should be understood clearly [40]. A study published in the *International Journal of Clothing Science and Technology* describes design expectation for a wearable system, called e-nose. Wearable electronic nose (e-nose) is a microenvironment that surrounds the human body in the form of clothes, can detect gas emitted from nose and skin and provide useful information for monitoring diabetes. They identified 15 design factors to meet consumer preferences. In their case, 209 diabetic participants attended an online survey. Among the 15 design factors, the participants rated safety, data accuracy, comfort in movement and portability as the most important factors [25]. Other investigation from [41] demonstrates that coupled antennas on multiple clothing layers could be a new approach to smart clothing.

In the interview-based study by Campling et al. [30] with a focus group, the researcher team determined the view of existing wearable products. They had chosen wearable devices for different use cases, such as fall detection, pill dispensing and vital sign monitoring. The lack of focus on end-user needs and design characteristic issues emerged from the interviews. Good design and ease of use were particularly important for the focus group. All participants represented the same view that the current products were not covering end-user needs. Many devices were aesthetically not interesting and non-intuitive.

Regarding the factor accuracy, many studies proved that current wearable sensor technologies result in 95–100% precision in fall detection [32]. Furthermore, the meta-analysis by Cancela et al. [37] shows remote patient monitoring for chronic heart failure patient can reduce the risk of death. This leads to a growing interest in the cardiology community. In contrast, a comparison between smartphone and wearable activity tracking clarifies that wearable devices differed more than smartphone applications [33]. Body-worn applications often report noise which can be a result of “electromagnetic interference of power line, poor quality of contact between the electrode and the skin, baseline wander caused by respiration, electrosurgical instruments and movement of the patient's body.” The noise collected can lead to different data [8]. Further, study-related disturbances such as freezing of gait (FOG) detect a lack of consistency in outcomes [35]. This is also supported by findings from Jayaraman and colleagues [36]. The authors state that different kind of data can be collected even if the location of the sensors changes.

As stated before, comfort plays a role when using a wearable [31]. This includes the battery time. In 2013, wearable medical devices came in a big size and had a power constraint. The power was limited, and the costs for replacing batteries were high [26]. A similar study concerning the energy efficiency of motion-related activity

recognition, published one year later, said that the power consumption for activity recognition remains to be a challenge [27]. The product design of wearable can also be limited by the consumption of the device. A bigger battery will be needed, which takes place and therefore the wearable must have a certain size [23]. In 2016, the authors of a paper developed a wearable cardiac monitoring system. They used a rechargeable lithium battery as a battery supply. The device was able to stream data for 3.5 h, which was more than their determined session for 1 h. They combined the battery with a low-energy Bluetooth chip, which provided a balance between physical size and power capacity [28]. A promising solution toward the limited battery is a sensor that does not require batteries. It would receive its energy from the electromagnetic field generated by the radio-frequency identification antennas. However, this technology is at an early stage of its development and still needs improvements [29].

For remote healthcare monitoring, the data collected from the devices can be evaluated by applications. This involves a user interface like a smartphone, tablet or computer. A middleware platform called EcoHealth was developed for IoT. It connects patients, healthcare providers and devices [24].

3.4 *Big Data*

With the help of wireless sensor technology, wearables send an enormous amount of data to the monitoring device. This data can then be interpreted by machine learning so that effects can be predicted and hopefully prevented [42]. Therefore, big data is the technology that makes wearable technology so incredibly potent and interesting for science [47]. The health condition of a patient is no longer evaluated by single medical examinations when a patient [1, 42] seeks out a doctor but is monitored in real time for a long duration [43]. This allows to constantly following the condition of a patient. However, this arises questions about how much the privacy of the patient is threatened by this constant surveillance. All information of the person is stored, analyzed and monitored [46]. The authors suggest that protocols should be used to allow the stakeholder, such as caregivers, commercial entities, access to medical data for being used unauthorized. This allows companies like insurances to create a detailed account of the person monitored by the wearable [44]. It will also be hard to control the information that is gathered from the patient as wearables might collect data of the everyday life which is not only health related and the users right to keep his data private is no longer granted, as the wearable constantly tracks his every life [44]. This constraining factor was investigated by several papers, among them one that examined this issue under the aspect of the “privacy calculus framework.” The researchers there found out that the perceived privacy risk indeed plays an important role in the adoption of the wearable [45]. However, as long as users see a large benefit in using the wearable they are willing to ignore the privacy issues they face using wearables [45]. Furthermore, in some cases like fall detection for elderly people, wearables are a measure that enhances their feeling of privacy because they

would otherwise require surveillance with security cameras in the hospital, in order to guarantee that they do not get severely injured by a fall when they leave their beds. Considering these papers, the privacy issues still present a barrier to wearable technology that can be overcome by giving the user the feeling that the device does truly help them so that privacy does not play the biggest role as a constraining factor in big data.

The true problem that is still a barrier is that “IoT devices were not built with security in mind and have backdoors placed on them by manufacturers” [44]. Consequently, wearables face many issues regarding data protection as well, as they are part of the IoT world. Nonetheless, research has so far mainly focused on efficiency of wearable technology [26], even though data protection is of incredibly high importance in wearable technology as security breaches can lead to dire consequences [51]. Not only do wearables transfer high sensitive data of people, which can be stolen and disclosed by passive attacks such as eavesdropping, but also an attacker can pose a severe threat to a patient by, for example, jamming the wearables’ radio frequencies [26]. It was discovered that one paper in the literature review which addresses this issue and suggests a framework that helps to detect anomalies in the data of the wearable to defend against wireless attacks on radio-frequency which resulted in a feasible and effective protection against those, as long as the attacker does not adapt to them [26]. Additionally, the sensors are often located at public places and an attacker can simply physically destroy them [48]. Another disadvantage of wearables so far is that only a fraction of the wireless sensors’ memory can be used for traditional security mechanisms like encryption, as the sensors are very small and therefore limited in power supply [48]. It is thus difficult to find a way to encrypt the data from sensors as the overall efficiency of the wearable device should not be negatively influenced by security measures [48]. This issue has been addressed by one paper, which tried different encryption techniques and found out that the blowfish algorithm is the most performant encryption code for wearables in combination with CP-ABE [49]. However, encryption itself already creates a problem, as data must be accessible in case of emergencies that decide about the life and death of a person. So far, no solution has been found. Researchers still try to find an efficient key management system. In summary, researchers have so far not been able to identify a solution that can guarantee the safety of a wearable device, as research so far can only exclude specific security issues without providing a framework that guarantees safety for all possible events. Therefore, security was identified as the main constraining factor for wearables in big data.

4 Conclusion

This paper analyzed more than 40 papers about the current state of research in wearable technology in order to identify enabling and constraining factors for wearables in healthcare. Four core areas in which healthcare applications must excel in order

to be market ready have been identified. Especially the quality of life plays a key enabling factor for wearables in healthcare.

With the help of wearables, the overall lifestyle of a person can become healthier thanks to the tracking of physical activity and calories. Considering that obesity has a huge influence on chronic illnesses like diabetes or kidney disease [52] and also heart attacks, the quality of life can be improved for people affected by those issues by constantly tracking and reminding them of eating healthy and engage in physical activity. However, in combination with new treatment possibilities also the quality of life of people can be improved as they can spend more time at home instead of the hospital. Thanks to the possibility to track the health condition in real time, ill people can spend more time at home instead of a hospital as they can be monitored remotely [53], as certain events like a heart attack can be predicted by analysis of the patient's current health condition. In addition, the response of medical teams can be triggered by those devices in case of emergencies and the general quality of treatment is improved as doctors can access long-time data of a patient for their diagnosis. Therefore, the improved treatment is an enabling factor for wearables in healthcare.

This is supported by the increasingly better results wearables in healthcare deliver. While a few years ago it was still a theory that wearables can support patients with heart conditions, the products have become better in supporting patients with their conditions. Falls can be predicted with almost 100% accuracy [34] and the amount of patients dying from heart attacks can be reduced significantly [54, 55]. Additionally, a lot of research has gone into smart clothing [56], which is also becoming more and more effective so that people are no longer aware that they are even connected to a monitoring system. Therefore, one constraining factor, which was the need to apply wearables on more than one place of the body, which influences also the accuracy of the sensor [36], is currently being eliminated by researchers and enables wearables to become more accepted by users.

Unfortunately, there is still one issue that keeps wearables in healthcare from being successful. While it would be impossible to achieve all those results with the huge amount of data sensors send for monitoring, the security issues big data brings along prevents wearables in healthcare from being applied for everyday use [19]. The fact that a relatively simple attack on such a device could lead to a person's death is one of the biggest reasons why it is not possible to use such devices. So far, no solution for the security aspect has been found. Even more research still focuses a lot on product design. Only a handful of papers were trying to find a solution for the security of wearables. Another aspect, which should be considered for research, is what doctors would need to apply wearables for their diagnosis. Research mainly focuses on patients but mostly neglects what doctors require in order to consider wearables helpful.

References

1. NZZ, 2018 dürfte jede Person erstmals über 10 000 Franken für ihre Gesundheit ausgeben, 6/13/2017, <https://www.nzz.ch/schweiz/gesundheitskosten-in-der-schweiz-2018-duerfte-jede-person-erstmals-ueber-10-000-franken-fuer-ihre-gesundheit-ausgeben-ld.1300609>
2. T. Yilmaz, R. Foster, Y. Hao, Detecting vital signs with wearable wireless sensors. *Sensors* **10**(12), 10837–10862 (2010)
3. Y. Gao, H. Li, Y. Luo, An empirical study of wearable technology acceptance in healthcare. *Ind. Manag. & Data Syst.* **115**(9), 1704–1723 (2015)
4. E.L. Mahoney, D.F. Mahoney, Acceptance of wearable technology by people with Alzheimer's disease: issues and accommodations. *Am. J. Alzheimer's Dis. Other Dement.* **25**(6), 527–531 (2010)
5. J. Clover, Study confirms apple watch can detect abnormal heart rhythm with 97% Accuracy. <https://www.macrumors.com/2018/03/21/apple-watch-abnormal-heart-rhythm/>
6. G.H. Tison, J.M. Sanchez, B. Ballinger et al., Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol.* **3**(5), 409 (2018)
7. K.E. Britton, J.D. Britton-Colonnese, Privacy and security issues surrounding the protection of data generated by continuous glucose monitors. *J. Diabetes Sci. Technol.* **11**(2), 216–219 (2017)
8. M.M. Baig, H. GholamHosseini, A.A. Moqem et al., A systematic review of wearable patient monitoring systems—current challenges and opportunities for clinical adoption. *J. Med. Syst.* **41**(7), 115 (2017)
9. S. Patel, H. Park, P. Bonato et al., A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **9**, 21 (2012)
10. M. Schukat, D. McCaldin, K. Wang et al., Unintended consequences of wearable sensor use in healthcare. contribution of the IMIA wearable sensors in healthcare WG. *Yearb. Med. Inform.* **1**, 73–86 (2016)
11. K.R. Evenson, M.M. Goto, R.D. Furberg, Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int. J. Behav. Nutr. Phys. Act.* **12**(1), e192 (2015)
12. L. Lee, S. Egelman, J.H. Lee, et al., Risk perceptions for wearable devices. 4/22/2015, <http://arxiv.org/pdf/1504.05694v1>
13. K.C. Preusse, T.L. Mitzner, C.B. Fausset et al., Older adults' acceptance of activity trackers. *J. Appl. Gerontol.* **36**(2), 127–155 (2017)
14. H.S. Koo, K. Fallon, Preferences in tracking dimensions for wearable technology. *Int. J. Cloth. Sci. Technol.* **29**(2), 180–199 (2017)
15. J. Lee, D. Kim, H.-Y. Ryoo et al., Sustainable wearables: wearable technology for enhancing the quality of human life. *Sustainability* **8**(5), 466 (2016)
16. S.J. Strath, T.W. Rowley, Wearables for promoting physical activity. *Clin. Chem.* **64**(1), 53–63 (2018)
17. E. Park, K.J. Kim, S.J. Kwon, Understanding the emergence of wearable devices as next-generation tools for health communication. *Inf. Technol. People* **29**(4), 717–732 (2016)
18. S.H. Koo, K. Fallon, Explorations of wearable technology for tracking self and others. *Fashion and Textiles* **5**(1), 141 (2018)
19. M.M. Baig, H. Gholam Hosseini, M.J. Connolly, A comprehensive survey of wearable and wireless ECG monitoring systems for older adults. *Med. & Biol. Eng. & Comput* **51**(5), 485–495 (2013)
20. U. Varshney, Pervasive healthcare and wireless health monitoring. *Mob. Netw. Appl.* **12**(2–3), 113–127 (2007)
21. M. Blount, V.M. Batra, A.N. Capella et al., Remote health-care monitoring using personal care connect. *IBM Syst. J.* **46**(1), 95–113 (2007)
22. P. Kostkova, H. Brewer, S. de Lusignan et al., Who owns the data? open data for healthcare. *Front. Public Health* **4**, 7 (2016)
23. J. Sun, Y. Guo, X. Wang et al., mHealth for aging China: opportunities and challenges. *Aging Dis.* **7**(1), 53–67 (2016)

24. J.J.P.C. Rodrigues, D.B. de Rezende Segundo, H.A. Junqueira, et al., Enabling technologies for the internet of health things. *IEEE Access*, **6**, 13129–13141 (2018)
25. H.S. Koo, D. Michaelson, K. Teel et al., Design preferences on wearable e-nose systems for diabetes. *Int. J. Cloth. Sci. Technol.* **28**(2), 216–232 (2016)
26. M. Zhang, A. Raghunathan, N.K. Jha, MedMon: securing medical devices through wireless monitoring and anomaly detection. *IEEE Trans. Biomed. Circuits Syst.* **7**(6), 871–881 (2013)
27. Y. Liang, X. Zhou, Z. Yu et al., Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare. *Mob. Netw. Appl.* **19**(3), 303–317 (2014)
28. A.A. Uddin, P.P. Morita, K. Tallevi et al., Development of a wearable cardiac monitoring system for behavioral neurocardiac training: a usability study. *JMIR mHealth and uHealth* **4**(2), e45 (2016)
29. R.L. Shinmoto Torres, R. Visvanathan, D. Abbott, et al., A battery-less and wireless wearable sensor system for identifying bed and chair exits in a pilot trial in hospitalized older people. *PLoS ONE*, **12**(10) (2017)
30. N.C. Campling, D.G. Pitts, P.V. Knight et al., A qualitative analysis of the effectiveness of telehealthcare devices (i) are they meeting the needs of end-users? *BMC Health Serv. Res.* **17**(1), 455 (2017)
31. J. Cancela, M. Pastorino, A. Tzallas et al., Wearability assessment of a wearable system for parkinson's disease remote monitoring based on a body area network of sensors. *Sensors* **14**(9), 17235–17255 (2014)
32. O. Aziz, S.N. Robinovitch, An analysis of the accuracy of wearable sensors for classifying the causes of falls in humans. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**(6), 670–676 (2011)
33. M.A. Case, H.A. Burwick, K.G. Volpp et al., Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA* **313**(6), 625 (2015)
34. O. Aziz, E.J. Park, G. Mori, et al., Distinguishing near-falls from daily activities with wearable accelerometers and gyroscopes using support vector machines, in *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2012 (2012) pp. 5837–5840
35. A.L. Silva de Lima, L.J.W. Evers, T. Hahn et al., Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. *J. Neurol.* **264**(8), 1642–1654 (2017)
36. C. Jayaraman, C.K. Mummidisetty, A. Mannix-Slobig et al., Variables influencing wearable sensor outcome estimates in individuals with stroke and incomplete spinal cord injury: a pilot investigation validating two research grade sensors. *J. NeuroEngineering Rehabil.* **15**(1), 788 (2018)
37. C. Klersy, A. de Silvestri, G. Gabutti et al., A meta-analysis of remote monitoring of heart failure patients. *J. Am. Coll. Cardiol.* **54**(18), 1683–1694 (2009)
38. J.H.M. Bergmann, A.H. McGregor, Body-worn sensor design: what do patients and clinicians want? *Ann. Biomed. Eng.* **39**(9), 2299–2312 (2011)
39. N.D. Weinstein, Testing four competing theories of health-protective behavior, *Health psychology: official journal of the division of health psychology. Am. Psychol. Assoc.* **12**(4), 324–333 (1993)
40. F. Axisa, P.M. Schmitt, C. Gehin, et al., Flexible technologies and smart clothing for citizen medicine, home healthcare, and disease prevention, in *IEEE Transactions on Information Technology in Biomedicine: A Publication Of The IEEE Engineering in Medicine and Biology Society*, vol. 9, no. 3. (2005), pp. 325–336
41. M. Suh, K.E. Carroll, E. Grant et al., Investigation into the feasibility of inductively coupled antenna for use in smart clothing. *Int. J. Cloth. Sci. Technol.* **26**(1), 25–37 (2014)
42. L. Clifton, D.A. Clifton, M.A.F. Pimentel et al., Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. *IEEE J. Biomed. Health Inform.* **18**(3), 722–730 (2014)
43. A. Lymberis, Wearable smart systems: from technologies to integrated systems, in *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2011 (2011), pp. 3503–3506

44. M.-H. Maras, Internet of things: security and privacy implications. *Int. Data Priv. Law* **5**(2), 99–104 (2015)
45. H. Li, J. Wu, Y., Examining individuals' adoption of healthcare wearable devices: an empirical study from privacy calculus perspective. *Int. J. Med. Inform.* **88**, 8–17 (2016)
46. M. Milutinovic, B. de Decker, Ethical aspects in eHealth—design of a privacy-friendly system. *J. Inf., Commun. Ethics Soc.* **14**(1), 49–69 (2016)
47. J. Couturier, D. Sola, G. Scarso Borioli, et al. How can the internet of things help to overcome current healthcare challenges (2012)
48. H.S. Ng, M.L. Sim, C.M. Tan, Security issues of wireless sensor networks in healthcare applications. *BT Technol. J.* **24**(2), 138–144 (2006)
49. D. Sathya, P. Ganesh Kumar, Secured remote health monitoring system. *Healthc. Technol. Lett.* **4**(6), 228–232 (2017)
50. M.A.D. Brodie, M.J.M. Coppens, S.R. Lord et al., Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Med. Biol. Eng. Compu.* **54**(4), 663–674 (2016)
51. M. Shin, Secure remote health monitoring with unreliable mobile devices. *J. Biomed. & Biotechnol.* **2012**, 546021 (2012)
52. F.P. Wieringa, N.J.H. Broers, J.P. Kooman et al., Wearable sensors: can they benefit patients with chronic kidney disease? *Expert Rev. Med. Devices* **14**(7), 505–519 (2017)
53. K.-Y. Lam, N.W.-H. Tsang, S. Han et al., Activity tracking and monitoring of patients with alzheimer's disease. *Multimed. Tools Appl.* **76**(1), 489–521 (2017)
54. M. Ehn, L.C. Eriksson, N. Åkerberg et al., Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. *JMIR mHealth and uHealth* **6**(2), e34 (2018)
55. W.K. Lim, S. Davila, J.X. Teo et al., Beyond fitness tracking: The use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. *PLoS Biol.* **16**(2), e2004285 (2018)
56. M. Chen, Y. Ma, J. Song et al., Smart clothing: connecting human with clouds and big data for sustainable health monitoring. *Mob. Netw. Appl.* **21**(5), 825–845 (2016)

Proposed System for Effective Adoption of E-government to Obtain Construction Permit in Egypt



Heba Fawzy and Dalia A. Magdi

Abstract This study aims to develop a new automated system for obtaining construction permit in Egypt instead of the old processes which depend on paper work and consumes more than 200 days through 19 procedures affecting the ranking of Egypt in Doing Business Report in dealing with construction permit pillar. Using this new system will have a positive impact on enhancing the rank of Egypt in this international report by reducing time and the number of procedures needed to obtain the permit in addition to enhancing the process efficiency and linking all the stakeholders involved in it with minimum human interaction which eliminate corruption level as well. The system also helps in obtaining regular reports and feedback from the employees and the citizens. This study analyzes old system of obtaining construction permit in Egypt including time, number of procedures, and their impact on the Egyptian ranking of Doing Business Report. It also describes the proposed system design and the impact of using the system on enhancing Egypt rank in dealing with construction permit index in Doing Business Report.

Keywords E-government · Information system · Doing Business Report · Construction permit

1 Introduction

The emerging purpose of information technology and communication networks is to expend and support the development of the economic, social, cultural, and political sectors that took the attention of most of developing countries [1]. The number of

H. Fawzy
Ministry of Housing, Utilities and Urban Communities, Cairo, Egypt
e-mail: enheba2016@yahoo.com

D. A. Magdi (✉)
Information System Department, French University in Egypt, Cairo, Egypt
e-mail: daliomagdi@gmail.com

Computer and Information System Department, Sadat Academy for Management Sciences, Cairo, Egypt

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_2

governments who take advantage of new information and communication technologies is rapidly increasing in the last few years. The emerging use of information and communication technology for the transformation in creating new perceptions about government and governance is a developing approach toward public sector [2].

E-government gains the interest and the governmental focus in different countries around the world. A lot of governments introduced and enforced e-government systems in order to reduce cost, improve the services, save time as well as to increase the effectiveness and the efficiency in different public sectors. The e-government associated the Internet to create fundamental amendments within the entire structure of the society that may include provided values, culture, and the ways of driving business using the information technology as a tool within the quotidian work. The e-government purpose extends the idea of transforming the traditional information into a digital form accessible through different Websites or automating traditional processes to an electronic platform, it additionally aims to rethink about the ways in which the government operates in order to enhance processes and integration [3].

2 Problem Definition

According the current status analysis of obtaining construction permit in Egypt, the process consumes high cost, long times to obtain the permit, lack transparency, and flexibility. It focuses on processes instead of results as well. This old procedures affect the Egyptian economic status that is why Egypt lags behind other countries in Doing Business Report and Global Competitiveness Report.

Thus, the problem statement will be as follow: “How to develop an effective and efficient e-government system to obtain a construction permit in Egypt?”

3 Objective

The study aims to present an e-government enhanced system that automates the system of obtaining construction permit in Egypt in order to improve productivity, performance, quality of service, and reduce time, cost and number of process required to obtain the permit, these improvements may lead to improve the Egyptian ranking of Doing Business Report to be within top 30 countries by 2030 [4].

So the proposed system’s objectives are as follow:

1. To improve efficiency and effectiveness so it could help the municipality authority to provide good service quality in order to save time and money.
2. Assembling all the approvals for the construction license in one place that involve all the entities in order to provide the service while saving time and money needed to obtain the permit.

3. Enhance coordination between all entities responsible for issuing permit (Municipality—Syndicate of Licensed Engineers—Electricity Authority—Civil Defense and Firefighting Authority—Inspection Municipal Authority—Water and Sewerage Company—Real Estate Registry) by providing a network based on ICT technology.
4. Reduce the number of procedures needed for issuing the permit in accordance with the construction law and ICT technology.

4 Literature Review

4.1 E-government

The transformation toward e-government raised in the late 1990s along with the beginning of the Internet age and the emergence of e-commerce. The UN definition thinks that the e-government includes the use of the public sector virtually by all means of applications and platforms of information technology and communication networks. Kitaw also defined e-government in 2006 as the use of new information technology to raise governmental efficiency, effectiveness, and to facilitate access to governmental services, and gain greater public access [1, 5].

Governments can be distinguished according to their roles into multiple definitions as follows:

- “E-government refers to facilitating governmental services provided to citizens by means of information technology and communication networks, especially over the Internet.”
- Digital government refers to all the services provided by the public sector by the use of different means of new information technologies and communication networks [6].
- E-governance refers to the use information technology and communication networks by the organizations that offer political activity within different countries.

Developing countries showed high potential of e-government that assist individuals to develop their full potential and productivity according to their interests and different needs [2]. As for Egypt, there are some studies associated with the initiatives of the Egyptian data society, the development and evaluation of information technology sector in Egypt, the increased use of e-commerce taking into consideration the legal, technological, social, and financial issues.

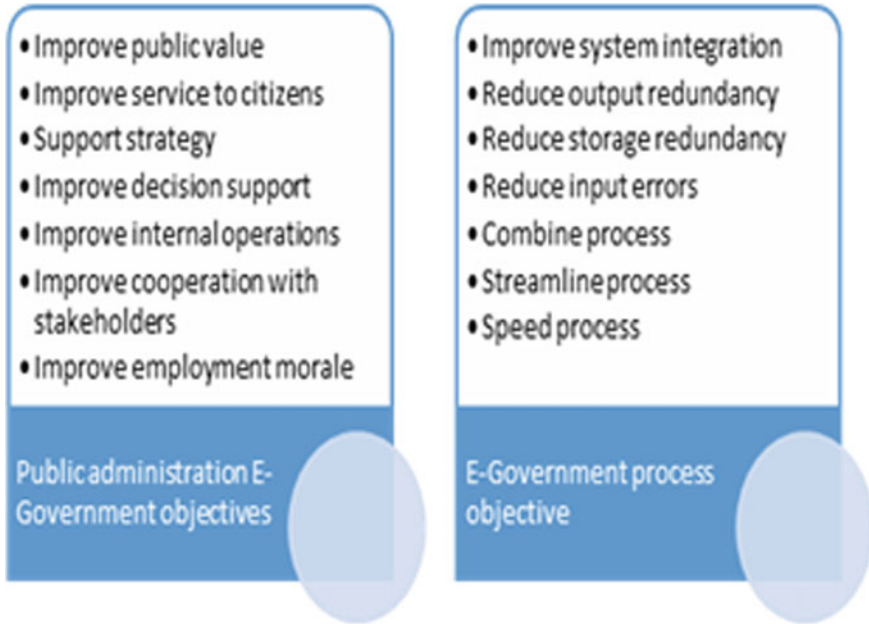


Fig. 1 E-government objectives

4.2 Goals of E-government

E-government goals vary significantly worldwide from one government to another. Usually, the e-government goals are locally determined according to the political leadership of every government. However, these goals may be influenced by the key institutional stakeholders among these countries [7].

As declared recently by the Ministry of Communication and Information Technology in collaboration with Ministry of Planning, Follow-up, and Administrative Reform who emphasized on the need to make every effort in order to provide businesses and citizens with digital public services that is open, efficient and inclusive, providing borderless, interoperable, personalized, user-friendly, and end-to end-at all levels of public administration (see Fig. 1).

4.3 E-government Adoption in Egypt

Egyptian government recently has recognized the importance of e-government at different levels since 2004 when it launched Egypt's e-government portal. It has been acknowledged that e-government plays an important role in providing effective and

acceptable services to Egyptian citizens, promoting the economy, and enhancing communication and information exchange among different governmental sectors.

The Egyptian Information and Decision Support Center (IDSC) was founded in 1985 in order to develop the industry of information technology in Egypt as well as the infrastructure required for governmental decision support. The main objective of this center is to help in public access to different types of information, in addition to enforcing the facilitation of business and investment processes [8]. For the past 33 years, the center succeeded to execute a lot of information technology projects in terms of governmental improvement, alteration of public sector, improvement of human resources, and facilitating the use of the commercial Internet, effective resource management, conservation of traditional culture, city planning, as well as the development of different projects in all governmental sectors and levels. Now, the IDS center focuses on decision support for the Cabinet [9]. The Ministry of Communications and Information Technology (MoCIT) was developed in 1999 in order to accelerate the creation of a society of information in addition to improvement of the information infrastructure [10]. After a short time of its formation, the Ministry demonstrated the Egyptian National Communications and Information Technology Plan (NCITP) [8]. The NCITP has paved the road for launching the Egyptian Information Society Initiative (EISI) that was designed for approximately seven important mechanisms, in order to facilitate the evolution of Egypt into the new society of information [11].

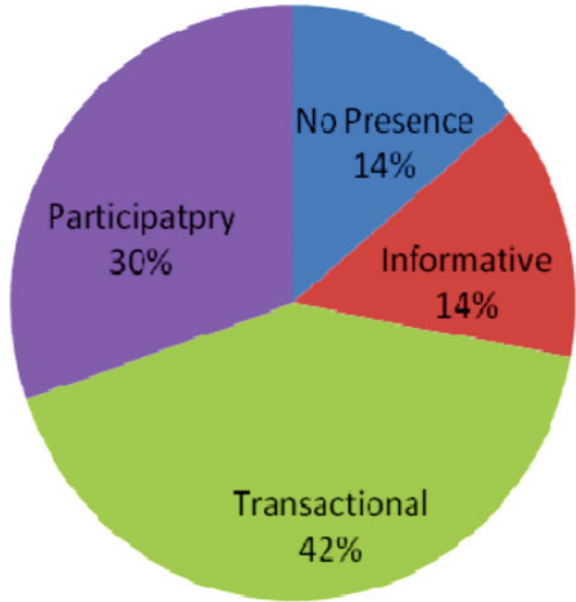
Unfortunately, there is not a lot of studies that discussed the general issues of e-government, such as different frameworks of e-government and the strategies needed to accelerate the interoperability in e-government [12–14]. Furthermore, the e-government implementation faced a lot of challenges due to the lake of evaluation of governmental readiness. This is where our contribution fits. Our study aims to evaluate the Egyptian readiness to apply e-government services and strategies, as well as the challenges that face its implementation when delivering e-government services. Hence, it comes to fill a gap in literature concerning e-government and implementation challenges [8].

4.4 E-government Program in Egypt

In 2004, the Egyptian government started different strategies and implementation plans for the development of governmental services, thus it established the following programs [16, 17]:

1. Institutional Development Program which contains different plans, policies, and regulations to improve the environmental and human resources work.
2. Governmental Services Development Program which provided citizens, private sectors, and governmental entities with services in an effective and efficient form.

Fig. 2 Egypt e-government portfolio summary



3. Enterprise Resource Planning Program which focuses on the improvement of the governmental work flows processes and the automation of the governmental procedures by means of information and communication technologies.
4. Establishing and Integrating National Databases Program which aimed to create an integrated national database for an efficient and safe exchange of information between governmental entities [8].

4.5 Egypt's E-government Portfolio Summary

Figure 2 shows e-government portfolio summary which is based on calcifying services provided by the government [8].

4.6 Egyptian Service Development Program (GSDP)

As declared by the minister's advisor for international relations of the Ministry of State for Administrative Development in Egypt, the GSDP program vision aims to provide effective governmental services framework that satisfy citizens/providers, with the declaration of transparency and integrity in order to facilitate the delivery of services provided by the government by different means of simplification of access

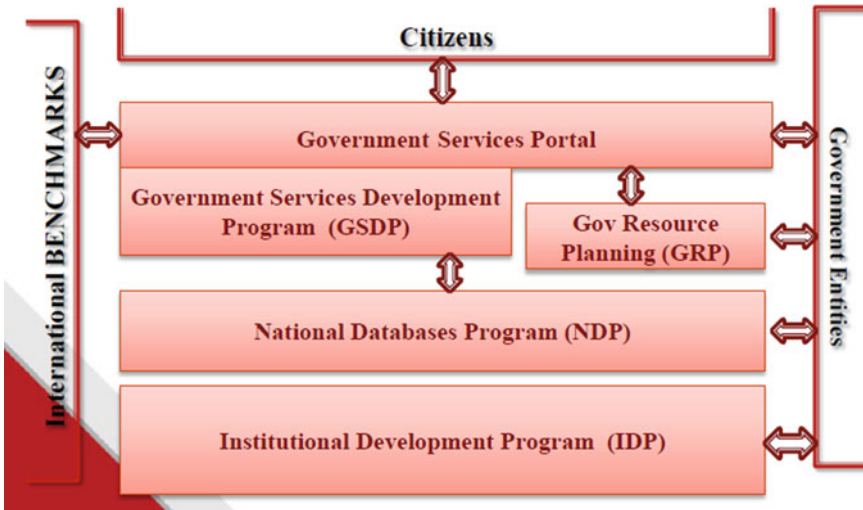


Fig. 3 Ministry of state for administrative development (MSAD)

and process behind it, and by using the advantage of the success of different MSAD programs to enhance the efficiency and effectiveness of the government units in order to achieve citizen satisfaction along with a transparent environment (see Fig. 3).

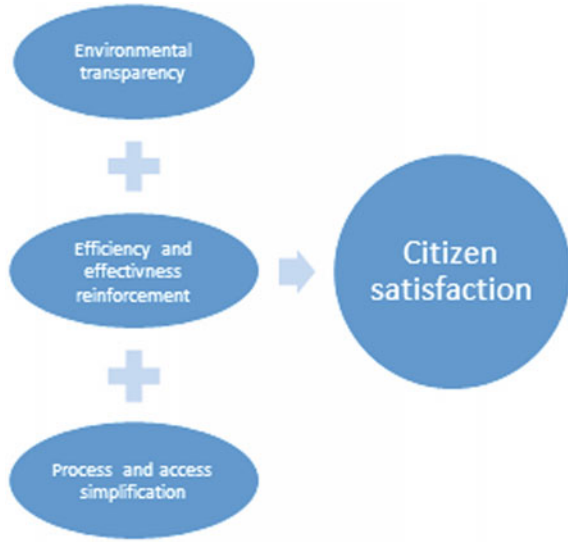
The adoption of GSDP in Egypt led to:

1. Re-engineering of process and documentation.
2. Full tracking of process by means of automated workflow system.
3. Applying the strategy of separation between service provider and acquirer through a one-window stop.
4. Enhanced environment of work.
5. Issuing 12 Governorates portals to provide service by 88 municipalities.
6. In October 2008, it won “All Africa Public Service Innovation Award (AAPSIA)” (Fig. 4).

5 Current Status Analysis as per the Global Competitiveness Report (GCR) and Doing Business Report

As known, global economy is currently based on competitiveness both in products and services. There are many international reports that examine and analyze the ranking of countries in these areas, especially the field of services provided to individuals, whether governmental or nongovernmental, such as the “Global Competitiveness Report” and “Doing Business Report.”

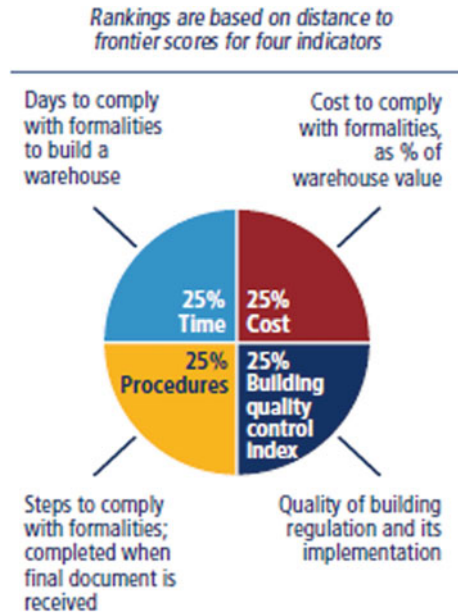
Fig. 4 Elements of e-government customer satisfaction



- The World Economic Forum publishes yearly the Global Competitiveness Report (GCR). Different countries are ranked, since 2004, and the Global Competitiveness Report is created on the basis of the Global Competitiveness Index. The report evaluates countries’ ability to provide citizens with high levels of prosperity. This report is affected by the productivity of a country when using available resources. That is why it can be seen that the Global Competitiveness Index measures the set of institutions, policies, and factors that precis the maintainable current and medium-term levels of economic prosperity [18].
- All required official procedures, or usually done in practice are recorded in the Doing Business Report for start-up entrepreneurs and for those who formally operate in industrial business or commercial environment, in addition to the calculation of time and cost consumed in order to complete all procedures and the paid-in minimum capital requirement [19]. All licenses and permits procedure must be included when obtaining and completing any company inscriptions, notifications required, verifications for the company, and employees with relevant authorities [17].

Dealing with construction permit is one of its nine pillars which we are going to focus on in this study. It is responsible of procedures tracking, calculates the cost consumed, and time spent to get permits and licenses of warehouse building, notifications submission, the request and collection of inspections, and the procuration of all necessary utility connections. More and above, handling the indicator of construction permits that are supposed to measure the quality control evidence of building process, the measure of regulation quality in building process, the quality control empowerment as well as safety mechanisms, legal responsibilities and insurance systems, in addition to all required certification (see Fig. 5).

Fig. 5 Dealing with construction permit indicators, Doing Business Report 2018



6 Evaluation of Current Status of Obtaining Construction Permit in Egypt

In 2018, Egypt was ranked 66 among 142 countries in the world according to Doing Business Report in dealing with construction permits pillar due to long time and expensive procedures compared to other countries which have simpler and less expensive procedures [15].

Egypt has established a new building law in order to reduce the time to obtain a building permit, thanks to that code Since 2009–2018 Egypt’s rank jumped from 165 to 66 which is a great positive change but Egypt still has a lot to do in order to improve that rank especially in reducing time required to get the permit and number of procedures.

Obtaining construction permit in Egypt consumes too much money and time as it still uses a lot of paper work to obtain construction permit. On the other hand, the processes take a lot of steps related to many entities governmental and non-governmental (Municipality Authority, Syndicate of Licensed Engineers, Electricity Company, Water and Sewerage Company, Contractors union).

Obtaining construction permit is the responsibility of the Municipality Authority represented in Municipality offices in all cities and villages all over Egypt under the supervision and monitoring of Housing Ministry as it is the responsible of technical and legal inspection of all procedures according to the Egyptian Construction law No119/2009 (see Table 1).

Table 1 Dealing with construction permit in Egypt, Doing Business Report 2018

Standardized company				
Estimated value of warehouse	EGP 1,411,914.00			
City covered	Cairo			
Indicator	Egypt, Arab Rep.	Middle East & North Africa	OECD high income	Overall best performer
Procedures (number)	19	16.2	12.5	7.00 (Denmark)
Time (days)	172	132.1	154.6	27.5 (Korea, Rep.)
Cost (% of warehouse value)	1.9	4.3	1.6	0.10 (5 Economies)
Building quality control index (0–15)	14	11.8	11.4	15.00 (3 Economies)

This can be explained in detail in the following table (Doing Business Report 2018).

These processes consume 173 days and EGP 25950 which equal 1442 \$. As shown in Fig. 6, anyone wants to obtain a construction permit has to pass through all these entities and a long journey consuming time and effort to go to each and every entity to obtain or submit a single certificate or document with a lot of paper work which could include corruption due to the human interface between him and the employees.

These procedures can be illustrated in the following blueprint that identifies customer actions, back-stage, actions, support processes, and physical line of visibility that separates front-stage and back-stage actions in addition to lines of interaction

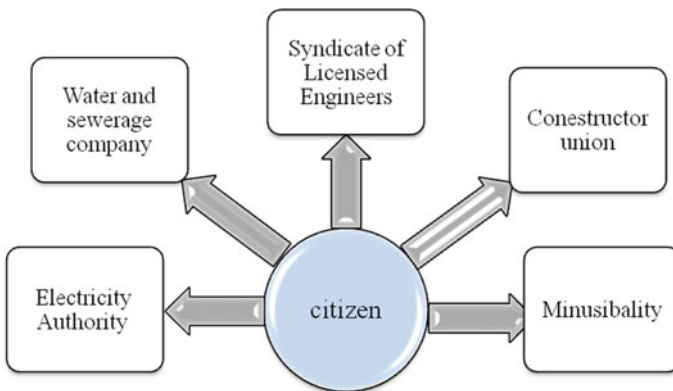


Fig. 6 Dealing with construction permit in Egypt stakeholders relations

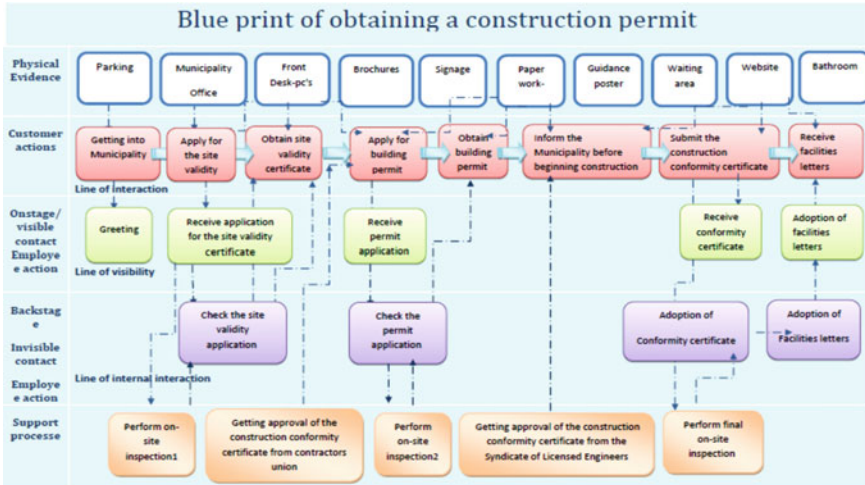


Fig. 7 Dealing with construction permit in Egypt blueprint, Doing Business Report 2018

which separates customer actions from service provider actions is designed as shown in Fig. 7 (Table 2).

As concluded from previous data, Egypt has a lot to do in order to improve its international rank especially in reducing time required to get the permit and number of procedures by improving the efficiency of obtaining the construction permit process to reduce time needed to obtain this service and providing high quality of service in the same time.

7 Proposed System Description

7.1 System Concept

This system is designed in order to automate the old process of obtaining construction permit process in Egypt which depends on managerial and routine paper work. The new model will change this aged philosophy into one-stop single window based on an umbrella organization which collect information and processes the information. As legislation and policy have a strong impact on the sharing of information and knowledge between organizations, there is a need of regulatory framework which consists of the scale, content, and standards of electronic information sharing between government organizations depending on formal policies and regulations.

So, instead of spending more than 170 days passing through long procedure and many entities (Municipality—Syndicate of Licensed Engineers—Electricity Authority—water and sewerage company- contractor union) in order to obtain the permit,

Table 2 Dealing with construction permit in Egypt steps, Doing Business Report 2018

1	Apply for the site validity certificate	1 day	EGP 200
2	Receive on-site inspection from the municipality	1 day	No charge
3	Obtain site validity certificate from the municipality	15 days	No charge
4	Obtain a geotechnical study/soil test (private sector)	9 days	EGP 4500
5	Request and obtain building permit from the municipality	30 days	EGP 2638
6	Hire an external engineer to supervise the construction site	1 day	EGP 2000
7	Obtain approval of the execution supervision certificate from the syndicate of licensed engineers	1 day	EGP 312
8	Inform the municipality before beginning construction	1 day	No charge
9	Receive set-back inspection from the municipality II	1 day	No charge
10	Receive set-back inspection from the municipality III	1 day	No charge
11	Receive set-back inspection from the municipality	1 day	No charge
12	Obtain approval of the construction conformity certificate from the syndicate of licensed engineers	1 day	EGP 300
13	Receive on-site inspection from the civil defense and firefighting authority	15 days	No charge
14	Submit the construction conformity certificate and receive final inspection from the municipal authority	15 days	No charge
15	Register the building with the real estate registry	60 days	EGP 2000
16	Request and obtain sewerage connection	30 days	EGP 5,000
17	Request water connection	1 day	No charge
18	Receive site inspection by Utilities to assess work and cost	7 days	No charge
19	Obtain water connection	22 days	10,000

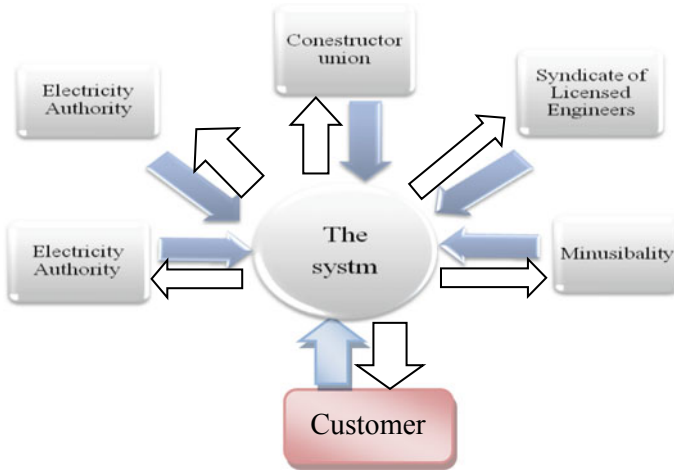


Fig. 8 Proposed system stakeholders relations

the service will be delivered automatically in one place (single window) or via the Internet (see Fig. 8).

In order to create this system, all the documents and procedures required for the issuance of building permits have been listed in accordance with the Consolidated Building Law No. 119/2008 promulgated by the Egyptian government. The concept of the new system is based on one-stop window concept as all the procedures will be done within it and the customer will not need to do any physical action except for request and obtain the permit through the system with no interface with all entities involved in the process which will be the system job to make the document circulation until the final stage which include obtaining the permit.

7.2 Proposed System Design

The proposed system will save all time and effort consumed in submitting and obtaining documents from each and every entity in the old system as it will be done automatically within the system. The new system will lead to minimum human interaction as employees which insure the maximum level of accuracy and efficiency in addition to the minimum level of corruption. Figure 9 shows the context diagram that presents all the users of the system and the documents flow through the system as well as system processes.

In order to implement the proposed model in all municipality offices all over Egypt, a five years plan has been developed by replacing paper work with this technology. By this improvement time, many steps could be eliminated as shown in Table 3.

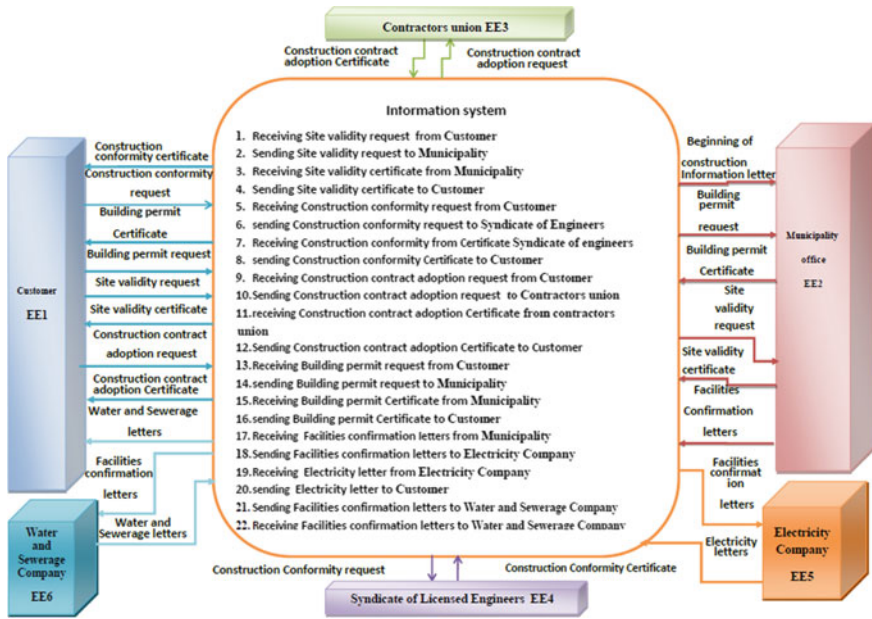


Fig. 9 Proposed system context diagram

The above table (Table 3) shows the reduced time by eliminating process that involve obtaining and submitting documents as it will be automatically done by the system.

Time consumed by other process will be reduced by half as it will be done easier with the help of the system. By this elimination and reduction, the whole process will consume only 142 days, with only 12 procedures. The system also helps in enhancing building quality as it provides transparency during all process including inspection process which will be done by the municipality employees and submitting their report into the system with ability to be checked by the higher control and evaluation entity which may prevent any corruption.

It also will have apposite effect on the cost as it saves some of cost due the long period consumed to obtain all paper work and documents needed, while there are some fees can not be reduced as it is approved in accordance with the law or it involves private sector which fees could not be controlled by government. These eliminations and reductions in time and cost should have a positive effect on Egypt rank Doing Business Report Egypt rank.

Table 3 Time and the number of procedures in case of implementing the proposed system

1	Apply for the site validity certificate	1 day	EGP 200
2	Receive on-site inspection from the municipality	1 day	No charge
3	Obtain site validity certificate from the municipality	7 days reduced by half	No charge
4	Obtain a geotechnical study/soil test (private sector)	9 days	EGP 4500
5	Request and obtain building permit from the municipality	15 days reduced by half	EGP 2638
6	Hire an external engineer to supervise the construction site	1 day	EGP 2000
7	Obtain approval of the execution supervision certificate from the syndicate of licensed engineers	1 day	EGP 312
8	Inform the municipality before beginning construction	1 day	No charge
9	Receive set-back inspection from the municipality	1 day	No charge
10	Receive set-back inspection from the municipality II	1 day	No charge
11	Receive set-back inspection from the municipality III	1 day	No charge
12	Obtain approval of the construction conformity certificate from the syndicate of licensed engineers	1 day	EGP 300
13	Receive on-site inspection from the civil defense and firefighting authority	15 days	No charge
14	Submit the inspection conformity certificate and receive final inspection from the municipal authority	7 days reduced by half	No charge
15	Register the building with the real estate registry	60 days	EGP 2000
16	Request and obtain sewerage connection	15 days reduced by half	EGP 5000
17	Request water connection	1 day	No charge
18	Receive site inspection by utilities to assess work and cost	7 days	No charge
19	Obtain water connection	11 days reduced by half	10,000

8 Results

As shown in this study, adopting this system, the number of procedures will be 12 with 142 days required to obtain the permit that will enhance Egypt rank to be compared with Kyrgyz Republic which occupied the 31st position in Doing Business Report 2018 with eleven procedures and 142 days (as shown in Table 4).

The following table (Table 5) shows the impact of using the proposed system on Egypt's rank in Doing Business Report in a form of comparison between before and after situation as it will be enhanced to the 30th position with 12 procedures and 142 days which accomplish the goals of sustainable development strategy Egypt 2030.

Table 4 Dealing with construction permit—Kyrgyz Republic, doing business 2018

Standardized company				
Estimated value of warehouse	KGS 3,601,278.90			
City covered	Bishkek			
Indicator	Kyrgyz Republic	Europe & Central Asia	OECD high income	Overall best performer
Procedures (number)	11	16.0	12.5	7.00 (Denmark)
Time (days)	142	168.3	154.6	27,5 (Korea, Rep.)
Cost (% of warehouse value)	1.7	4.0	1.6	0.10 (5 Economies)
Building quality control index (0–15)	11.0	11.4	11.4	15.00 (3 Economies)

Table 5 Comparison between Egypt ranking before and after adopting of the proposed system

EGTPY	Before adopting the system	After adopting the system
Rank	66	30
Number procedures	19	173
Time in days	12	142

9 Conclusion

The main objective of this study is developing a new automated model for obtaining construction permit in Egypt instead of the old process which depends on paper work and consumed more than 172 days through 19 procedures affecting the ranking of Egypt in Doing Business Report in Dealing with construction permit pillar. Using this new model will have a positive impact on enhancing the rank of Egypt in this international report by reducing time and number of procedures needed to obtain the permit in addition to enhancing the process efficiency and linking all the enteritis involved in it with minimum human interaction which eliminate corruption level as well. The system also helps in obtaining regular reports and feedback from the employees and the customers.

The study analyzes old system of obtaining construction permit in Egypt including the time, the number of procedures and their impact on Egypt ranking in Doing Business Report.

It also describes the proposed system through planning stage which includes developing vision, mission, objectives, and KPIs in addition to system designing graphs, implementation, and monitoring feedback stages within the frame work of legal and technological principals based on organization, people technology pillars. By adopting this new model, Egypt ranking in doing business report will be enhanced to be one of top thirty country which satisfy sustainable development strategy Egypt vision 2030.

References

1. Y.N. Chen, H.M. Chen, W. Huang, R.K. Ching, E-government strategies in developed and developing countries: An implementation framework and case study. *J. Glob. Inf. Manag. (JGIM)* **14**(1), 23–46 (2006)
2. V. Ndou, E-government for developing countries: opportunities and challenges. *Electron. J. Inf. Syst. Dev. Ctries.* **18**(1), 1–24 (2004)
3. D. Stamoulis, D. Gouscos, P. Georgiadis, D. Martakos, Revisiting public information management for effective e-government services. *Inf. Manag. & Comput. Secur.* **9**(4), 146–153 (2001)
4. M. Backus, E-governance in developing countries (2002)
5. Y. Kitaw, E-government in Africa: Prospects, challenges and practices. International Telecommunication Union (2006)
6. S. Bhatnagar, *E-government: From Vision To Implementation-A Practical Guide with Case Studies*, vol. 21, No. 1 (Sage 2004)
7. S.C. Bhatnagar, The egyptian cabinet information and decision support center (IDSC) (1993)
8. T.R. Gebba, M.R. Zakaria, E-government in Egypt: an analysis of practices and challenges. *Int. J. Bus. Res. Dev.* **4**(2) (2012)
9. Ministry of State for Administrative development (2010). Available online at: www.egptpt.gov.eg. Accessed 15 May 2012
10. L. Hassanin, Africa ICT policy monitor project: Egypt ICT country report. The Association of Progressive Communications (APC) (2003)
11. Ministry of Communication and Information Technology report 2004

12. Azab, N., Ali, M., G. Dafoulas, Incorporating CRM in e-government: case of Egypt, in *Proceedings of the IADIS International Conference e-commerce*, p. 247 (2006)
13. R. Klischewski, Architectures for tinkering?: contextual strategies towards interoperability in e-government. *J. Theor. Appl. Electron. Commer. Res.* **6**(1), 26–42 (2011)
14. A.M. Riad, H.M. El-Bakry, G.H. El-Adl, E-government frameworks survey. *Int. J. Comput. Sci. Issues (IJCSI)* **8**(3), 319 (2011)
15. A. Am Abdelkader, A manifest of barriers to successful e-government: cases from the Egyptian programme. *Int. J. Bus. Soc. Sci.* **6**(1) (2015)
16. X. Sala-i-Martin, K. Schwab, M.E. Porter (eds.), *The global competitiveness report 2003–2004* (Oxford University Press, USA, 2004)
17. M.M. Abbassy, S. Mesbah, Effective e-government and citizens adoption in Egypt. *Channels* **2**(3), 4 (2016)
18. World Bank, *Doing Business 2017: Equal Opportunity for All*. World Bank Publications (2016)
19. Doing business Homepage, <http://www.doingbusiness.org>. Last accessed 29 Dec 2018
20. A. Rorissa, D. Demissie, T. Pardo, Benchmarking e-government: a comparison of frameworks for computing e-government index and ranking. *Gov. Inf. Q.* **28**(3), 354–362 (2011)
21. B.W. Wirtz, L. Mory, R. Piehler, P. Daiser, E-government: a citizen relationship marketing analysis (IRPN-D-16-00005). *Int. Rev. Public Nonprofit Mark.* **14**(2), 149–178 (2017)
22. M.R. Zakaria, Towards categorizing e-government services: the case of Egypt. *Int. J. Bus. Res. Dev.* **3**(3), 16–28 (2015)
23. M.E.S. Wahed, E.M. El Gohary, The future vision for the design of e-government in Egypt. *Int. J. Comput. Sci. Issues (IJCSI)* **10**(3), 292 (2013)
24. T. Schuppan, E-government in developing countries: experiences from sub-saharan Africa. *Gov. Inf. Q.* **26**(1), 118–127 (2009)

Potential Use of Bitcoin in B2C E-commerce



Ralf-Christian Härting and Christopher Reichstein

Abstract Bitcoin is a new digital currency with a very high visibility in media and research. Therefore, different aspects of potentials of Bitcoin are explored. In a prior investigation, several manifest indicators like transaction velocity have been identified as important influencing factors for the perceived use of digital currency. The focus of this paper is an empirical study, which examines factors of the potential use of Bitcoin in a B2C E-Commerce environment. More than 100 online merchants were interviewed in 2016. Based on a structural equation model (SEM), the results of the analysis show that the low transaction costs and acceptance are the main factors that influence the potential benefits of Bitcoin in B2C E-Commerce. The study also gives ideas for the relevance of further indicators.

Keywords Bitcoin · Digital currency · E-Commerce · Empirical research

1 Introduction

The digital currency Bitcoin has existed since 2009 [1]. Based on new approaches to digital transformation, digital or crypto currencies have become more widespread and valuable [2]. Numerous alternative currencies such as Ethereum, Iota, Ripple, or Stellar have also emerged [3]. Accordingly, a few empirical studies or research papers are available on the subject of Bitcoins. Most studies on this have examined the general knowledge and use of the digital currency of private users.

Among the companies that may want to accept or have already accepted the digital currency, there are very few empirical studies. Therefore, companies engaged in E-Commerce and not private users are examined in this paper. The impact of different criteria on the potential benefits of Bitcoin in B2C E-Commerce should be considered

R.-C. Härting (✉) · C. Reichstein
Business Science, Aalen University of Applied Sciences, Aalen, Germany
e-mail: ralf.haerting@hs-aalen.de

C. Reichstein
e-mail: christopher.reichstein@kmu-aalen.de

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_3

in more detail. These criteria include, among other things, the velocity of transactions [4], the security of transactions [5], and the volatility of the Bitcoin price [6].

2 Determinants of the Digital Currency Bitcoin

In order to carry out an empirical study, an extensive literature review was first required on the topic of Bitcoin. This was sought from international journals, papers, and empirical studies. In addition, the results of a prior research project [7] on benefits of using Bitcoin in European countries were analyzed. Based on a structured literature review, relevant determinants of Bitcoin could be identified. The following five determinants are the starting point for generating hypothesis for further investigation (Table 1).

The determinant ‘safety’ [8–10] is focusing on two effects of Bitcoins: (1) The irreversibility of transactions and therefore the protection against chargeback fraud; (2) The opportunity to pay anonymously, based on Bitcoin addresses.

The determinant ‘transaction velocity’ [11] states that the transactions are processed in almost real time. The transaction is carried out within seconds and takes only about ten minutes to be confirmed [12].

The determinant ‘acceptance’ [6, 13, 14] describes the growing number of users who use Bitcoin. In addition, the number of online stores that accept Bitcoins as payments is growing. This also increases the overall acceptance of Bitcoin.

The next determinant is ‘transaction costs’ [15, 16]. When you pay with Bitcoins, the transaction costs are significantly lower than with other payment options.

The final determinant comprises the heavily jittery nature of Bitcoin prices. Because of specific events, such as government intervention or theft of Bitcoins, there is a great volatility of the Bitcoin course [17, 18].

Table 1 Determinants

Determinant	Short description	References
Safety	Irreversibility of transactions	[8]
	Anonymity	[9]
	Cryptography as a security aspect	[10]
Transaction velocity	Transactions are near real time	[11]
Acceptance	Growing number of users	[13]
	Growing number of merchants which accept Bitcoins	[6]
	Growing acceptance	[14]
Transaction costs	Very low transaction costs	[15]
Volatility	Fluctuating Bitcoin prices	[17]

3 Research Design and Methods

This chapter specifies the research design of the study. It includes generating of hypothesis, research methods, and data collection (Fig. 1).

Hypothesis 1 Safety positively influences the benefit of using Bitcoins.

As soon as a payment is carried out with Bitcoins, it is irreversible. This means that the transaction cannot be undone. Through this, the dealer has a greater measure of safety than otherwise, since deceit cannot take place with reversed entries anymore [19, 20]. Online dealers do not have to bear the costs of deceit through this any longer.

Hypothesis 2 A fast transaction positively influences the benefit of using Bitcoins.

As Bitcoins have no actual physical location, it is possible to transfer them without delay and limits to any place in the world. Thus, it is possible to transfer Bitcoins from A to B in a few minutes. By the peer-to-peer network, a real-time transaction is quite feasible. So, a transaction with Bitcoin is significantly faster than the usual payment methods [8, 21].

Hypothesis 3 A higher acceptance positively influences the benefit of using Bitcoins.

The greater the acceptance of digital currencies, the higher is its rate of success. A greater acceptance of Bitcoin too, therefore, leads to greater success. The number of users rises through this, and more dealers (storekeepers) accept payments made with Bitcoin [13, 22].

Hypothesis 4 Low transaction costs positively influence the benefit of using Bitcoins.

Online dealers try to bait more and more customers by a variety of payment possibilities. More recently, mobile payments (m-payments), electronic wallets (Azimo,

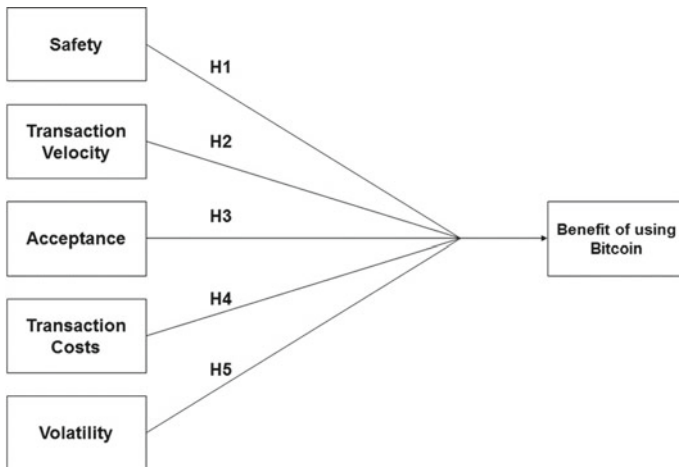


Fig. 1 Research model

PayPal, SolidTrustPay), and crypto currencies have shaped the new high-tech landscape of payment systems to this day [3]. Conceptually, m-payments and crypto currencies like Bitcoins are a new form of value transfer [23]. They rely more on the advanced features of mobile devices and the Blockchain technology. The customer does not have to pay any transaction costs. The storekeeper must pay these transaction costs. For example, PayPal asks 1.9% of the selling price for every transaction. Transaction costs, in this range, do not arise with Bitcoin [6, 24].

Hypothesis 5 A lower volatility of the Bitcoin price positively influences the benefit of using Bitcoins.

It is the aim of online shops to sell products and services. They try to improve their offers permanently. Therefore, the chief attention of enterprises is not on managing the exchange risks. It is likely that the company's profits are considerably reduced by price fluctuations. Enterprises, correspondingly, need a stable currency [10, 25]

4 Research Methods and Data Collection

Data was collected by means of an online questionnaire, which was created with the software 'LimeSurvey' [26]. The questionnaire contained 16 questions and was generated in accordance with the hypothesis created. After a short pretest, the link to the questionnaire was emailed to German online shops. Overall, more than 5000 online shops or operators, during the period from 31/05/2016 to 21/06/2016, were contacted. In order to participate in the empirical study successfully, only full-completed questionnaires were evaluated. The study had 173 returns. 61 questionnaires were not fully completed. A sample of $n = 112$ was thereby generated.

Of 112 companies surveyed, about 53 could consider accepting Bitcoins as payments. Eleven companies already accept Bitcoins. The remaining 48 companies decline to accept Bitcoins as payment.

These are mainly small and medium enterprises among the companies surveyed. In around 65% of the participating 112 companies, there are fewer than ten employees. For around a quarter (about 24%), there are up to 50 employees. 12 (about 10%) respondents employ up to 250 employees. Two companies have up to or more than 500 employees. In order to analyze the causal model with the obtained data, the method of structural equation modeling (PLS-SEM) was used [27].

PLS-SEM is a method of multivariate data analysis, which provides a deep insight into the analysis of the data as it concentrates in particular on individual relationships between existing variables [28]. Before the structural equation model (SEM) and the hypothetical relationships between the latent variables can be calculated, the measurement model must first be evaluated, which can have either formative or reflective variables or constructs [28]. The statistical software SmartPLS 3 was used to test the measurement and structural model due to its particular robustness and low data requirements [28–30]. The advantage of this method is both the identification of causal relationships and the measurement of direct as well as indirect

dependencies without the necessary condition of large samples [28]. In addition, Smart PLS, in contrast to the two methods AMOS or LISREL for example, enables the calculation of structural path coefficients and tests their statistical significance using bootstrapping [28].

5 Results

In the first hypothesis (safety positively influences the benefit of using Bitcoins), the analysis by SEM has a value of 0.246 at the significance level 0.806. This means that the security of transactions has no significant effect on the potential benefits of Bitcoins in B2C E-Commerce. The hypothesis has to be rejected. One possible explanation is that the security of Bitcoin payments based on a Blockchain technology is already large and many companies are no longer afraid of fraud (Fig. 2).

With reference to the second hypothesis (a fast transaction positively influences the benefit of Bitcoin application), a result of 0.841 for the path coefficient at the significance level 0.401 could be observed. That implicates, that the speed of a transaction has no significant positive impact on the potential benefits of Bitcoins. Thus, this hypothesis must also be rejected. One reason for this could be that all transactions already take place rapidly. For most companies, it makes no difference whether they receive their money after a few minutes or a day.

The third hypothesis (a higher acceptance positively influences the benefit of using Bitcoins) has a value of 1.864 and a significance level of 0.063. That variable has a positive impact on the use of Bitcoins. Thus, the hypothesis can be confirmed.

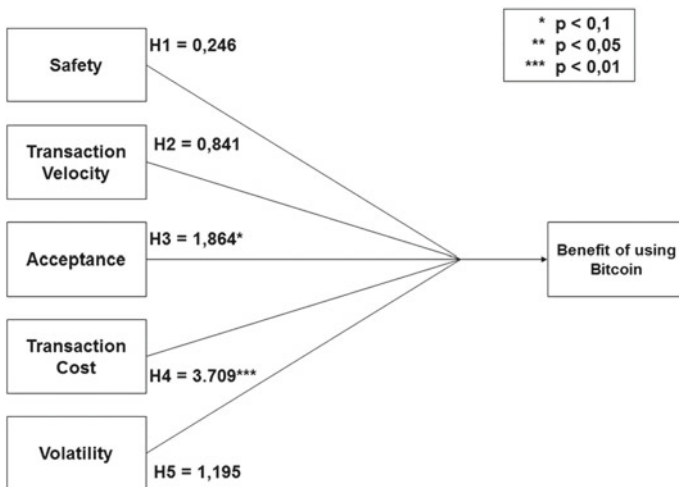


Fig. 2 Structural equation model with coefficients

Table 2 SEM coefficient

Path	Path coefficient	Significance (P-value)
Safety [Symbol] benefit of using Bitcoin in B2C E-Commerce	0.246	0.806
Transaction velocity [Symbol] benefit of using Bitcoin in B2C E-Commerce	0.841	0.401
Acceptance [Symbol] benefit of using Bitcoin in B2C E-Commerce	1.864	0.063
Transaction costs [Symbol] benefit of using Bitcoin in B2C E-Commerce	3.709	0.000
Volatility [Symbol] benefit of using Bitcoin in B2C E-Commerce	1.195	0.233

This means that a higher acceptance of Bitcoins influences the potential benefits of Bitcoin in B2C E-Commerce.

In the fourth hypothesis (low transaction costs positively influence the benefit of using Bitcoins), the analysis by SEM has a value of 3.709 at the significance level of 0.000. This represents a considerable positive influence. Thus, this hypothesis can be confirmed. The lower the transaction costs, the greater the potential benefits of Bitcoins in B2C E-Commerce.

The last hypothesis (a lower volatility of the Bitcoin price positively influences the benefit of using Bitcoins) gives a value of 1.195 at the significance level 0.233. This hypothesis must be rejected since the volatility of Bitcoin exchange rate does not significantly affect the potential benefits of Bitcoins in B2C E-Commerce. One reason for this could be that the price is no longer significantly changed, as was the case in the past.

Generally, it can be said that the acceptance of digital currency and the low transaction costs of Bitcoins are the key success factors for this digital currency.

The important values of the SEM are summarized in Table 2. The coefficient of determination is satisfactory $R^2 = 0.298 > 0.19$.

6 Conclusion

This paper explores different aspects of the potential benefits of Bitcoins in a B2C E-Commerce environment. The acceptance and the transaction costs are important indicators, which have a significant influence on the potential use of Bitcoins in E-Commerce. The investigation also gives ideas for the relevance of further factors. Our study can help academics understand and develop some of the key aspects of Bitcoins and combine them with other currencies or even new security concepts. We contribute to the literature on how the potential use of Bitcoin is affected.

This study is subject to certain restrictions. First, the survey was conducted only within Germany. Therefore, no global statement can be made. Second, no qualitative survey was carried out. Further, Bitcoin is a new topic for online traders who were interviewed. Because of the fundamental rejection of online surveys for large companies, mostly small and medium enterprises have participated in this study.

Because of these limitations, there may be a need for further research. The survey could be conducted internationally. This would make it possible to compare differences of each country. Further, a qualitative study may also be carried out to online merchants, who have experience with Bitcoin for a certain period or already accept Bitcoin as a payment method. The study should take place on a regular basis to make changes in payment through digital currency more visible.

Acknowledgements We thank Tobias Rieger and Sebastian Schmid for supporting our research.

References

1. S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system (2008)
2. C. Reichstein, R. Härting, P. Neumaier, Understanding the potential value of digitization for business, in *Agents and Multi-Agent Systems: Technologies and Applications*, vol. 96 (Springer, Berlin, Heidelberg, 2018), pp. 287–298
3. S. Blakstad, R. Allen, New payments landscape, in *FinTech Revolution*. (Palgrave Macmillan, Cham 2018)
4. E. Murphy, M. Murphy, M. Seitzinger, Bitcoin: questions, answers, and analysis of legal issues. *Congr. Res. Serv.* (2015)
5. A. Rogojanu, L. Badea, The issue of competing currencies, case study-bitcoin. *Theor. & Appl. Econ.* **21**(1), 103–114 (2014)
6. M. Van Alstyne, Why bitcoin has value. *Commun. ACM* **57**(5), 30–32 (2014)
7. R. Schmidt, M. Möhring, D. Glück, R. Haerting, B. Keller, C. Reichstein, Benefits from using Bitcoin: empirical evidence from a European country. *Int. J. Serv. Sci., Manag., Eng., Technol.* **7**(4), 48–62 (2016)
8. T. Bamert, C. Decker, L. Elsen, R. Wattenhofer, S. Welten, Have a snack, pay with Bitcoins, in *Peer-to-Peer Computing (P2P), IEEE Thirteenth International Conference (IEEE, Trento, Italy, 2013)*, pp. 1–5
9. E. Androulaki, G.O. Karame, M. Roeschlin, T. Scherer, S. Capkun, Evaluating user privacy in bitcoin, In, *International Conference on Financial Cryptography and Data Security 2013*. (Springer, Berlin, Heidelberg, 2013), pp. 34–51
10. M. Polasik, A.I. Piotrowska, T.P. Wisniewski, R. Kotkowski, G. Lightfoot, Price fluctuations and the use of Bitcoin: an empirical inquiry. *Int. J. Electron. Commer.* **20**(1), 9–49 (2015)
11. S. Barber, X. Boyen, E. Shi, E. Uzun, Bitter to better—how to make bitcoin a better currency, in *International Conference on Financial Cryptography and Data Security 2012, LNCS*, volume 7397 (Springer, Berlin, Heidelberg, 2012), pp. 399–414
12. G.O. Karame, E. Androulaki, S. Capkun, Double-spending fast payments in bitcoin, in *Proceedings of the 2012 ACM conference on Computer and communications security (ACM, New York, NY, USA, 2012)*, pp. 906–917
13. J. Brito, A. Castillo, Bitcoin: A primer for Policymakers. Mercatus Center at George Mason University. (Mercatus Center at George Mason University, 2013)
14. C. Decker, R. Wattenhofer, Information propagation in the bitcoin network, in *Peer-to-Peer Computing (P2P), IEEE Thirteenth International Conference*. (IEEE, Trento, Italy 2013), pp. 1–10

15. M. Andrychowicz, S. Dziembowski, D. Malinowski, L. Mazurek, Secure multiparty computations on bitcoin, in *Security and Privacy (SP)* (IEEE, San Jose, CA, USA, 2014), pp. 443–458
16. E.B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, M. Virza, Zerocash: decentralized anonymous payments from bitcoin, in *IEEE Symposium on Security and Privacy (SP)* (IEEE, Berkeley, CA, USA, 2014), pp. 459–474
17. European Central Bank, *Virtual Currency Schemes—A Further Analysis*. (Frankfurt am Main, 2015)
18. W.J. Luther, Bitcoin and the future of digital payments. *Indep. Rev.* **20**(3), 397–404 (2016)
19. D. Ron, A. Shamir, Quantitative analysis of the full bitcoin transaction graph, in *International Conference on Financial Cryptography and Data Security* (Springer, Berlin, Heidelberg, 2013)
20. S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G.M. Voelker, S. Savage, A fistful of bitcoins: characterizing payments among men with no names, in *Proceedings of the 2013 conference on Internet measurement conference* (ACM, Barcelona, Spain, 2013), pp. 127–140
21. I. Miers, C. Garman, M. Green, A.D. Rubin, Zerocoin: anonymous distributed e-cash from bitcoin, in *Security and Privacy (SP)*. (IEEE Berkeley, CA, USA, 2013), pp. 397–411
22. E. McCullum, N. Paul, Bitcoin: Property or Currency? *Tax Notes*, **148**(8) (2015)
23. J. Liu, R.J. Kauffman, D. Ma, Competition, cooperation, and regulation: Understanding the evolution of the mobile payments technology ecosystem. *Electron. Commer. Res. Appl.* **14**(5), 372–391 (2015)
24. R. Böhme, N. Christin, B. Edelman, T. Moore, Bitcoin: Economics, technology, and governance. *J. Econ. Perspect.* **29**(2), 213–238 (2015)
25. P. Ciaian, M. Rajcaniova, D.A. Kancs, The economics of BitCoin price formation. *Appl. Econ.* **48**(19), 1799–1815 (2016)
26. LimeSurvey: The most popular Free Open Source Software survey tool on the web, <https://www.limesurvey.org/>. Last accessed 20 Oct 2018
27. D. Hooper, J. Coughlan, M. Mullen, Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res. Methods* **6**(1), 53–60 (2008)
28. J.F. Hair Jr., G.T.M. Hult, C. Ringle, M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. (Sage Publications, Thousands Oaks, 2016)
29. C.M. Ringle, S. Wende, J.M. Becker. SmartPLS 3. Boenningstedt: SmartPLS GmbH (2015). <https://www.smartpls.com/>. Last accessed 10 Apr 2018
30. W.W. Chin, The partial least squares approach to structural equation modeling, in *Modern methods for business research*, vol. 295, no. 2 (Psychology Press, New York, 1998), pp. 295–336

Smart Cabin: A Semantic-Based Framework for Indoor Comfort Customization Inside a Cruise Cabin



Atieh Mahroo , Daniele Spoladore , Massimiliano Nolich, Raol Buqi, Sara Carciotti and Marco Sacco

Abstract This paper introduces Smart Cabin, a semantic-based framework for indoor comfort metrics customization inside a cruise cabin. Smart Cabin merges Ambient Intelligence, Ambient Assisted Living and Context Awareness perspectives to provide customized comfort experience to the cruise's passengers. Considering that passengers may be afflicted by some impairments, Smart Cabin aims at mitigating discomfort situations by providing tailored comfort settings. The framework leverages ontological representations of passengers' health conditions, activities, cabin's environment, and available devices to provide passengers with indoor temperature, humidity rate, CO₂ concentration, and luminance suitable for their health conditions and to the activities they want to perform inside the cruise cabin. Passengers' interactions with Smart Cabin are performed with a simple smartphone application, while the ontologies composing the knowledge base are reasoned and hosted on a semantic repository. Two use cases depict the framework's functioning in two typical scenarios: saving energy when the passenger leaves the cabin while reestablishing customized comfort when she/he returns, and adapting indoor comfort metrics when two or more passengers decide to perform different activities inside the same cabin.

A. Mahroo (✉) · D. Spoladore · M. Sacco
Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing (STIIMA), National Research Council of Italy (CNR), 23900 Lecco, Italy
e-mail: atieh.mahroo@stiima.cnr.it

D. Spoladore
e-mail: daniele.spoladore@stiima.cnr.it

M. Sacco
e-mail: marco.sacco@stiima.cnr.it

M. Nolich · R. Buqi · S. Carciotti
DIA-University of Trieste, 34127 Trieste, Italy
e-mail: mnolich@units.it

R. Buqi
e-mail: rbuqi@units.it

S. Carciotti
e-mail: scarciotti@units.it

Keywords Ontology · Indoor comfort customization · Ambient intelligence · Ambient assisted living · Internet of things

1 Introduction

In recent decades, the naval industry has faced a growing concern regarding the passenger's comfort inside the cruise ship in general and inside the cabin in particular. There have been many studies concerning the indoor comfort inside other transportation vehicles; however, little study has been conducted in regards to the comfort inside cruise ships. Indoor comfort metrics have been the topic of many studies and much is known about an "able-bodied" person inside any indoor environment. However, little is known about the comfort requirements for a person with disabilities. Nevertheless, in the context of Ambient Assisted Living (AAL) [1], it is important to ensure Ambient Intelligence (AmI) [2] and Context Awareness (CA) [3] to guarantee smart services.

The aim of these services is to help elderlies and people with disabilities to live more independently, in a safe, healthy, and socially connected environment [4]. AmI encompasses the ideas of ubiquitous computing, by adding intelligent automation and human-computer intuitive interaction. This requires incorporating sensors and actuators as physical means and artificial intelligence (AI) as a reasoning brain behind the system.

This work describes Smart Cabin, a framework for passenger's comfort in a cruise ship cabin, including people with disabilities. Smart Cabin exploits Semantic Web technologies and the protocol of Internet of things (IoT) [5], to enable the interoperability among different knowledge domains. This can be done through the exchange of information between the ubiquitous devices mounted inside the cabin, interchanging and manipulating the knowledge underlying the framework, and near real-time reasoning over data to retrieve the most relevant information and adaptability.

In order to make the cabin smarter, several factors are considered including passenger presence, health conditions, preferences, activities, and feedbacks. Therefore, the cabin must be equipped with a synergic system consisting of heterogeneous devices (sensors and actuators) mounted inside the cabin to gather data about the current status of the indoor environment, a comprehensive knowledge base provisioning passenger's information, a reasoning system to make decisions according to the gathered information coming from sensors and actuators—which actually make the decisions happen.

In this context, Semantic Web technologies could be a promising solution to tackle the knowledge representation of information coming from different domains. Knowledge needs to be captured, processed, recycled, and interconnected to be considered as relevant. Thus, exploiting the ontology to manage knowledge bases, and enriching the knowledge bases by deriving new facts using reasoning techniques, can be a robust solution.

Smart Cabin relies on domain ontologies—formal and explicit specifications of shared conceptualizations [6]—modeled with Resource Description Framework (RDF) [7], Web Ontology Language (OWL) [8] and Semantic Web Rule Language (SWRL) [9]—which are W3C-endorsed standard languages.

Smart Cabin allows some indoor comfort metrics to be adjusted according to the passenger's health conditions and needs; in particular, the framework takes into account indoor temperature and humidity rate, CO₂ concentration, and lighting.

The remainder of this paper is organized as follows: Section. 2 highlights some of the relevant works in the field of indoor comfort customization; Section. 3 delves into Smart Cabin's architecture; Section. 4 depicts two use cases and Smart Cabin's features; finally, the Conclusions summarize the main outcomes of this paper and sketch the future works.

2 Related Works

Passengers' comfort on a cruise ship is a little-debated topic in scientific literature; there are works addressing the cabin as a social space [10], investigating the acoustic comfort for passengers [11], and examining the leisure cruise service environment [12], but only a few papers addressed the issue of indoor cabin comfort. Recently, Buqi et al. [13] addressed the possibility to provide customized indoor comfort to cruise passengers relying on a smartphone app and a decision support system; similarly, Spoladore et al. [14] envisioned the possibility to rely on Semantic Web Technologies to provide tailored comfort inside a cabin. These two works highlight how a cabin can be seen as a—downsized—living environment, in which AmI and CA technologies can be adopted to foster indoor comfort personalization.

There exists a variety of works related to the provision of customized indoor comfort metrics, but only few of them adopt the user's health condition and preferences as one of the determinants for comfort customization and actuation. In [15], the authors leveraged semantic representations of domestic environment, comfort metrics, and dwellers' health conditions to configure the indoor comfort settings of a smart home. Tila et al. [16] adopted a context ontology-based approach to manage indoor comfort metrics, providing also the means for the description of actuators and sensors. Frešer et al. [17] developed a decision support system to improve the quality of temperature, humidity rate, and CO₂ concentration in an indoor environment, leveraging on reasoning processes enabled by the exploitation of Semantic Web Technologies. Adeleke et al. [18] proposed an ontology for indoor air quality monitoring and control, formalizing some of the knowledge of the standard ISO 7730:2005. Stavropoulos et al. [19] developed an ontology for smart building and AmI, mainly focusing on services, hardware energy management, and some concepts regarding the context. In the context of AmI and CA, Smart Cabin provides a semantic-based framework able to deliver customized comfort to passengers; furthermore, encompassing AAL vision, it takes into accounts passengers' specific disabilities to provide them with a tailored and comfortable experience inside the cruise cabin.

3 The Architecture of the Smart Cabin

This section describes the architecture of the Smart Cabin framework and its modules in detail. As mentioned in Sect. 1, in the context of AAL, the possibility of enabling indoor environment responsiveness to the inhabitants' needs and comfort requirements plays a pivotal role. In this context, this work proposes the Smart Cabin, a system that aims at providing the maximum level of comfort within the cruise cabin considering the diverse groups of people including people with disabilities, and adjusting the system according to the various activities they might be performing. In this regard, the first issue to be addressed is the physical environment and its substantial smart devices to ensure the passengers' comfort. The cabin must be equipped not only with proper living furniture, but also with necessary smart and ubiquitous devices to be prepared for exploitation by the Smart Cabin framework. Thus, a strong and solid network of sensors and actuators is needed to sense, measure, and exchange the data both from sensors to the application—where a decision is made—and from the application to the actuators. The application designed in this project is a user interface mobile application that is installed on the passenger's smartphone prior to boarding. The passenger may log in for more secure and tailored services and provide the basic information to the application such as name, gender, and disability status. The Smart Cabin considers a range of activities to be done by the passenger inside the cabin, thus it modifies the indoor environment to leverage the comfort metrics according to that specific passenger's needs and preferences.

For example, the passenger decides to read: she/he selects the type of reading she/he would like to dedicate to (relax, study or work) on her/his smartphone application. She/he also chooses the position inside the cabin where she/he would like to read (bed, desk, or balcony). The cabin system is aware of passenger's impairment, and adapts the comfort metrics according to her/his need, activity and position. In order to make this happen, Smart Cabin needs a comprehensive framework composed of four different layers as follow:

- a. Physical cabin equipped with smart devices to be connected to the application;
- b. User interface smartphone application where the passenger can log in and interact with the system;
- c. Knowledge base including the domain ontologies, hosted on a Stardog semantic repository equipped with SL reasoner to run SPARQL Protocol and RDF Query Language (SPARQL) [20]; and
- d. Middleware Java program to communicate between (b) and (c), translating the information from Smart Cabin application to semantic repository and vice versa.

Figure 1 sketches the high-level perspective of the framework architecture and communication between its modules, which enables the interoperability between different layers of the platforms. Data lifecycle in this scenario starts from the smart devices (physical layer), routes to the Smart Cabin application, passes through a middleware program, is manipulated inside the semantic repository, and then comes back through the middleware, and is submitted back to the application to trigger the actuators in the cabin.

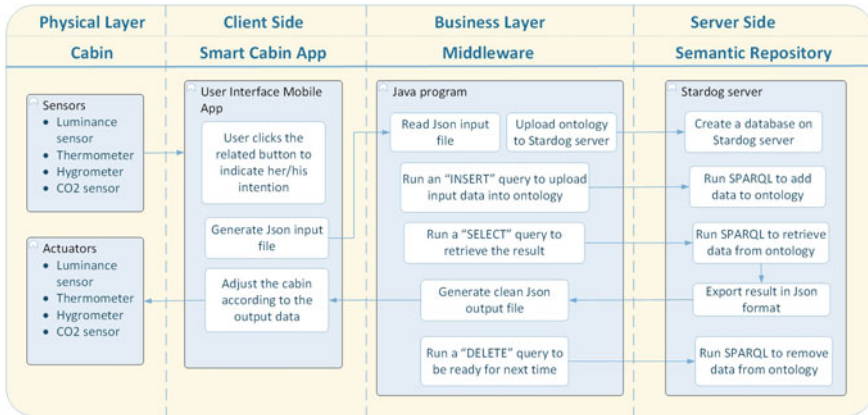


Fig. 1 Overall architecture of the framework divided into four main modules

3.1 Real Cabin

In this work, a cabin of $3 \times 6 \text{ m}^2$ is being considered and it has been provided by Fincantieri company in the University of Trieste for system deployment. As it is shown in Fig. 2, the cabin is divided into four different functional areas—entrance and bathroom; living area (including sofa, wardrobe, TV, desk, and a chair); sleeping area (including a double bed, and two bedside tables); and balcony (including a chair and a small table).

In addition to the cabin’s furniture, the physical layer of this architecture consists of a network of interconnected devices, which produces data streams transmitted by means of active transponders to proper receivers—thus, making the cabin equipped

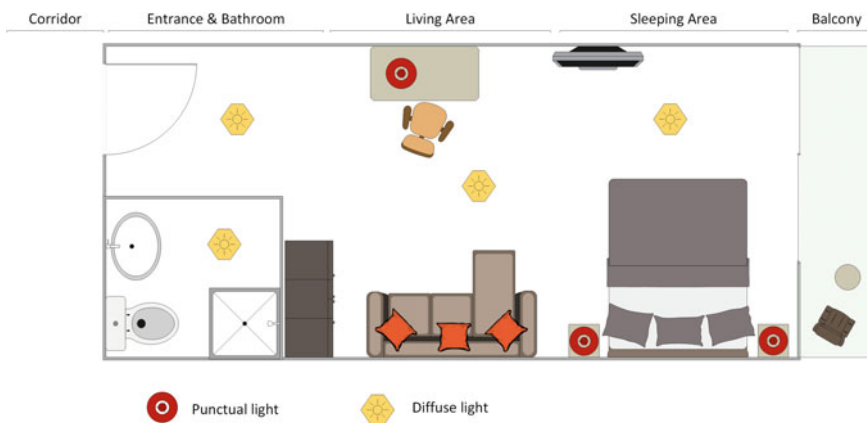


Fig. 2 Cabin plan divided into four functional areas

with a set of sensors and actuators to communicate between the environment and the framework. These smart devices could be divided into two categories as follow:

- a. Sensors (illuminance, temperature, humidity, and CO₂ concentration sensors);
- b. Actuators (Philips Hue lamps, and PCE thermo-hygrometer).

3.2 Smart Cabin User Interface Application

Smart Cabin App is an application developed using Android Studio. Its aim is to help the passengers to reach a state of comfort inside the cruise cabin. User Interface (UI) design focuses on anticipating what users need; the visual interface is original, minimal, intuitive, and functional. The Home Page layout of the application is divided into four parts:

- a. General settings,
- b. Weather (weather situation, location, outside and inside temperature),
- c. Functionality (cabin, activity, atmosphere, devices), and
- d. User's settings.

The UI has elements that are easy to access, understand, and use, thanks to the grid's disposition [21]. The four "functionality" icons of the Home Page are arranged in the main part of the screen. The central position allows the user to view the situation of the devices, to check the cabin's characteristics, to choose the desired atmosphere, or to initiate a new activity.

Figure 3 shows three different activities that passenger can choose to perform in order to have different light settings within different functional areas of the cabin.

The simplicity and ease of access are two main factors for designing the application, due to the fact that one of the main objectives of the project is to include the elderly and people with disabilities.

3.3 Knowledge Base and Semantic Repository

As discussed in Sect. 3.1, a set of sensors and actuators has to be mounted within the cabin to measure, analyze, and adapt the indoor comfort metrics. However, there must be a control center—with special sets of rules and reasoning logic—between the sensors and the actuators in which decisions have to be made. In other words, sensors measure the data about indoor comfort, send them to the semantic repository to be saved and reasoned over in the knowledge base, and finally, decisions provided by reasoning process are sent to the actuators to initiate the required action. Smart Cabin's knowledge base, its different domains' ontologies, semantic repository, and

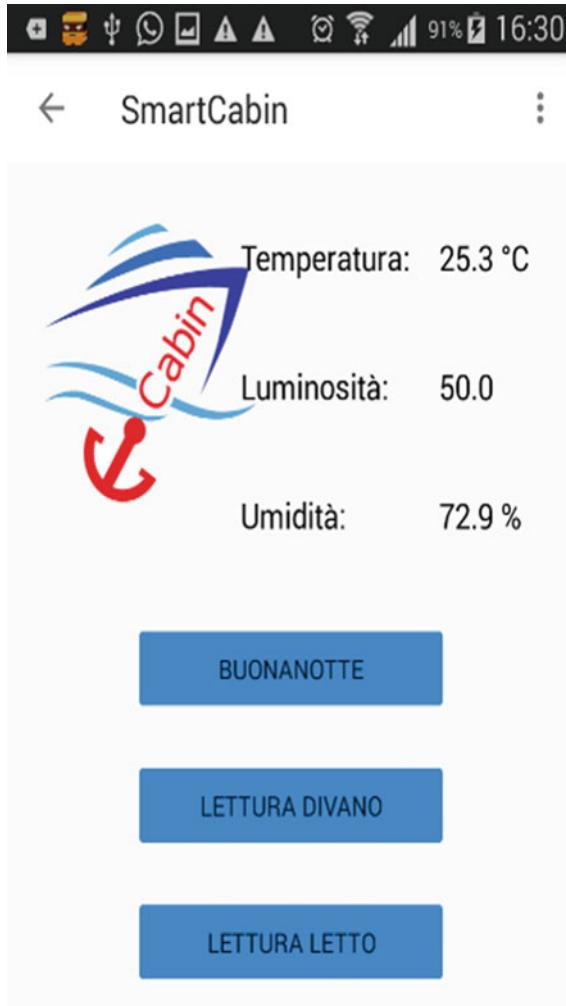


Fig. 3 A snapshot of Smart Cabin application indicating some indoor comfort metrics

reasoner are stored on a private server to be available anytime while being protected. Thus, the second layer of this framework architecture consists of the following:

- a. A set of ontologies for semantic representation of knowledge about both the cabin environment and the passenger’s characteristics and health conditions modeled with RDF and OWL;
- b. A set of rules defined in SWRL to infer new pieces of information;
- c. A reasoner engine to reason over data, rules (SWRL), ontologies’ constraints, and axioms to infer new data;

- d. A semantic repository to upload the ontologies on the server to allow querying and retrieving data; and
- e. SPARQL to query over the semantic repository and allowing to insert, retrieve, and delete the information modeled in the ontologies.

The semantic model—discussed in [14]—is composed of several modules to describe different domains of knowledge, composing a comprehensive knowledge representation of the cabin and its passengers. The ontologies developed for the Smart Cabin framework are:

- a. Passenger’s status ontology. This semantic model provides the means to describe the passenger, her/his registry records (gathered from the purchase she/he has made for the cruise ticket) and her/his health condition. The modeling of the health condition of the passengers relies on the International Classification of Functioning, Disability and Health (ICF) [22]—which allows Smart Cabin knowledge base to reuse an existing and WHO-endorsed ontology. In this way, it is possible to describe the passenger’s functional impairments: in fact, ICF is a shareable and standard language for disability description that conceptualizes the functioning of an individual as a “dynamic interaction between a person’s health condition, environmental factors, and personal factors” [23]. ICF provides a set of codes, each indicating a specific impairment that can be completed with qualifiers in order to state the magnitude of the impairment (1st qualifier) and—only for impairments in body structures—the origin of the impairment (2nd qualifier) and its location (3rd qualifier). Figure 4 provides an example of the passenger’s health condition modeling. In this case, the passenger suffers from light sensitivity and has an impairment in the eyelid of her/his right eye.

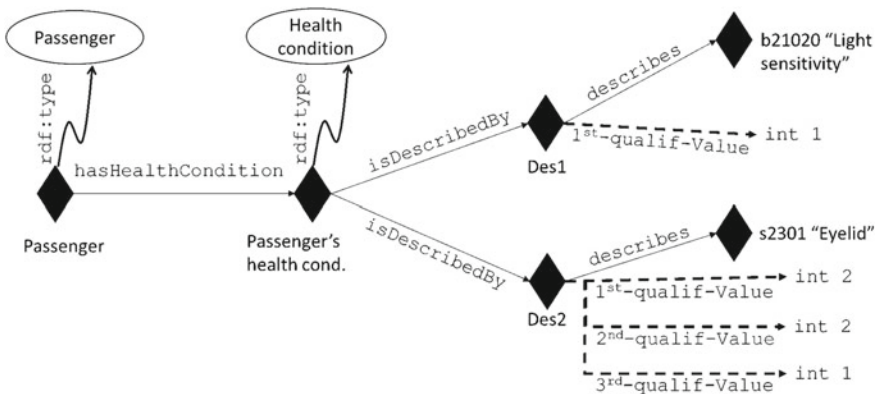


Fig. 4 An example of Passenger’s health condition modeling with ICF. Diamonds represent individuals, circles represent concepts, arrows represent roles (dashed for datatype property, full-line for object properties); the type of an individual is stated with a curved arrow

- b. Passenger's preferences ontology. Passenger's preferences regarding comfort metrics (such as light intensity, indoor temperature, etc.) are also modeled and saved relying on a set of datatype properties.
- c. Activities performable by a passenger. A list of typical activities that a passenger can perform inside a cabin is modeled in an application ontology (reading, relaxing, sleeping, watching TV, etc.).
- d. Cabin and devices ontology. The space composing the cabin is described leveraging on a simple model created from scratch, which provides the means to describe the areas composing a cabin. The devices deployed in the cabins (sensors, appliances, and actuators) are described by reusing the Smart Appliances REference (SAREF) ontology [24], an ontology providing the means to formally describe the appliances and their measurements; measurements provided by sensors are also modeled with reusing an existing ontology design pattern [25].

3.4 *Middleware Java Program*

This section describes the middleware program between the semantic models and the smartphone application. This middleware program is developed to connect the smartphone application to the semantic knowledge base and make them communicate with each other and exchange information in various ways.

The Smart Cabin application needs to access the semantic repository to retrieve the information, insert new pieces of data coming from sensors, and/or delete the old information. However, modifying the ontology, which is already on the server, is not a trivial task to do due to the Open-World Assumption (OWA) of monotonic nature of Description Logic (DL) [26]. One of the consequences of the adoption of OWA is the impossibility for a deductive reasoner to infer the existence of a new instance unless it is already modeled in the knowledge base. As a result, inserting a new piece of information into the semantic repository is not a task supported by DL-based technologies [27]. On the other hand, retrieving precise information from the semantic repository must be done through the execution of SPARQL queries in which the exact passenger's data should be provided to run the query. Moreover, running the proper query to retrieve the most accurate and relative information at each moment, is something that has to be done automatically.

As a result, Smart Cabin relies on a Java-based program acting as a middleware between the smartphone application and the semantic repository to solve all of these issues. This program runs each time the passenger decides to perform an activity inside the cabin and taps the related button on the smartphone application to state her/his intention. The moment the passenger taps the related button on the application to perform an activity, the middleware program runs to generate a proper query with necessary input data to be able to retrieve information regarding the passenger's preferences and indoor comfort. These input data are passed from application to the middleware through a JSON file indicating the passenger's unique ID, the activity

she/he would like to perform, and the place in which she/he would like to perform that activity. This JSON file is fed into the middleware as an input to generate the precise SPARQL queries to insert that specific situation inside the ontology, retrieve the inferred data according to the new data inserted, and finally, delete the inserted data to have the ontology ready for the next execution. After generating the correct SPARQL query, the middleware program runs the Stardog server—in which the ontology is saved—and executes three different SPARQL queries generated by the program in the following order:

- a. INSERT query, to add new triple data about the passenger’s activity she/he intends to perform, and the place in which she/he would like to perform her/his activity;
- b. SELECT query, to retrieve triple data about the passenger’s comfort metrics and preferences based on her/his health condition, the activity she/he intends to perform, and the place in which she/he would like to perform her/his activity; and
- c. DELETE query to remove triple data from ontology on Stardog repository which had been inserted in (a), to be free for further reuse.

4 Use Cases Scenarios

Smart Cabin framework combines various technologies to enable the AAL, AmI, and CA environment running. Considering the complexity of the technological problem, it is important to define some use case scenarios to depict the interaction between the passengers and the Smart Cabin to demonstrate how these smart features could help users to perform their activities more conveniently. In the following use cases, a cabin is considered to be related to a middle-aged couple traveling together in order to demonstrate how these smart features could help them to perform their activities inside the cabin more conveniently.

4.1 “Find Indoor Comfort as You Left It” Mode

The first use case illustrates how Smart Cabin optimizes the energy consumption within the cabin. Smart Cabin is equipped with the presence sensors to be able to detect the presence of the passengers inside the cabin while one or both of them are inside the cabin. The cruise ship is equipped with a network of Bluetooth Low Energy (BLE) beacons as a proximity sensor to locate the passengers on the ship due to the fact that it is not possible to rely on Global Positioning System (GPS) as there is no Internet connection on the ocean. When the passengers leave the cabin and reach a certain distance from her/his cabin, the system switches the status of the Smart Cabin to the power-saving mode to maximize energy saving. However, when the passengers return to the cabin, Smart Cabin sets the environment prior to

passenger's arrival when they reach a certain distance from the cabin. The system resumes the indoor comfort metrics as the passengers left them before leaving. In this way, Smart Cabin offers the possibility to save energy and preserve passenger's comfort.

4.2 Having Two Passengers Inside a Cabin Performing Different Activities

The second use case demonstrates how Smart Cabin adapts itself to the different needs of the passengers performing different activities to boost the level of comfort and welfare. Smart Cabin is able to customize the comfort requirements based on the activity the passenger is performing. Considering only one passenger in a cabin, it could be done easily, however, having two passengers traveling together in a cabin makes it more challenging. Assume a middle-aged couple is traveling together. While the wife has no particular impairment, the husband suffers from light sensitivity—a condition modeled in his ontology with ICF, as depicted in Fig. 4. In this case, the Smart Cabin provides him with the specific light with a precise amount of intensity, the right direction, and color to support his need. This feature is important when it comes to activities that need focus and alertness like reading. When the husband decides to read at the desk, he has to reveal his intention to the system through the smartphone application indicating the type of reading he would like to perform (such as study, work, or hobby), the particular type of support for reading (such as book, magazine, newspaper, digital book, etc.), and his position within the cabin (such as sitting behind the desk, sitting on the couch, lying on the bed, or sitting in the balcony area). Smart Cabin then tunes the closest punctual light to him with proper settings (180 lx intensity, 2500 K temperature) adjusted to his need to support the entire reading session for him with constant and suitable luminance. In this way, the wife can relax while listening to music in another area of the cabin, and Smart Cabin provides her with a relaxing luminous setting (120 lx intensity, 1800 K).

5 Conclusion and Future Works

This work introduces Smart Cabin, a semantic-based framework aimed at enhancing indoor comfort for cruise cabin passengers. The framework relies on Context Awareness, Ambient Intelligence, and Ambient Assisted Living paradigms to encompass also passengers with disabilities' needs; according to evidences [25, 28], Semantic-based technologies provide a sharable and machine-understandable representation of passengers' health condition and can trigger environmental actuation to help them in performing several activities. Smart Cabin relies on a server-based architecture and can be operated by the passengers via a simple smartphone application.

Future works foresee the validation of the Smart Cabin framework and, in particular, the validation of the smartphone application using standard questionnaires and tests—such as the Mobile App Rating Scale (MARS) test, the Technology Acceptance Model 2 (TAM2), and the System Usability Scale (SUS).

Acknowledgements This study is part of Project AGORÀ, a Research and Innovation project coordinated by Fincantieri S.p.A. with the participation of the National Research Council (CNR); the project received grants from the Italian Ministry of Infrastructures and Transport (MIT). The authors would like to acknowledge Daniel Celotti, Alessandro Trotta, Nicola Bassan, and Paolo Guglia from Fincantieri S.p.A. for the support provided for this paper.

References

1. P. Rashidi, A. Mihailidis, A survey on ambient-assisted living tools for older adults. *IEEE J. Biomed. Health Inform.* **17**(3), 579–590 (2013)
2. E. Aarts, R. Wichert, Ambient intelligence, in *Technology Guide* (Springer, 2009), pp. 244–249
3. M. Baldauf, S. Dustdar, F. Rosenberg, A survey on context-aware systems. *Int. J. Ad Hoc Ubiquitous Comput.* **2**(4), 263–277 (2007)
4. H. Sun, V. De Florio, N. Gui, C. Blondia, Promises and challenges of ambient assisted living systems, in *Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on*, 2009, pp. 1201–1207
5. A. Dohr, R. Modre-Opsrian, M. Drobics, D. Hayn, G. Schreier, The internet of things for ambient assisted living, in *2010 seventh international conference on Information technology: New generations (ITNG)*, 2010, pp. 804–809
6. N. Guarino, D. Oberle, S. Staab, “What is an ontology?”, in *Handbook on ontologies*, Springer, 2009, pp. 1–17
7. J.Z. Pan, Resource description framework, in *Handbook on ontologies* (Springer, 2009), pp. 71–90
8. D.L. McGuinness, F. Van Harmelen, others, OWL web ontology language overview. W3C recommendation, **10**(10), 2004 (2004)
9. I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Groszof, M. Dean, others, SWRL: a semantic web rule language combining OWL and RuleML. W3C Member Submission, **21**, 79 (2004)
10. C.M. Yarnal, D. Kerstetter, Casting off: An exploration of cruise ship space, group tour behavior, and social interaction. *J. Travel. Res.* **43**(4), 368–379 (2005)
11. B. Goujard, A. Sakout, V. Valeau, Acoustic comfort on board ships: an evaluation based on a questionnaire. *Appl. Acoust.* **66**(9), 1063–1073 (2005)
12. R.J. Kwortnik, Shipscape influence on the leisure cruise experience. *Int. J. Cult., Tour. Hosp. Res.* **2**(4), 289–311 (2008)
13. R. Buqi, S. Carciotti, M. Cipriano, P. Ferrari, M. Nolich, A. Rinaldi, W. Ukovich, D. Celotti, P. Guglia, Cruise cabin as a home: smart approaches to improve cabin comfort, in *International Workshop Formal Ontologies Meet Industries*, 2015, pp. 100–112
14. D. Spoladore, S. Arlati, M. Nolich, S. Carciotti, A. Rinaldi, D. Celotti, P. Guglia, Ontologies’ definition for modeling the cabin comfort on cruise ships, in *Proceedings of NAV 2018: 19th International Conference on Ship and Maritime Research*, 2018, pp. 100–112
15. D. Spoladore, S. Arlati, M. Sacco, Semantic and virtual reality-enhanced configuration of domestic environments: the smart home simulator. *Mob. Inf. Syst.* **2017** (2017)
16. F. Tila, D.H. Kim, Semantic IoT system for indoor environment control—a sparql and SQL based hybrid model. *Adv. Sci. Technol. Lett.* **120**, 678–683 (2015)

17. M. Frešer, B. Cvetkovi, A. Gradišek, M. Luštrek, Anticipatory system for T–H–C dynamics in room with real and virtual sensors, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1267–1274
18. J.A. Adeleke, D. Moodley, An ontology for proactive indoor environmental quality monitoring and control, in *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists*, 2015, p. 2
19. T.G. Stavropoulos, D. Vrakas, D. Vlachava, N. Bassiliades, BOnSAI: a smart building ontology for ambient intelligence, in *Proceedings of the 2nd international conference on web intelligence, mining and semantics*, 2012, p. 30
20. E. Prud, A. Seaborne, others, SPARQL query language for RDF (2006)
21. J.J. Garrett, *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education (2010)
22. W.H. Organization, others, *International Classification of Functioning, Disability and Health: ICF*. (World Health Organization, Geneva, 2001)
23. D. Spoladore, Ontology-based decision support systems for health data management to support collaboration in ambient assisted living and work reintegration, in *Working Conference on Virtual Enterprises*, 2017, pp. 341–352
24. L. Daniele, F. den Hartog, J. Roes, Created in close interaction with the industry: the smart appliances reference (SAREF) ontology, in *International Workshop Formal Ontologies Meet Industries*, 2015, pp. 100–112
25. D. Spoladore, M. Sacco, Semantic and dweller-based decision support system for the reconfiguration of domestic environments: RecAAL. *Electronics* 7(9), 179 (2018)
26. F. Baader, I. Horrocks, U. Sattler, Description logics as ontology languages for the semantic web, in *Mechanizing Mathematical Reasoning*, (Springer, 2005), pp. 228–248
27. A. Mahroo, D. Spoladore, E.G. Caldarola, G.E. Modoni, M. Sacco, Enabling the smart home through a semantic-based context-aware system, in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2018, pp. 543–548
28. D. Spoladore, S. Arlati, S. Carciotti, M. Nolich, M. Sacco, RoomFort: an ontology-based comfort management application for hotels. *Electronics* 7(12), 345 (2018)

Formal Modeling and Analysis of Probabilistic Real-Time Systems



Christian Nigro, Libero Nigro and Paolo F. Sciammarella

Abstract This paper considers formal modeling and analysis of distributed timed and stochastic real-time systems. The approach is based on Stochastic Time Petri Nets (sTPN) which offer a readable yet powerful modeling language. sTPN are supported by special case tools which can ensure accuracy in the results by numerical methods and the enumeration of stochastic state classes. These techniques, though, can suffer of state explosion problems when facing large models. In this work, a reduction of sTPN onto the popular Uppaal model checkers is developed which permits both exhaustive non-deterministic analysis, which ignores stochastic aspects and it is useful for functional and temporal assessment of system behavior, and quantitative analysis through statistical model checking, useful for estimating by automated simulation runs probability measures of event occurrence. The paper provides the formal definition of sTPN and its embedding into Uppaal. A sensor network case study is used as a running example throughout the paper to demonstrate the practical applicability of the approach.

Keywords Stochastic time petri nets · Probabilistic real-time systems · Timing constraints · Model checking · Statistical model checking · Uppaal

C. Nigro
Independent Computer Professional, Rende, Italy
e-mail: christian.nigro21@gmail.com

L. Nigro (✉) · P. F. Sciammarella
DIMES, University of Calabria, Rende, Italy
e-mail: l.nigro@unical.it

P. F. Sciammarella
e-mail: p.sciammarella@dimes.unical.it

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_5

1 Introduction

Many software systems built today are concurrent/distributed in character and have timed and probabilistic/stochastic aspects. For proper operation of these systems, both functional and non-functional (e.g., reliability, timing constraints) correctness have to be checked early in a development. Building a system with a timing violation, in fact, can have severe consequences in the practical case. Therefore, the use of formal tools both for modeling and property analysis is strongly recommended [1, 2].

In this paper, the Stochastic Time Petri Nets (sTPN) formalism [3–5] is adopted for abstracting the behavior of a timed and probabilistic system. The modeling language is supported as a special case by the Oris toolbox [6] which admits generally distributed timers for transitions (activities), that is not necessarily Markovian, and can exploit *numerical methods* and the *enumeration of stochastic state classes* (state graph or transition system) [3, 4] for quantitative analysis. The approach, although accurate in the estimation of system properties, can suffer of *state explosion problems* when facing complex realistic systems.

The work described in this paper claims that a more practical yet general solution for supporting sTPN is possible by using the popular and efficient Uppaal model checkers [7], in particular the symbolic model checker [8] for non-deterministic analysis and/or the statistical model checker (SMC) [9, 10], for quantitative evaluation of probability measures of event occurrences. Uppaal SMC does not build the model state graph but rather depends on simulation runs which are automated according to the desired level of accuracy in the results. Therefore, the memory consumption is linear with the model size, and thus, large systems can be modeled and analyzed. Although potentially less accurate than a method which uses numerical techniques, the SMC approach is anyway capable of generating results which are of value from the engineering practical point of view.

This paper extends the preliminary authors' work reported in [5], by improving the reduction of sTPN onto Uppaal and by focusing on performance prediction of complex distributed probabilistic real-time systems [1, 2].

The paper structure is as follows. Section 2 describes the syntax and semantics of the sTPN modeling language. Section 3 proposes a realistic real-time sensor network case study which is studied throughout the paper. Section 4 details a mapping of sTPN onto Uppaal which opens to both non-deterministic exhaustive model checking and to quantitative statistical model checking of an sTPN model. Section 5 describes the experimental analysis work carried out on the chosen case study. Section 6 concludes the paper and indicates some directions of further work.

2 The Formalism of Stochastic Time Petri Nets

2.1 Basic Concepts

As in classical Petri nets [11], an sTPN [3–5] is composed of a set of *places* (circles in Fig. 1), a set of *transitions* (bars in Fig. 1), and *arcs* (arrows in Fig. 1) connecting places to transitions or transitions to places only. A place is an *input* or *output* place of a transition, depending on if an arc exists which goes from the place to the transition (input arc), or vice versa (output arc). All the net objects have attributes. Places can have *tokens* (small black dots in Fig. 1), arcs have *weights* (natural numbers, by default 1) to condition transition enabling on the basis of the tokens into the input places, and transitions have *temporal* and *probabilistic/stochastic* information which constrain their firing. A transition is *enabled* if sufficient tokens exist in its input places, as required by the input arc weights. When a transition is enabled, it can fire. At the fire time, a number of tokens are withdrawn from the input places according to the input arc weights, and a number of tokens are deposited into the output places, always in a measure stated by the output arc weights. The firing of a transition is an *atomic* event and can influence the enabling status of the other transitions in the net model.

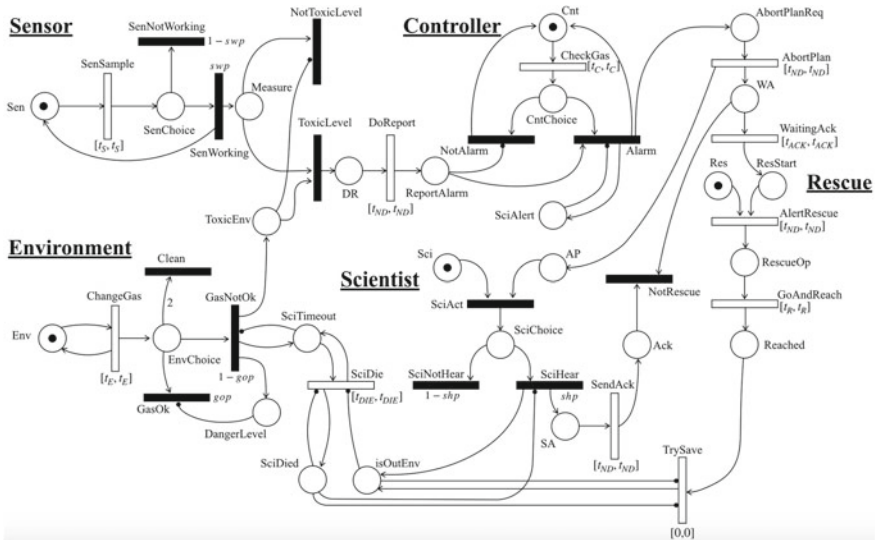


Fig. 1 An sTPN model for a sensor network real-time system

2.2 Syntax

An sTPN is a tuple $(P, T, B, F, I_{\text{nh}}, m_0, \text{EFT}, \text{LFT}, \pi, DF)$ where P is a set of places; T a set of transitions, with $P \cup T \neq \emptyset$ and $P \cap T = \emptyset$; B is the *backward function*: $B : P \times T \rightarrow N^+$ which associates an input arc $(p, t) \in B$ with its natural (not zero) weight (default is 1); F is the *forward function*: $F : T \times P \rightarrow N^+$ which associates to an output arc $(t, p) \in F$ its natural (not zero) weight (default 1); I_{nh} is the set of inhibitor arcs (input arcs ending with a black dot in Fig. 1); $I_{\text{nh}} : P \times T$, which have an implicit weight of 0. m_0 is the *initial marking* of the sTPN model, which assigns a number of tokens (also 0) to each place: $m_0 : P \rightarrow N$. EFT and LFT are, respectively, the *earliest firing time* and the *latest firing time* of a transition, as in the basic Time Petri nets [12, 13]: $\text{EFT} : T \rightarrow Q^+$, where Q^+ is the set of positive rational numbers including 0, $\text{LFT} : T \rightarrow Q^+ \cup \{\infty\}$, with $\text{EFT} \leq \text{LFT}$. The set of transitions consists of two disjoint subsets: $T = T_i \cup T_s$, $T_i \cap T_s = \emptyset$, where T_i is the set of *immediate transitions* (black bars in Fig. 1), T_s is the set of *stochastic transitions* (white bars in Fig. 1). Immediate transitions are implicitly associated with the times $\text{EFT} = \text{LFT} = 0$. In addition, π is a function which associates to each immediate transition a *probabilistic weight*: $\pi : T_i \rightarrow [0, 1]$, where $[0, 1]$ is the dense interval of real numbers between 0 and 1. DF is a function which associates to each stochastic transition a *probability distribution function* (pdf), which is constrained in the timing interval $[\text{EFT}, \text{LFT}]$ which acts as the *support* for the pdf: $DF : T_s \rightarrow \text{pdf}$. By default, the pdf of a stochastic transition is the uniform distribution function defined on the support $[\text{EFT}, \text{LFT}]$ of the transition. The pdf can be an exponential distribution function (EXP) defined by its rate parameter λ , or it can be a generally distributed non-Markovian pdf.

The sTPN formalism adopted in this paper differs from the definitions in [3, 4] because arcs can have an arbitrary weight. Moreover, our sTPN language clearly distinguishes the immediate from the timed/stochastic transitions. Only to immediate transitions, a probabilistic weight can be attached, whereas in [3, 4], each transition can have its weight.

2.3 Semantics

Enabling. A transition t is enabled in a marking m , denoted by: $m[t >, \text{iff}$:

$$\begin{aligned} \forall p \in P, (p, t) \in I_{\text{nh}} \Rightarrow M(p) = 0 \wedge \\ B(p, t) > 0 \Rightarrow M(p) \geq B(p, t) \end{aligned}$$

Firing. An enabled transition can fire. When $t \in T$ fires, it modifies the current marking m into a new marking m' as follows:

$$m \sim(p) = m(p) - B(p, t) (\text{withdraw sub} - \text{phase})$$

$$m'(p) = m^{\sim}(p) + F(p, t)(\text{deposit sub} - \text{phase})$$

where m^{\sim} is the *intermediate marking* determined by the withdrawal sub-phase. The two sub-phases (withdraw and deposit) are executed atomically. They are explicitly indicated because the firing of t can change the enabling status (from not enabled to enabled or vice versa) of other transitions in the model, due to the sharing of some input places (conflict situations), both just after the withdraw or after the deposit sub-phase (also considering the existence of inhibitor arcs). A transition t' is said *persistent* to the firing of t iff: $m[t' > \wedge m^{\sim}[t' > \wedge m'[t' > >$. Transition t' is said *newly enabled*, since the firing of t , if: $m'[t' >$. sTPN assumes that transitions are regulated by *single server* firing semantics. In other terms, each transition will fire its enablings one at a time, sequentially.

Immediate transitions always are fired before stochastic transitions. Let C_i^m be the candidate set of immediate transitions enabled in marking m : $C_i^m = \{t_i \in T_i | m[t_i >\}$. Transition t_i is chosen for firing with probability:

$$\frac{\pi(t_i)}{\sum_{t_j \in C_i^m} \pi(t_j)}$$

Fireability of timed/stochastic transitions. The firing process of transitions in a sTPN model is now described in more details. Each transition, except for the immediate transitions, has a built-in *timer* which is reset at its enabling and automatically advances toward the firing time. An sTPN model rests on *global time* and on the fact that *all* the timers increase with the same rate.

Under *non-deterministic semantics*, as in classic Time Petri Nets [12, 13], probabilistic weights and pdf s are ignored, and a transition is said *fireable* as long as the timer value is *within* the [EFT, LFT] interval of the transition. Therefore, the transition cannot fire when timer < EFT, but it should fire when the timer has reached the EFT and it is less than or equal to the LFT (*last time point*). It is worth noting that, due to conflicts, an enabled transition can lose its enabling at any instant of the timer and even at the last time LFTa. It is not possible for a continually enabled transition, to fire beyond the LFT (*strong firing model*). In the case multiple transitions are fireable, one of them is chosen non-deterministically and fires.

Under *stochastic semantics*, at its enabling, a transition gets a sample d (duration) from its associated pdf which must be: $\text{EFT} \leq d \leq \text{LFT}$, then it resets its timer. The transition is fireable when the timer == d , provided the transition does not loose its enabling in the meanwhile. When multiple stochastic transitions are fireable, one of them is chosen non-deterministically and it is fired.

3 A Real-Time Modeling Example

In the following, a realistic distributed probabilistic and timed system is considered as a case study. For validation purposes, the example is adapted from [1, 2] and concerns modeling the behavior of a sensor network. In [1, 2], the model was achieved by timed actors and asynchronous message passing.

There is a laboratory wherein a scientist is working. In the environment of the laboratory, a gas level can grow so as to become toxic. One or multiple sensors are used to monitor the gas level. In the case a toxic gas level is sensed, the life of the scientist is threatened and thus, she/he has to be immediately asked to abandon the laboratory. However, due to non-deterministic and probabilistic behavior, the request to abandon the laboratory can possibly not occur at all (the sensor can be faulty), or even that the dangerous situation is correctly sensed, the scientist can be in a position of not hearing the request. As a consequence, following a given deadline for the scientist to acknowledge the request to abandon the laboratory, a rescue team is asked to go and reach the laboratory so as to possibly save the scientist. There is a deadline, tied to the dangerous character of the gas, measured from the time a toxic gas level occurs, for the scientist to be saved. Failing to rescue the scientist, in a way or another, within the deadline causes the scientist to die.

The case study was modeled as an sTPN as shown in Fig. 1 together with its initial marking. Black bars denote immediate transitions. White bars indicate timed/stochastic transitions. In the model in Fig. 1, all the stochastic transitions are deterministic. The model was designed to be a *good* abstraction [14] of the chosen system, in the sense that only the relevant actions are reproduced. Not essential aspects are omitted. In particular, the model focusses on the reaction of the system to one *single* toxic gas level. The model consists of the following components: the Environment, the Sensor (can be multiple instantiated), the Controller, the Rescue and the Scientist. Each component is mirrored by a distinct place (Env, Sen, Cnt, Sci and Res) whose token represents its ability to perform actions.

At each period t_E , the environment chooses if the gas level is ok or not ok. In the case the gas is not ok, one token is generated in the place ToxicEnv and in the place SciTimeout. Transition SciDie, with time t_{DIE} , activates the main deadline for the scientist saving process. Would SciDie fire, the scientist dies (one token is generated in the place SciDied). The Sensor, with period t_S , samples the environment gas. However, with probability swp (sensor working probability) the sensor is actually working, and with probability $1 - \text{swp}$ it can become not working. A not working sensor remains faulty forever and will not be able to inform the controller about a toxic gas. All of this was achieved by a *random switch* between the conflicting SenWorking and SenNotWorking immediate transitions of the sensor. In a similar way, the environment decides between gas ok or gas not ok during its operation. It should be noted, though, that following the first firing of GasNotOk, the transition will no longer become enabled. A toxic gas level is reported by the sensor to the controller through a firing of the ToxicLevel transition (which can fire only one time), which is followed by a firing of the DoReport transition whose timing expresses the

net communication delay (t_{ND}). One token in the ReportAlarm place, causes an alarm event to be signaled to the controller. However, also the controller has a cyclic behavior with period t_C . Hence, an Alarm is actually heard at the *next* period of the controller. Only one firing of the Alarm transition can occur during a model reaction. Alarm deposits one token in the AbortPlanReq which, after one net delay, triggers an AbortPlan event thus depositing one token in the AP place and in the WaitingAck place. Following a t_{ACK} time, the controller knows the scientist was not responding to the abort request. Therefore, the rescue is alerted to go and reach the scientist (the GoAndReach transition fires). For realistic modeling, the scientist *can* hear a request to abort plan with probability shp and with probability $1 - shp$ she/he cannot hear the request. If the scientist is hearing, then an ack is sent to the controller through the transition SendAck which definitely causes the rescue team to be *not* activated. In addition, hearing the abort plan request, determines the scientist to generate one token in the IsOutEnv which mirrors the scientist exited the laboratory and she/he is saved. In the case the scientist is not hearing, the rescue will try to save the scientist by a firing of the TrySave transition. Such a transition does not fire if the scientist was already saved or she/he died. As a subtle point, TrySave was made timed with $[0, 0]$ time interval, to give priority to the event of responding to the abort plan request (see NotRescue) would the ack be generated at the same time.

As a final remark, generating one token in SciDied or IsOutEnv terminates the response of the system to the environmental stimulus of a gas toxic level. The model in Fig. 1 can easily be extended to accommodate for multiple sensors. It is sufficient to replicate the SenSample transition and to adjust the initial marking of the Sen place to reflect the required number of sensors. All the SenSample transitions share Sen as the input place and SenChoice as the output place.

4 Mapping STPN onto Uppaal

sTPN are supported by the TPN Designer toolbox [15]. This way an sTPN model can be graphically edited and preliminarily simulated. An sTPN model can then be translated into the terms of the timed automata of Uppaal for model checking. A translated sTPN model can be decorated for it to be more convenient for the analysis.

The availability of a high-level modeling language with basic types (int and bool), arrays and structures of basic types (under statistical model checking is also permitted the type double), and C-like functions, greatly facilitated the reduction process. The main points of the translation are summarized in the following:

- The number of places (P) and the number of transitions (T) are determined. In particular, the number ST of stochastic transitions, and the number IT of immediate transitions are defined, with $T = ST + IT$. In addition, constant names for places and transitions are introduced as in the source model.

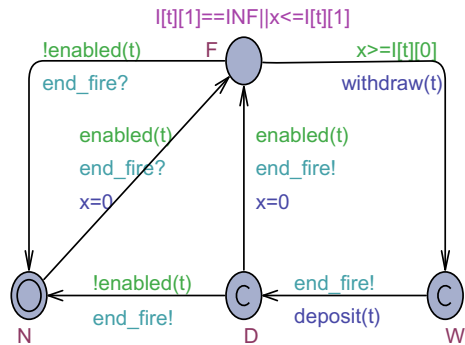
- The B and F functions of formal syntax are realized by two corresponding constant matrices whose elements are pairs of a place id and its associated weight of the input arc or output arc. An inhibitor arc has weight 0.
- The $[EFT, LFT]$ intervals of all the transitions are collected into a (constant) matrix $I: T \times 2$. If t is a transition id, $I[t][0]$ holds the EFT and $I[t][1]$ stores the LFT of t . An infinite bound for LFT is coded by the constant $INF = -1$.
- The pdf of timed transitions are implemented in a double $f(stid)$ which receives the id of a stochastic transition and returns a sample of the corresponding pdf, constrained into the associated $[EFT, LFT]$ support interval.
- The random switch in a not empty candidate set C_i^m of enabled immediate transitions in current marking, is realized by a function $rank()$ which returns the id of the Next Immediate Transition (NIT) to fire. A stochastic transition can fire provided $NIT == NONE$ that is there are no immediate transitions to fire.
- The enabling, withdraw, and deposit operations of transitions are, respectively, implemented by the bool $enabled(const tid t)$, void $withdraw(const tid t)$, and void $deposit(const tid t)$ functions. Such fundamental functions receive as parameter the id of a generic transition.

4.1 Timed Automata for Transitions

The active part of a reduced sTPN model into Uppaal is constituted by timed automata each one corresponding to a distinct transition. Basic automata are: $ndTransition(const tid t)$, $sTransition(const stid t)$, and $iTransition(const itid t)$ whose parameter is the unique id of the transition. The $ndTransition$ is used when an sTPN model is analyzed by the exhaustive symbolic model checker of Uppaal. The $sTransition$ and $iTransition$ are instead used when an sTPN model is evaluated with the statistical model checker.

Figure 2 depicts the $ndTransition$ automaton of a non-deterministic Time Petri Net transition [12], which captures the basic behavior of any sTPN transition. An

Fig. 2 $ndTransition$ automaton



ndTransition owns a locally declared clock x which implements the *timer* explained in the Sect. 2.3. The transition starts in the N (Not enabled) location. From N it moves to the F (under Firing) location as soon as the transition finds itself enabled. When moving from N to F , the clock x is reset to measure the time-to-fire. In F , the transition can stay as permitted by the LFT time. In the case LFT is infinite, the dwell-time in F is arbitrary. A finite stay in F is imposed by the invariant $x \leq I[t]$ [1] attached to F , in the case of a strict [EFT, LFT] interval. At any instant in time, the transition moves from F to N as it finds itself disabled.

As soon as the clock x reaches the earliest firing time EFT, and the transition keeps continually enabled, the transition can terminate its firing by initiating the withdraw sub-phase and switching to the W location. The complete firing process is achieved by a pair of synchronizations using the end_fire broadcast channel, raised, respectively, from the committed locations W and D (a committed location has to be left immediately, without time passage; committed locations have priority over urgent locations, for example, the F locations of transitions which have reached the LFT time). Being broadcast, end_fire is heard by *all* the remaining transitions in the model, which thus can evaluate their enabling status following, respectively, the withdraw and the deposit sub-phase of the transition which is completing its firing.

A subtle point in Fig. 2 refers to the fact that at the first end_fire synchronization (following the withdraw sub-phase), all the remaining transitions call enabled() as a guard and check effectively their enabling status in the intermediate marking established by the withdrawal of tokens (see Sect. 2.3). In the second end_fire synchronization, the influenced transitions evaluate their status in the final marking reached after the deposit sub-phase. From D , the transition will move to N if it is not enabled, or come back to F , by resetting the clock x , would it be still enabled.

For bootstrapping purposes, a Starter automaton (see Fig. 3) is used which initially launches a first end_fire synchronization and let transitions to reach the F location or remain in the N location would they be, respectively, enabled or disabled in the initial marking. After entering the $S1$ location, the Starter will take no further part in the model behavior.

Figures 4 and 5 show, respectively, the iTransition and the sTransition automata whose basic behavior coincides with that described for the ndTransition.

The F location in iTransition is a committed location, meaning that an enabled immediate transition has to complete its firing without time passage. In order to guarantee the atomicity of the firing process of a stochastic transition st , it is fundamental to forbid immediate transitions to exit F before the completion of the firing process of st . The global bool fire variable is set to true by a transition st at its exiting from F (see Fig. 5), and put to false at the end of the firing. But an immediate transition can conclude its firing only when it is selected by the rank() function which applies probability weights and realizes the random switch (see Sect. 4). The automata in

Fig. 3 Starter automaton

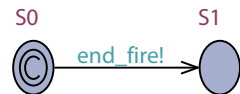


Fig. 4 iTransition automaton

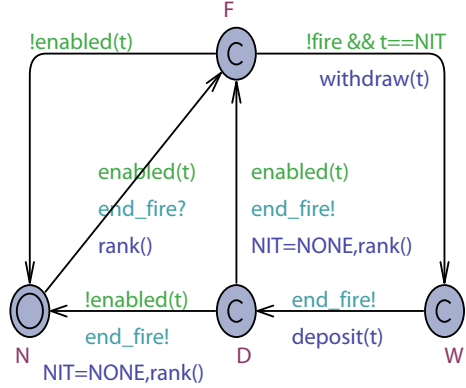
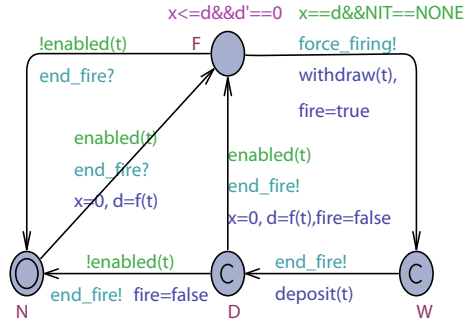


Fig. 5 sTransition automaton



Figs. 4 and 5 assume that transition conflicts are always *homogeneous*, that is they are composed by the same type of transitions: immediate or stochastic.

The stTransition uses two clocks: x and d (delay). The clock d is assigned the next sample of the pdf of the stochastic transition, returned by the $f(t)$ function. In F , the transition remains until x reaches d . During this time, the clock d is frozen by putting its first derivative to 0. The pattern exploited in Fig. 5 is suggested in the Uppaal SMC tutorial [9].

Since Uppaal SMC can have problems exiting the F location of a stochastic transition whose pdf is, e.g., deterministic, a `force_firing` broadcast and urgent channel is used in Fig. 5. This way as soon as the delay is elapsed, F is forced to be exited. The `force_firing` broadcast synchronization is non-blocking and it is heard by no one. Only its urgent character is exploited. Both designs in Figs. 4 and 5 improve previous authors' work described in [5].

The automata in Figs. 4 and 5 implement in a natural way the semantics of sTPN transitions discussed in Sect. 2.3.

5 Analysis of the Case Study Model

An sTPN model like that in Fig. 1 naturally requires to be quantitatively analyzed, e.g., by a statistical model checker [10] in order to estimate, for example, the probability for the scientist to be not saved in time when a toxic gas level occurs, when some scenario parameters like those shown in Table 1 are assumed. It is important to note that the EFT and LFT bounds of transitions *must* be integral values when the non-deterministic model checker of Uppaal is used.

Some preliminary experiments were carried out using the non-deterministic symbolic model checker of Uppaal [8] which ignores probability and pdfs. A perfect sensor was assumed (the SenNotWorking transition in Fig. 1 was omitted). However, the scientist was kept capable of perceiving or not an abort plan request, as well as the environment can choice at each period if the gas level is ok or not. It is worthy of note that by ignoring probability weights, alternate model paths which could occur with very different probabilities, are handled non-deterministically, in the sense that the model checker considers and visits them as occurring with the same probability.

5.1 Non-deterministic Analysis

The sTPN model of Fig. 1 reduced to Uppaal, and with a perfect never faulty sensor, was configured using only the ndTransition template of Fig. 2 as follows (implicit instantiations of processes occur):

system Starter, ndTransition;

Then the exhaustive model checker of Uppaal was used which relies on the construction of the model state graph. Properties are specified in the supported subset of the TCTL temporal logic [8].

Table 1 Scenario parameters for the sTPN model of Fig. 1

Parameter	Name	Value
Sensor period	t_S	2
Controller period	t_C	3
Environment period	t_E	5
Scientist saving deadline	t_{DIE}	14
Network delay	t_{ND}	1
Scientist ack deadline	t_{ACK}	2
Rescue time	t_R	3
Sensor working probability	swp	0.99
Gas ok probability	gop	0.98
Scientist hearing probability	shp	0.90

An important concern was checking that the system does not admit deadlock states:

$$A[]!deadlock$$

This query was found satisfied. This in turn also proved that the model is 2-bounded. In fact, except for the EnvChoice place of the Environment, all the other places will have at most 1 token during system evolution. This property was checked by the (satisfied) query:

$$A[]forall(p : pid)(p == EnvChoice || M[p] <= 1)$$

Since the use of a not faulty sensor, the next checked property was knowing if the intrinsic timing behavior of the model (see parameters in Table 1) can guarantee the scientist can always be saved when a toxic gas level occurs. The following queries (both satisfied) were used:

$$\begin{aligned} &ndTransition(GasNotOk).W -- > M[IsOutEnv] == 1 \\ &A[]M[SciDied] == 0 \end{aligned}$$

The first one, based on the leads $-->$ operator, checks if starting from a firing of the GasNotOk transition, it inevitably follows that one token will be generated in the IsOutEnv place (i.e., the scientist is saved). The second query, similarly, checks if *invariantly*, that is in all the states of the state graph, the marking of the SciDied place is without tokens.

Since in the assumed operating conditions, the scientist gets saved, the following query (satisfied) was used to assess that the saving is effectively performed by the AbortPlan request or through the rescue team (see the TrySave transition in Fig. 1):

$$\begin{aligned} &ndTransition(GasNotOk).W -- > \\ &ndTransition(SciHear).W || ndTransition(TrySave).W \end{aligned}$$

A critical issue in the scenario parameters in Table 1 is the *scientist die deadline* (t_{DIE}) which obviously depends, e.g., on the sensor period. Therefore, by changing the sensor period from 1 to 15, and keeping unchanged all the other parameters except for the t_{DIE} value which was set to 30, it was determined the maximum *end-to-end delay* (EED) between the occurrence of a toxic gas level and the completion of the system reaction which saves the scientist. To this purposes, a decoration clock z was added to the Uppaal model, which is reset (in the deposit(tid) function) when the GasNotOk transition concludes its firing. Then, the clock z was checked when either the SciHear or the TrySave transition concludes its firing, by a query like the following:

$$A[](ndTransition(SciHear).W || ndTransition(TrySave).W) \&\&$$

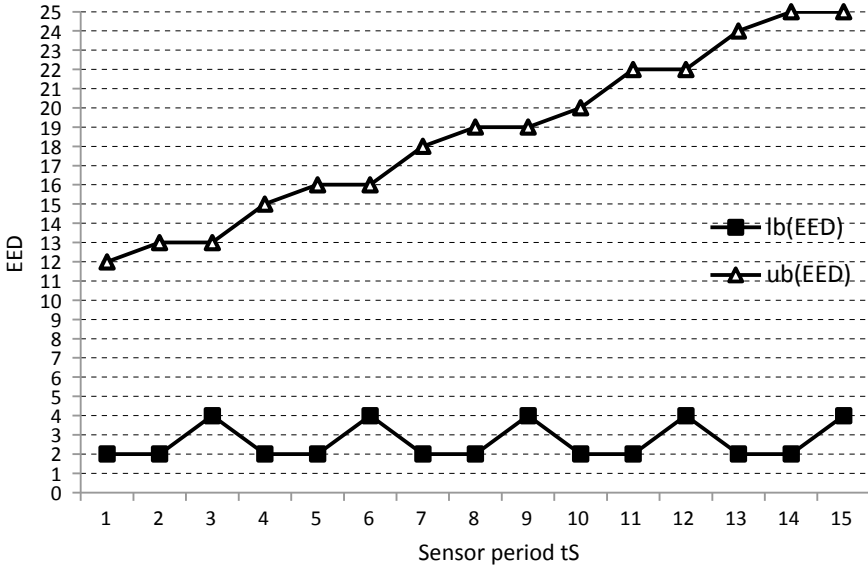


Fig. 6 Observed EED vs. sensor period

$$(M[\text{SciDied}] == 0 \ \&\& \ M[\text{IsOutEnv}] == 0) \text{ imply } z \text{ op bound}$$

where op can be \geq or \leq and the corresponding bound is the lower bound or the upper bound of the EED. More precisely, the lower bound (best-case response time) is assessed by the greatest value lb which satisfies the above query with the constraint $z \geq \text{lb}$. Similarly, the lowest value ub which satisfies the above query with the constraint $z \leq \text{ub}$ establishes the upper bound (worst-case response time). Figure 6 shows the observed lb and ub values for the monitored EED. The results coincide with those achieved in [2] using actors for modeling the case study.

Some further checks were carried on the non-deterministic model with the sensor which can fail, and thus is not able to inform the Controller about a toxic gas level. It is observed that it is the logic of exhaustive verification that of checking all the state paths, then also the path where the sensor fails.

$$E \langle \rangle \text{ndTransition}(\text{SenNotWorking}).W$$

This query is satisfied. As one expected, even assuming a t_{DIE} value greater than the worst-case value of the EED emerged, for a given sensor period, in the analysis on the optimistic model, the scientist can now die.

The query:

$$E \langle \rangle M[\text{SciDied}] == 1$$

is satisfied and clearly indicates that exists at least one state where the place SciDied holds one token. For correctness of the model, the following query was also used to check that in no case the scientist can be both saved and died.

$$E \langle \rangle M[\text{SciDied}] == 1 \&\& M[\text{IsOutEnv}] == 1$$

Such query is *not* satisfied.

From the non-deterministic analysis, it emerges that the sTPN model of the case study is compliant with the actor model developed in [2]. This in turn talks about correctness of the sTPN reduction into Uppaal.

The analysis confirms the scientist *can* possibly be not saved in the event of a dangerous gas level. This fact raises the important concern of estimating a probability measure for the scientist to be effectively saved under different operating conditions.

5.2 Quantitative Analysis

The statistical model checker (SMC) of Uppaal rests on a stochastic interpretation of timed automata. Properties are specified using the Metric Interval Temporal Logic (MITL) and its weighted extension (WMITL) [9]. An sTPN model can be naturally analyzed under SMC because it depends on broadcast synchronizations only [9].

Basically, SMC uses simulation runs whose number is dynamically adjusted according to the property to check. Important is the number of time units assigned to each experiment for reaching a conclusion. Such time units should guarantee a dangerous gas level occurs in the environment and sufficient time exists to produce a system response, a sensor not working can happen, and the scientist receiving an abort plan request can possibly be not hearing it. The sTPN model in Fig. 1 was configured to exploit stochastic and immediate transitions, thus:

system Starter, sTransition, iTransition;

The following query:

$$\text{Pr}[\leq 1000](\langle \rangle \text{iTransition}(\text{GasNotOk}).W)$$

asks to quantify the probability of occurrence of a toxic gas by using a certain number of experiments each one lasting after (at maximum) 1000 time units (tu). Each run is actually stopped as soon as the event occurs. By setting the uncertainty error of a confidence interval (CI) as $\varepsilon = 0.005$, 3013 runs were used with parameters as in Table 1. Then, the following CI (0.95 confidence degree) was proposed [0.975932, 0.98593] which confirms the expected probability of the event. The query:

$$\text{Pr}[\leq 1000](\langle \rangle M[\text{SciDied}] == 1 \&\& M[\text{IsOutEnv}] == 1)$$

checks that never should happen that the scientist can be found (absurd) both saved and died. After 368 runs, Uppaal SMC suggests a CI of $[0, 0.00997405]$ thus witnessing the event is almost impossible. For the sake of simplicity, in the subsequent SMC analysis work, the default error of $\epsilon = 0.05$ was adopted, which implies fewer runs but anyway an acceptable level of accuracy in the results.

In the hypothesis that the toxic gas requires the scientist to be saved within a t_{DIE} deadline of 10 tu, the following query based on the until (U) operator [9], with the sensor period varied from 1 to 15 and other parameters as in Table 1, was used to estimate the probability of the event: “Would an instant in time in $[0, 1000]$ exists where a token is deposited in the *SciTimeout* place and no token is present in the *IsOutEnv* place, will it happen that a token is put in the *IsOutEnv* place within the next 10 time units?”.

$$\Pr(\langle \rangle [0, 1000]((M[SciTimeout] == 1 \& \& M[IsOutEnv] == 0) \ U[0, 10]M[IsOutEnv] == 1))$$

Each execution of the query uses 738 runs. The observed confidence interval bounds, when only one sensor is used, are collected in Fig. 7 and are in good agreement with similar results reported in [2].

As one can see from Fig. 7, the scientist saving probability is low for small and for high sensor periods. In fact, the more frequent is the sensor sampling the more likely the sensor can be faulty, and thus unable to notify the controller about a dangerous gas level. On the other hand, a sensor with a high period can capture late a toxic gas level thus delaying the controller intervention and putting at a severe risk the life of the scientist.

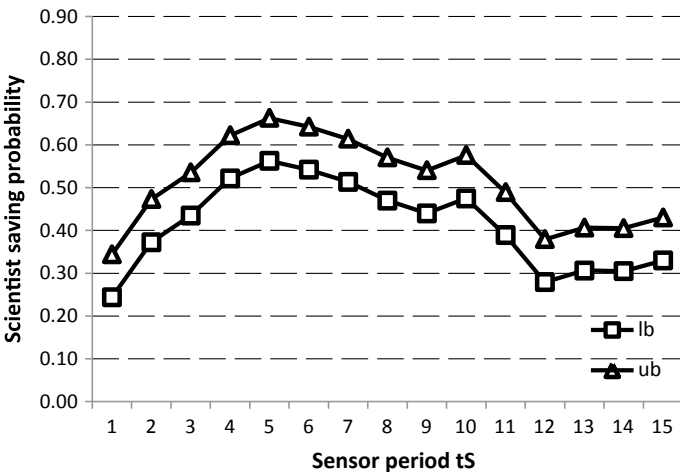


Fig. 7 Scientist saving probability versus sensor period (one sensor)

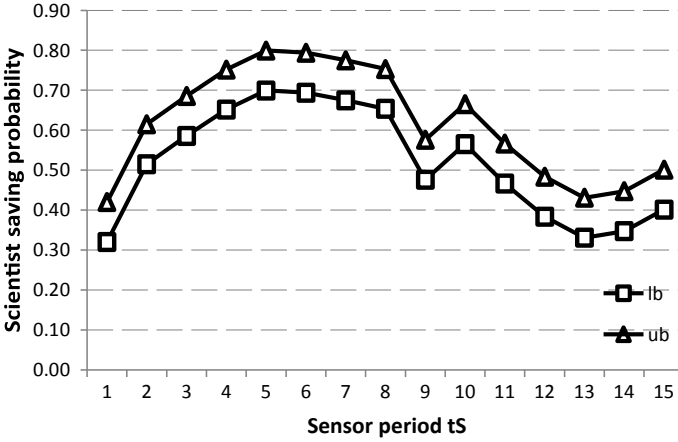


Fig. 8 Scientist saving probability versus sensor period (two sensors)

The scientist saving probability increases as the sensor period augments by taking a maximum when the period reaches the value 5. Other local maxima occur at 10 and 15 and so forth. These maximum points correspond to *parameter alignment situations* (the sensor period is a multiple of the environment changing period t_E , see Table 1), also observed in [2], when the sensor can perceive a bad gas level at the same moment the environment signals it. Near to local maxima, the saving probability keeps high because although the sensor can capture with a small delay a toxic gas, sufficient time remains for the controller to execute the rescue operations.

Figure 8 portrays the scientist saving probability when two sensors are used. This experiment was not carried in [2]. Figure 8 confirms the expectation that the more are the sensors, the less likely is the circumstance that all the sensors become faulty simultaneously. The same behavior with multiple maxima observed in Fig. 7 is also present in Fig. 8 although now the scientist saving probability is higher.

The following queries evaluate specifically the probability that the rescue operations are completed, respectively, by the abort plan request of the controller or through the rescue team intervention. The model is configured with two sensors, the sensor period is $t_S = 5$ and the scientist die deadline is $t_{DIE} = 10$. All the other parameters are as in Table 1.

$$\begin{aligned} & \Pr(\langle \rangle [0, 1000]((M[\text{SciTimeout}] == 1 \& \& M[\text{IsOutEnv}] == 0) \\ & \quad U[0, 10]i\text{Transition}(\text{SciHear}). W)) \\ & \Pr(\langle \rangle [0, 1000]((M[\text{SciTimeout}] == 1 \& \& M[\text{IsOutEnv}] == 0) \\ & \quad U[0, 10]s\text{Transition}(\text{TrySave}). W)) \end{aligned}$$

The first query, after 738 runs, generates a CI of [0.645122, 0.745122]. The second query proposes a CI of [0, 0.0838753] thus confirming that the scientist is mostly saved through the first abort plan request issued by the controller.

All the experiments were carried out on a Win 7 station with 4 GB RAM using the Uppaal version 4.1.19 and the 4.1.20.beta 25 development version.

6 Conclusions

Modeling and formal verification are fundamental tools for the development of distributed and probabilistic timed systems [1, 2]. In this paper, the Stochastic Time Petri Nets (sTPN) modeling language [3–5] is adopted. A reduction of sTPN onto the Uppaal model checkers [8, 9] is developed which enables both non-deterministic exhaustive analysis, and quantitative evaluation of system properties through statistical model checking [10, 16].

A non-trivial case study concerning a distributed and dependable real-time sensor network is introduced, modeled in sTPN, and thoroughly verified in the paper.

Prosecution of the research aims to:

- Building and making available in Uppaal SMC, a library of recurrent pdfs.
- Developing a structural approach to the operational semantics [17] of sTPN (see, e.g., [1, 2]) and formally showing the correctness of the sTPN reduction into Uppaal.
- Exploiting sTPN for performance prediction of general complex timed and stochastic systems. In fact, the particular adopted syntax of sTPN is close to the Generalized Stochastic Petri Nets (GSPN) formalism [18], when the default support interval $[0, \infty]$ is attached to each timed transition and a general distribution probability function (pdf) is chosen.
- Establishing a formal transformation of distributed probabilistic timed actors [1, 2] into sTPN so as to leverage the modeling and verification activities afforded by Uppaal. A transformed timed actor model, indeed, as demonstrated through the case study presented in this paper, can be more amenable to analysis due to its abstraction level [14] and greater efficiency and scalability during verification.

References

1. A. Jafari, E. Khamespanah, M. Sirjani, H. Hermanns, M. Cimini, PTRebecca: modeling and analysis of distributed and asynchronous systems. *Sci. Comput. Program.* **128**, 22–50 (2016)
2. L. Nigro, P.F. Sciammarella, Qualitative and quantitative model checking of distributed probabilistic timed actors. *Simul. Model. Pract. Theory* **87**, 343–368 (2018)
3. E. Vicario, L. Sassoli, L. Carnevali, Using stochastic state classes in quantitative evaluation of dense-time reactive systems. *IEEE Trans. on Soft. Eng.* **35**(5), 703–719 (2009)
4. L. Carnevali, L. Grassi, E. Vicario, State-density functions over DBM domains in the analysis of non-Markovian models. *IEEE Trans. on SW Eng.* **35**(2):178–194 (2009)
5. F. Cicirelli, C. Nigro, L. Nigro, Qualitative and quantitative evaluation of stochastic Time Petri Nets, in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, (IEEE, 2015), pp. 763–772

6. G. Bucci, L. Carnevali, L. Ridi, E. Vicario, ORIS: a tool for modeling, verification and evaluation of real-time systems. *Int. J. on Software Tools for Technology Transfer* **12**, 391–403 (2010)
7. Uppaal, on-line: www.uppaal.org
8. G. Behrmann, A. David, K.G. Larsen, A tutorial on UPPAAL, in *Formal Methods for the Design of Real-Time Systems*, vol. 3185, ed. by M. Bernardo, F. Corradini Lecture Notes in Computer Science (Springer-Verlag, 2004), pp. 200–236
9. A. David, K.G. Larsen, A. Legay, M. Mikucionis, D.B. Poulsen, Uppaal SMC tutorial. *Int. J. on Software Tools for Technology Transfer* **17**, 1–19 (2015). <https://doi.org/10.1007/s10009-014-0361-y>
10. Agha, A., Palmskog, K.: A survey of statistical model checking. *ACM Trans. Model. Comput. Simul.* **28**(1), 6:1–6:39 (2018). <http://doi.acm.org/10.1145/3158668>
11. T. Murata, Petri nets: properties, analysis and applications. *Proc. IEEE* **77**(4), 541–580 (1989)
12. P.M. Merlin, D.J. Farber, Recoverability of communication protocols: implications of a theoretical study. *IEEE Trans. Commun.* **24**(9), 1036–1043 (1976)
13. B. Berthomieu, M. Diaz, Modeling and verification of time dependent systems using Time Petri Nets. *IEEE Trans. Soft. Eng.* **17**(3), 259–273 (1991)
14. E.A. Lee, M. Sirjani, What good are models? in *International Conference on Formal Aspects of Component Software*, (Springer, Cham, 2018), pp. 3–31
15. L. Carullo, A. Furfaro, L. Nigro, F. Pupo, Modelling and simulation of complex systems using TPN designer. *Simul. Model. Pract. Theory* **11**(7-8), 503–532 (2003)
16. C. Nigro, L. Nigro, P.F. Sciammarella, Modelling and analysis of multi-agent systems using UPPAAL SMC. *Int. J. Simul. Process Model.* **13**(1), 73–87 (2018)
17. G.D. Plotkin, A structural approach to operational semantics (1981)
18. M.A. Marsan, G. Balbo, G. Conte, S. Donatelli, G. Franceschinis, *Modelling with generalized stochastic Petri nets* (Wiley, 2004)

Regional Agricultural Land Classification Based on Random Forest (RF), Decision Tree, and SVMs Techniques



Nassr Azeez, Wafa Yahya, Inas Al-Taie, Arwa Basbrain and Adrian Clark

Abstract Land cover observation based on remote sensing data demands robust classification techniques which give the precise complex land cover mapping. Scientists and researchers made great efforts in improving classification accuracy considerably. The aim of this paper is to show outcomes gained from the RF classifier and decision tree and to compare their effectiveness with the SVMs technique. The mentioned techniques are applied over the imagery we have captured with six different classes of ROI (Region Of Interest) images including unknown range. Results indicated that the performance of the random forest classifier outperforms the decision tree and SVMs techniques performance in terms of the number of mis-classifications instances and the classification accuracy with an overall accuracy of 86%, while the decision tree accuracy is 67%, and the SVMs accuracy is 56%, respectively.

Keywords Classification · RF · SVMs · Remote sensing

N. Azeez (✉) · I. Al-Taie · A. Basbrain · A. Clark
School of Computer Science and Electronic Engineering,
University of Essex, Colchester, UK
e-mail: naazee@essex.ac.uk

I. Al-Taie
e-mail: iyyalt@essex.ac.uk

A. Basbrain
e-mail: amabas@essex.ac.uk

A. Clark
e-mail: alien@essex.ac.uk

N. Azeez · I. Al-Taie
Faculty of Science, Computer Department, University of Baghdad, Baghdad, Iraq

A. Basbrain
Faculty of Computing and Information Technology,
University of King Abdul-Aziz, Jeddah, Saudi Arabia

W. Yahya
Faculty of Education, Mathematics Department,
University of Al-Hamdaniya, Ninawa, Iraq
e-mail: rwafa1993@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_6

1 Introduction

Land cover is known as the monitored structural cover on the earth's surface [9]. Land cover mapping and observing used for estimation of land cover variation, and it is considered as one of the essential applications of Earth monitoring remote sensing data. Increasing the numbers of remotely sensed images made land cover observing programs for big area mapping more easier [13]. Different classification techniques have been employed to map the land cover by using remote sensing data. Classification using remote sensing is a complicated process and requires the perception of different factors. The main image classification factors that can be taken into consideration to improve classification accuracy may include placement of a sufficient classification system, choice of training samples, extraction of features, and selection of appropriate classification techniques. Remotely sensed images classification is the process of locating land cover classes to pixels [19]. Mapping and monitoring of land cover are fundamental to estimate land cover variation [24]. In land cover classification, it is desired to use multisource or multisensor remote sensing data to extract as much data as possible from the desired area to be classified [15]. Classification based on remote sensing data is a challenging since an appropriate multivariate statistical pattern is not obtainable for such as data. Thus, the traditional statistical methods that have been utilized in the remote sensing data are not appropriate for such classification [23]. Consequently, other methods have been proposed. In 1990, Benediktsson et al. [4] used neural network classifiers. Swain and Benediktsson, 1992, proposed a statistical theory that used heuristic weighting techniques on the data source for classification [3]. In 1997, Benediktsson et al. used a hybrid approach by combining the statistical theory with the neural network classifiers and accomplished improved accuracies as compared to previous researches [2]. However, this technique is complicated in implementation. Therefore, the attention has more been turned to the ensemble classification approaches. The basic idea behind the ensemble classification is that various classifiers are trained, and the outcomes comprised by using a voting approach. Different ensemble classification methods have been introduced, and the most commonly used are boosting and bagging methods. In 1994, Breiman suggested the Bagging technique that is known as bootstrap aggregating [5]. This technique is established on training different classifiers on bootstrapped pattern taken from a training set to minimize the classification disparity. Schapire, 1999 suggested the boosting technique that utilizes iterative re-training [11, 12]. As the iterations progress, samples of incorrectly classified are given increased weight. Computationally, boosting is less speed than bagging, but in most cases, it is more precise than bagging. However, boosting has many drawbacks: it is extremely slow, sensitive to noise, and it can overtrain [7]. Random forests classifiers are kind of tree classifiers that utilizes an improved technique of bootstrapping as bagging. Random forests are the improvement to the boosting method in terms of accuracy and without the drawbacks [6]. In 2005, Ham et al. performed Random Forests classification to remote sensing data [16] and achieved good results for this kind of data set. Parametric methods are not convenient for classification of multisource data [14]. Therefore, the

Random forest technique has received great attention for multisource classification because it is nonparametric [10], and it also estimates the significance of the data channels. The remainder of the paper is orderly as follows. Section 2 looks into the following components related to the remote sensing images principles and classification in remote sensing. It is followed by a description of feature extraction based on support vector machines classifier (SVMs), decision tree, and random forest classifier. Experimental results are showed in Sect. 6, and the paper completes with our conclusions in Sect. 7.

2 Image Classification in Remote Sensing

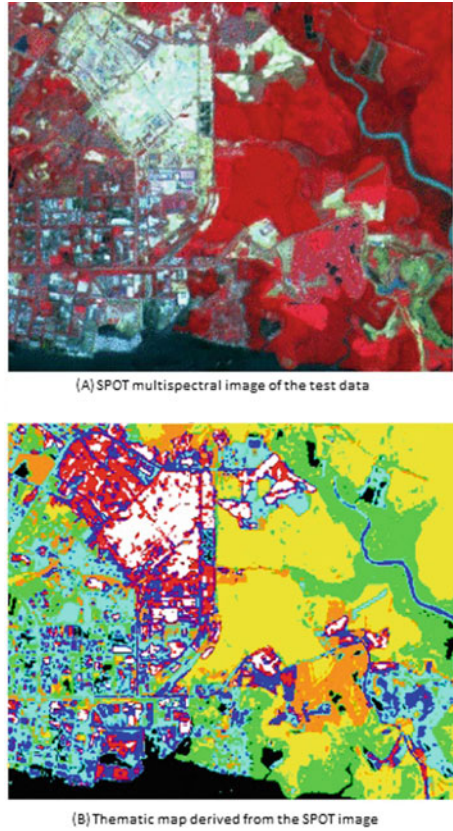
In general, remotely sensed aspect is the aggregating of information from an Earth's surface object without any direct contact with it. Remote sensing observation is made from over the object of interest, by using a sensor holed on a spaceborne platform or airborne [21, 22]. Normally, remotely sensed images are in the form of digital images, to extract important information from the images. Image processing methods can be used to enhance the image to support visual interpretation and to correct the image if it has been subjected to blurring or degradation distortion by other factors. There are various image analysis methods available used depending on the demands of the specific problem interested. In many cases, image classification techniques are employed to delineate different regions in an image into a thematic map of the study area. Usually, the thematic map can be used with other test area databases for additional analysis and utilization [22].

One of the popular functions of remotely sensed data is the production of land cover maps that can be used by a procedure known as image classification [1]. Remote sensing images classification built on the idea that various surface feature types of the earth have a various spectral reflectance, and these features are recognized by using the classification process [1]. In other words, image classification is the operation of classifying all image pixels or remote sensing satellite data to gain land cover themes [18]. As can be seen in Fig. 1

3 Support Vector Machines Classifier (SVMs)

SVMs are a supervised learning process with related learning techniques that analyse data that are used for regression and classification analysis. Recently, SVMs have become one of important techniques in the area of classification due to achieving well predicting to unseen samples with an acceptable degree of precision as compared to other traditional classifiers [27]. In addition, SVMs provide efficient and powerful classification algorithms which have the ability to deal with high-dimensional input features [26]. SVMs learning machine is defined by an optimal separating hyperplane,

Fig. 1 Example of remote sensing image classification [1]

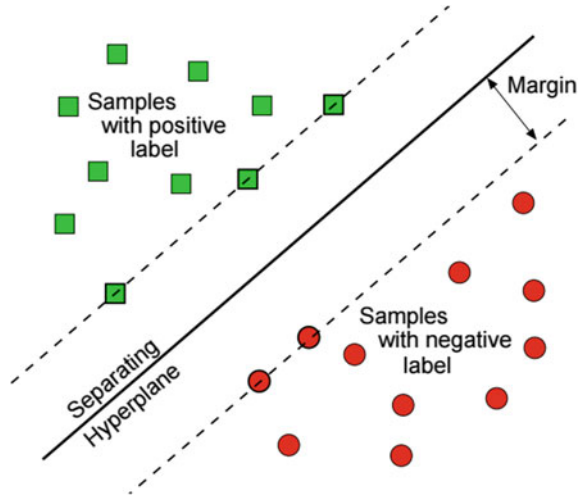


where the margin between the negative and positive samples is maximal. The solution of this technique is based on these data points, which are located on the margin [28]. These points are known as support vectors. See Fig. 2.

4 Decision Tree Classifier

It is a simple and commonly based rule system that used a supervised learning algorithm. This kind of classifier builds a classification or regression model in a tree structure. It breaks down a data set into subsets while an related decision tree is progressing incrementally at the same time. In other words, these classifiers organized a set of test conditions and questions in structure of tree to be used to perform the prediction on the test data set [25].

Fig. 2 Optimal separating hyperplane [20]



The decision trees learning algorithm is:

- Onset with the training data.
- Choose attribute along the dimension that produces the best split.
- Generate child nodes build on the split.
- Repeat on every child using child data till a stopping case is reached.

Build on the best decision tree is a key challenge in decision tree classifier. In general, many decision trees are constructed from a specific set of attributes, whereas some of these trees are more precise than others. However, many methods have been introduced to construct an accurate decision tree in an acceptable time. In general, these algorithms use a technique that grows a tree by applying a sequence of optimum decisions to choose which attribute to be employed for the best data split [25]

5 Random Forest Classification

Random forest algorithm is considered as a supervised classification algorithm. It is an improvement over the decision tree algorithm, in which a multiple numbers of small decision trees been generated from random subsets of the data [15]. Each of the decision trees builds a separated classifier where each of these classifier captures various trends in the data. Then comprises the predictions of all the trees to classify a class [8]. The random forest classifier achieves the high accuracy results with the highest number of the decision trees [24].

The random forests technique for regression and classification is summarized as follows [17]:

1. From the original data, draw n_{tree} bootstrap samples.
2. Expand a classification tree for each of the bootstrap samples with the following modulation: for every node, instead of selecting the best split in all predictors, select a random sample m_{try} of the predictors and take the optimal split from those variables.
3. Assembling the predictions of the n_{tree} trees to predict a new data. An estimation of the error rate based on the training data can be obtained by applying the following:
 - Predict new data which are not included in the bootstrap sample (OOB data) by using the tree expanded with the sample of the bootstrap.
 - Assemble the OOB predictions and calculate the OOB estimate of error rate.

6 The Experimental Work

The aim of this part is to present outcomes obtained with the RF classifier and decision tree and to compare their effectiveness with the SVMs technique. The mentioned techniques are applied over the imagery we have captured with six different classes of ROI (Region Of Interest) images including unknown range.

The Davidsons Farm, Peldon, Colchester is the research area for this study. It is situated in the south of Colchester near Mersea Island. RGB and NAR images are taken by a drone to characterize variation of vegetal cover.

Figures 3, 4, and 5 represent the outcomes of remotely sensed images classification based on RF, decision tree, and SVMs classifier, respectively.

In the experimental works, a matching matrix (confusion matrix) is used to give visualization of the algorithms performance. In another words, confusion matrix is showing mislabelling one class as another. As it is shown in Figs. 3, 4, and 5, the primary diagonal values of the table represent the number of instances that are classified correctly, while off-diagonal values represent the number of mis-classifications. Through observation, it is clarified that small numbers along the primary diagonal represent cases in which the performance of the classification is poor. Results indicate that the random forest classifier performance outperforms the decision tree and SVMs techniques performance in terms of the number of mis-classifications instances and the classification accuracy with an overall accuracy of 86%, while the decision tree accuracy is 67%, and the SVMs accuracy is 56, respectively.

```

We have 1401 samples
The training data include 6 classes: [1 2 3 4 5 6]
predict   1   2   3   4   5   6   All
truth
1         204  64  18  27  0  0   313
2         20  218 23  35  0  0   296
3          11  18 232  37  0  0   298
4           3  88  1 118  0  0   210
5          22 206  1  24  0  0   253
6           1  30  0  0  0  0    31
All       261 624 275 241 0  0  1401

Accuracy  0.56
    
```

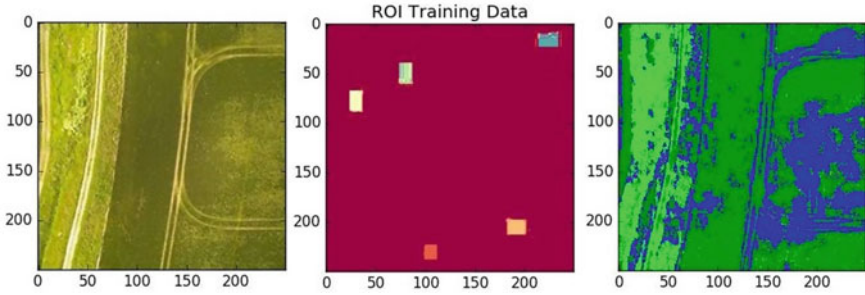


Fig. 3 Classification of remotely sensed images based on SVMs classifier

```

We have 1401 samples
The training data include 6 classes: [1 2 3 4 5 6]
predict   1   2   3   4   5   6   All
truth
1         223  29  17  18  26  0   313
2         20  199 22  18  37  0   296
3           5  11 259  18  5  0   298
4           7  50  20 112  21  0   210
5          21  57  6  19 150  0   253
6           5  18  0  3  5  0    31
All       281 364 324 188 244 0  1401

Accuracy  0.67
    
```

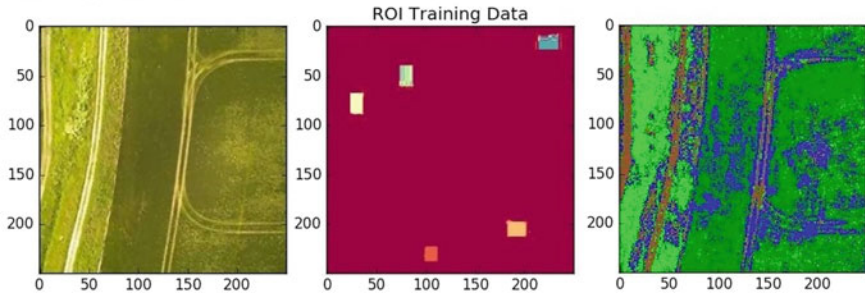


Fig. 4 Classification of remotely sensed images based on tree classifier


```

We have 1401 samples
The training data include 6 classes: [1 2 3 4 5 6]
predict   1    2    3    4    5    6    All
truth
1         262  18    4    9   17    3   313
2          7  228    6   20   35    0   296
3          0    3  286    3    5    1   298
4          1    7    3  186   13    0   210
5          8   19    1   11  211    3   253
6          1    6    0    3   10   11    31
All       279  281  300  232  291   18  1401
Accuracy  0.85
    
```

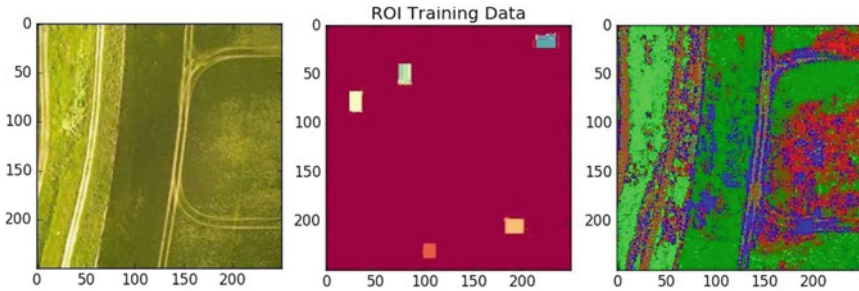


Fig. 5 Classification of remotely sensed images based on RF classifier

7 Conclusions

The information of earth’s surface is demanded in various areas. Land cover classification based on using remote sensing images is one of the popular applications in remote sensing, and many techniques have been improved and applied for remote sensing images classification. Recently, RF and SVMs, which are supervised classification techniques, have been employed in the remote sensing area. The aim of this paper is to present outcomes obtained with the RF classifier and decision tree and to compare their effectiveness with the SVMs technique. The mentioned techniques are applied over the imagery we have captured with six different classes of ROI (Region Of Interest) images including unknown range. Results indicated that the random forest classifier performance outperforms the decision tree and SVMs techniques performance regarding to the number of mis-classifications instances and the classification accuracy with an overall accuracy of 86%, while the decision tree accuracy is 67%, and the SVMs accuracy is 56%, respectively.

References

1. J. Al-doski, S.B. Mansori, H.Z.M. Shafri, *Image Classification in Remote Sensing* (Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia, 2013)
2. J.A. Benediktsson, J.R. Sveinsson, P.H. Swain, Hybrid consensus theoretic classification, in *Geoscience and Remote Sensing Symposium, 1996. IGARSS’96. Remote Sensing for a Sustainable Future.*, International, vol. 3 (IEEE, 1996), pp. 1848–1850

3. J.A. Benediktsson, P.H. Swain, Consensus theoretic classification methods. *IEEE Trans. Syst., Man, Cybern.* **22**(4), 688–704 (1992)
4. J.A. Benediktsson, P.H. Swain, O.K. Ersoy, Neural network approaches versus statistical methods in classification of multisource remote sensing data (1990)
5. L. Breiman, Bagging predictors. technicalreport421, department of statistics. (University of California, Berkeley, 1994)
6. L. Breiman, Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
7. G.J. Briem, J.A. Benediktsson, J.R. Sveinsson, Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote. Sens.* **40**(10), 2291–2299 (2002)
8. D.R. Cutler, T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J.J. Lawler, Random forests for classification in ecology. *Ecology* **88**(11), 2783–2792 (2007)
9. A. Di Gregorio, L.J. Jansen, *Land Cover Classification System (LCCS): Classification Concepts and User Manual* (FAO, Rome, 1998)
10. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern classification*, 2nd edn., vol. 58, p. 16 (Wiley, New York, 2001)
11. Y. Freund, R. Schapire, N. Abe, A short introduction to boosting. *J.-Jpn. Soc. Artif. Intell.* **14**(771–780), 1612 (1999)
12. Freund, Y., Schapire, R.E., et al.: Experiments with a new boosting algorithm. In: *Icml*, vol. 96. (Citeseer 1996), pp. 148–156
13. M.A. Friedl, D.K. McIver, J.C. Hodges, X. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper et al., Global land cover mapping from modis: algorithms and early results. *Remote. Sens. Environ.* **83**(1–2), 287–302 (2002)
14. P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, Random forest classification of multisource remote sensing and geographic data, in *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*, vol. 2 (IEEE 2004), pp. 1049–1052
15. P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, Random forests for land cover classification. *Pattern Recognit. Lett.* **27**(4), 294–300 (2006)
16. J. Ham, Y. Chen, M.M. Crawford, J. Ghosh, Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote. Sens.* **43**(3), 492–501 (2005)
17. A. Liaw, M. Wiener et al., Classification and regression by randomforest. *R news* **2**(3), 18–22 (2002)
18. T. Lillesand, R.W. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation* (Wiley, 2014)
19. D. Lu, Q. Weng, A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote. Sens.* **28**(5), 823–870 (2007)
20. F. Markowitz, L. Edler, M. Vingron, Support vector machines for protein fold class prediction. *Biom. J.* **45**(3), 377–389 (2003)
21. P. Pellikka, W.G. Rees, *Remote Sensing of Glaciers: Techniques for Topographic, Spatial and Thematic Mapping of Glaciers* (CRC Press, 2009)
22. W.G. Rees, P. Pellikka, *Principles of Remote Sensing* (Remote Sensing of Glaciers, London, 2010)
23. J.A. Richards, J. Richards, *Remote Sensing Digital Image Analysis*, vol. 3 (Springer, 1999)
24. V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote. Sens.* **67**, 93–104 (2012)
25. Rokach, L., Maimon, O.: Top-down induction of decision trees classifiers-a survey. *IEEE Trans. Syst., Man, Cybern., Part C (Appl. Rev.)* **35**(4), 476–487 (2005)
26. T.A. Salh, M.Z. Nayef, Face recognition system based on wavelet, pca-lda and svm. *Comput. Eng. Intell. Syst. J.* **4**(3) (2013)
27. S. Suralkar, A. Karode, P.W. Pawade et al., Texture image classification using support vector machine. *Int. J. Comput. Appl. Technol.* **3**(1), 71–75 (2012)
28. Weston, J.: Support vector machine (and statistical learning theory) tutorial, 4 independence way. Princeton, USA, dostupné (10.5. 2016) z http://www.cs.columbia.edu/~kathy/cs4701/documents/jason_svm_tutorial.pdf

On Hierarchical Classification Implicative and Cohesive M_{GK} -Based: Application on Analysis of the Computing Curricula and Students Abilities According the Anglo-Saxon Model



Hery Frédéric Rakotomalala and André Totohasina

Abstract Extracting association rules from a huge binary data according to a quality measure is an important pretreatment step in data analysis. Also, among unsupervised techniques, our approach for a hierarchical classification implicative and cohesive is based on the new measure of cohesion according to the interestigness measure M_{GK} . In this paper, we present, for the first time, a validation of this approach in the field of education, mainly in the computing curricula and the performance capabilities of students pursuing this curriculum in the Anglo-Saxon model.

Keywords Binary data · Quality measure M_{GK} · Implicative cohesion · Statistical implicative analysis · Oriented classification

1 Introduction and Motivations

Scientists all over the world are putting their energy into improving the ever-evolving computing curricula, thanks to technological performance and large volumes of data, as well as the constant demands of perfection of society. Considering the remarkable properties of the interestigness measure M_{GK} already studied in [1], we are moving toward Régis Gras's theory of statistical implicative analysis (SIA) to discover quasi-implicative, non-symmetric rules of type "if we have X, we have generally Y" in a binary context [2]; but this time, the extraction of association rule (AR) valid is done according to the interestigness measure M_{GK} [3]. The extracted AR-valid will then be classified in relation to the hierarchical classification implicative and cohesive (HCIC) method according to the new M_{GK} -based cohesion index [4, 5]

H. F. Rakotomalala (✉) · A. Totohasina
Department of Mathematics and Informatics High School Training for Technological Schools (ENSET), University of Antsirana, Postal Box 0 Antsirana, Madagascar
e-mail: fredericrakotomalala@yahoo.fr

A. Totohasina
e-mail: andre.totohasina@gmail.com

for a hierarchical representation of meta-rules [6]. This paper proposes a computing curricula and abilities analysis using the HCIC method according to M_{GK} .

Today, computing is almost always present in many areas related to human activities. In less advanced countries like Madagascar, the majority of the population does not use computers in a private setting as in developed countries. This has an impact on computing curricula and requires especially in-depth analyzes and innovations adapted to the country’s situation. So, it is opportune for us to see Anglo-Saxon’s model of the **computing curricula** and the **students abilities** following the programs offered.

In the USA, four state-supported organizations set standards and describe computing curricula in higher education in the Anglo-Saxon model. The ACM / AIS / IEEE-CS Technical Report [7] described five branches of computer science disciplines at the university: Computer Science (CS), Computer engineering (CE), information technology (IT), information systems (IS), and software engineering (SE). Each discipline has its own training objectives, but a lot of knowledge and performance capabilities are transversal. In CC2005 [7], the knowledge area offers 40 themes according to the consensus by the task force based on a review of the discipline-specific body of knowledge found in the most recent curriculum volume for each of the disciplines of computer science (CS, CE, IS, IT, SE) and indicates their weight in each. The weight is characterized by two values: 0 (low) and 5 (high) which represent the importance of a theme for the domain. Thus, we can construct a binary table of the relationship between the lower and higher weights of each theme of study programs in the five branches of computer science disciplines mentioned above (Cf. Table 1). The report also defines 60 performance capabilities and assesses their weight for each computer discipline (Cf. Table 2).

Table 1 Table of weight comparing computing theme through five types of study programs

	Programming fundamentals	...	Technical support
CEmin0	0	...	1
...
CEmin5	0	...	0
CEmax0	0	...	0
...
CEmax5	0	...	0
...
SEmax5	1	...	0

Table 2 Table of students abilities of computer science graduates per discipline

	Prove theoretical results	...	Evaluate new forms of search engine
CE0	0	...	1
...
CE5	0	...	0
CS0	0	...	0
...
CS5	0	...	0
...
SE5	1	...	0

2 Data Mining Tools

As it is mentioned in the title of our present paper, we essentially aim to apply the theoretical work recently published in [6] and partially published in [5]:

- extracting association rules is based on the so-called interestigness implicative measure M_{GK} ; in its favoring component M_{GK}^f which is implicative and closely related with the one degree of freedom χ^2 that enable to provide the critical values ($M_{GK(\alpha)}^f$) of M_{GK}^f ;
- support value of each extracted valid rule was given by $supp_{M_{GK}}$ [3]. To have a contrast of each value, a normalization process is necessary to calculate its respective normalized support $supp_{(n)M_{GK}}$ [4, 8];
- classifying items is based on the so-called implicative cohesion $coh_{supp_{(n)M_{GK}}}$.

According to the technical report [7], each discipline has its own training objectives, but a lot of knowledge and performance capabilities are transversal. The knowledge area offers 40 themes and indicates their weight in each of the five computer disciplines. The weight is characterized by two values: 0 (low) and 5 (high) which represent the importance of a theme for the domain. Thus, we can construct a binary table of the relationship between the lower and higher weights of each theme in the five types of study programs mentioned above (Cf. Table 1).

The technical report also defines 60 performance capabilities and assesses their weight for each computer discipline (Cf. Table 2).

The data consisting of sixty objects ($CE_{\min-\max} : 0 \div 5, CS_{\min-\max} : 0 \div 5, IS_{\min-\max} : 0 \div 5, IT_{\min-\max} : 0 \div 5, SE_{\min-\max} : 0 \div 5$) and forty variables (40 themes) for the computing curricula and thirty objects ($CE : 0 \div 5, CS : 0 \div 5, IS : 0 \div 5, IT : 0 \div 5, SE_{\min-\max} : 0 \div 5$) and sixty variables (60 abilities) for the performance capacities.

3 Results and Interpretations

After processing the binary data (Tables 1 and 2) with the HCIC- M_{GK} tool:

- In regard to the **computing curricula**, we had seven hundred eighty valid rules for a $\alpha = 10\%$ threshold set by ourself of which five hundred eight rules are positive and two hundred seventy-two are negative with $0.04 \leq \text{supp}_{(n)M_{GK}}^f \leq 1$ and formed thirty-two pairs of oriented variables, i.e., $\text{card}(\text{coh}_{\text{supp}_{(n)M_{GK}}}) = 32$ with $0.048 \leq \text{coh}_{\text{supp}_{(n)M_{GK}}} \leq 0.607$. Thus, we had fifteen meta-rules (Cf. Table 3);
- For the **students abilities**, we had one thousand seven hundred eleven valid rules for a $\alpha = 1\%$ threshold set by ourself of which one thousand nine rules are positive and seven hundred two are negative with $0.04 \leq \text{supp}_{(n)M_{GK}}^f \leq 1$ and formed seventy-eight pairs of oriented variables, i.e., $\text{card}(\text{coh}_{\text{supp}_{(n)M_{GK}}}) = 78$ with $0.607 \leq \text{coh}_{\text{supp}_{(n)M_{GK}}} \leq 1$. Thus, we had twenty-four meta-rules (Cf. Table 4).

Rules’s interpretation for computing curricula:

- **R(1):** the theme *Human-Computer Interaction* is the prerequisite of theme *Analysis of Requirements Technical*;
- **R(2):** those who master the *Software Quality* theme must master the *Software Evolution (Maintenance)* theme;

Table 3 The 15 meta-rules concerning the 40 themes in CE, CS, IS, IT, and SE

Meta-rule
R(1) = (Analysis of requirements technical \Rightarrow human-computer interaction)
R(2) = (Software quality \Rightarrow software evolution (maintenance))
R(3) = (Systems administration \Rightarrow security implementation and mgt)
R(4) = ((Software quality \Rightarrow software evolution (maintenance)) \Rightarrow software Verification and Validation)
R(5) = (Operating systems principles & Design \Rightarrow algorithms and complexity)
R(6) = (Network-centric design and principles \Rightarrow operating systems configuration & Use)
R(7) = (Technical support \Rightarrow digital media development)
R(8) = (Software process \Rightarrow ((software quality \Rightarrow software evolution (maintenance)) \Rightarrow software verification and validation))
R(9) = (Ethics/professional/legal/society \Rightarrow (analysis of requirements technical \Rightarrow human-computer interaction))
R(10) = (Software design \Rightarrow (operating systems principles & design \Rightarrow algorithms and complexity))
R(11) = (Network-centric use and configuration \Rightarrow integrative programming)
R(12) = (Intelligent systems (AI) \Rightarrow graphics and visualization)
R(13) = (Software modeling and analysis \Rightarrow information management (DB) Theory)
R(14) = (Management of info systems org \Rightarrow scientific computing (numerical mthds))
R(15) = (E-business \Rightarrow analysis of business requirements)

Table 4 The 24 meta-rules concerning the 60 performance capabilities in CE, CS, IS, IT, and SE

Meta-rule
R(1) = (Determine if faster solutions possible ⇒ develop solutions to programming problems)
R(2) = (Design a spreadsheet program e.g., excel ⇒ design a word processor program)
R(3) = (Train and support spreadsheet users ⇒ train and support word processor users)
R(4) = (Select network components ⇒ (train and support spreadsheet users ⇒ train and support word processor users))
R(5) = (Manage an organization’s web presence ⇒ (Select network components ⇒ (Train and support spreadsheet users ⇒ train and support word processor users)))
R(6) = (Design computer peripherals ⇒ design embedded systems)
R(7) = (Design complex sensor systems ⇒ (Design computer peripherals ⇒ design embedded systems))
R(8) = (Design a chip ⇒ (Design complex sensor systems ⇒ (Design computer peripherals ⇒ Design embedded systems)))
R(9) = (Program a chip ⇒ (Design a chip ⇒ (Design complex sensor systems ⇒ (Design computer peripherals ⇒ Design embedded systems))))
R(10) = (Design a computer ⇒ (Program a chip ⇒ (Design a chip ⇒ (Design complex sensor systems ⇒ (Design computer peripherals ⇒ Design embedded systems))))
R(11) = (Train and support database users ⇒ model and design a database)
R(12) = (Develop computer resource plan ⇒ (Train and support database users ⇒ model and design a database))
R(13) = (Schedule budget resource upgrades ⇒ (Develop computer resource plan ⇒ (Train and support database users ⇒ Model and design a database)))
R(14) = (Manage databases ⇒ configure database products)
R(15) = (Manage computer networks ⇒ install upgrade computer software)
R(16) = ((Schedule budget resource upgrades ⇒ (Develop computer resource plan ⇒ (Train and support database users ⇒ model and design a database))) ⇒ (Manage an organization’s web presence ⇒ (Select network components ⇒ (Train and support spreadsheet users ⇒ Train and support word processor users))))
R(17) = ((Determine if faster solutions possible ⇒ develop solutions to programming problems) ⇒ Prove theoretical results)
R(18) = (Develop business solutions ⇒ (Manage databases ⇒ configure database products))
R(19) = ((Manage computer networks ⇒ install upgrade computer software) ⇒ Install upgrade computers)
R(20) = (Maintain and modify information systems ⇒ use word processor features well)
R(21) = (Configure and integrate e-learning systems ⇒ use spreadsheet features well)
R(22) = (Implement communication software ⇒ do systems programming)
R(23) = (Implement intelligent systems ⇒ design auto reasoning systems)
R(24) = (Develop multimedia solutions ⇒ configure and integrate e-commerce software)

- **R(3)**: before addressing the topic *Systems Administration*, we must first go through the topic *Security: implementation and mgt*;
- **R(4)**: the prerequisite of *R(3)* is the *Software Verification and Validation* theme;
- **R(5)**: The theme *Algorithms and Complexity* is the crucial step before approaching the theme *Operating Systems Principles & Design*;
- **R(6)**: if we study the theme *Network Centric Design and Principles*, generally we study the theme *Operating Systems Configuration & Use*;
- **R(7)**: we must master the theme *Digital Media Development*, so as not to have difficulty learning the theme *Technical Support*;
- **R(8)**: the basis of the *Software Process* theme is to respect rule *R(4)*;
- **R(9)**: the prerequisite of the *Ethics / Professional / Legal / Society* theme is the respect of the rule *R(1)*;
- **R(10)**: *Software Design* theme depends on rule *R(5)*;
- **R(11)**: the prerequisites of the *Network Centric Use and Configuration* theme is the theme *Integrative Programming*;
- **R(12)**: *Graphics and Visualization* theme has an implicative link with *Intelligent Systems (AI)* theme;
- **R(13)**: the *Information Management Theory (DB)* theme is the prerequisite of the *Software Modeling and Analysis* theme;
- **R(14)**: if you learn the *Management of Information Systems Org.* theme, you usually learn *Scientific Computing (Digital Mthds)* theme;
- **R(15)**: to do the *E-business* theme, one needs to know generally about the *Analysis of Business Requirements* theme (Figs. 1 and 2)

Interpretation of some outstanding rules about students abilities:

- **R(1)**: all those who are able to *Determine if faster solutions possible* are able to *Develop solutions to programming problems*;
- **R(2)**: students who master the *Design of a spreadsheet users* generally seem to be able to *Design a word processor program*;
- **R(5)**: students able to *Manage an organization’s web presence* are usually able to *Select network components*, to *Train and support spreadsheet users* and to *Train and support word processor users*;

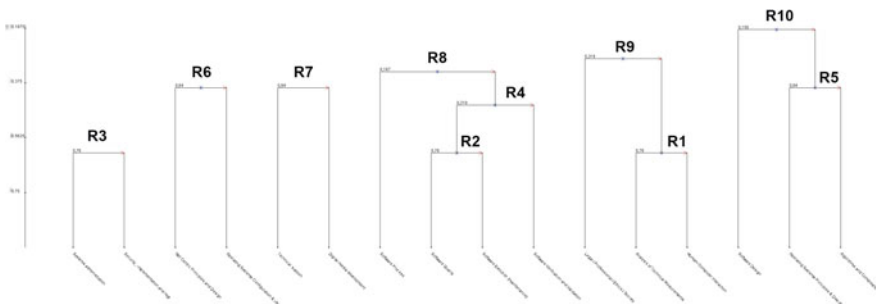


Fig. 1 Output as meta-rules on computing curricula

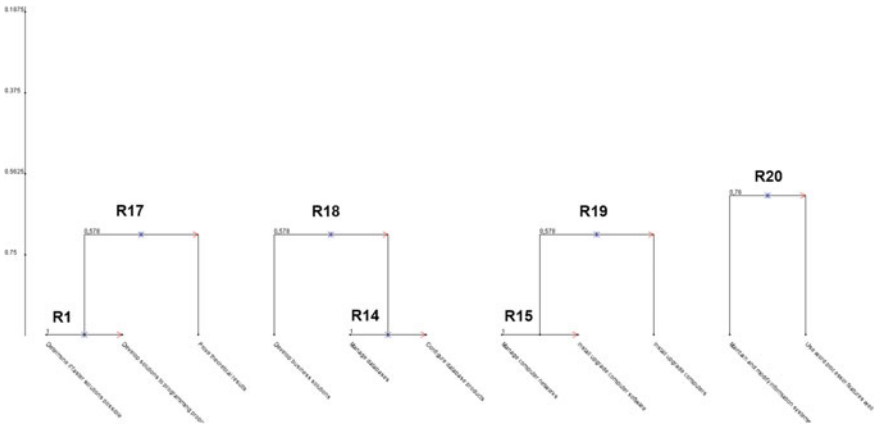


Fig. 2 Output as meta-rules for students abilities

- **R(10)**: Those who are able to *Design a Computer* usually arrive to *Program* and *Design a chip*, to *Design complex sensor systems*, *Computer Peripherals* and *embedded systems*;
- **R(13)**: those who are able to *Model and design a database* typically can *Training and supporting database users*, *Developing computer resource plan* and *Scheduling budget resource upgrades*;
- **R(14)**: those who are able to *Manage Databases* are usually able to *Configure database products*;
- **R(15)** : those who can *Manage computer networks* are generally able to *Install / Upgrade computer software*;
- **R(16)** : those who have the competence under rule *R(13)* are generally capable of assuring the rule *R(5)*;
- **R(17)**: to arrive at to *Prove theoretical results*, it was necessary to respect the rule *R(1)*;
- **R(18)**: those who can develop business solutions are generally complying with rule *R(14)*;
- **R(19)**: those who respect the rule *R(15)* are generally able to *Install / Upgrade computers*;
- **R(20)**: those who can *Maintain and modify information systems* generally manage to *Use word processor features well*;
- **R(21)**: if we are able to *Configure and integrate e-learning systems*, we are also able to *Use the spreadsheet features well*;
- **R(22)**: those who are able to *Implement communication software* usually have no problem to *Do systems programming*;
- **R(23)**: those who are able to *Implement intelligent systems* are usually able to *Design auto reasoning systems*;
- **R(24)**: those who *Develop multimedia solutions* usually manage to *Configure and integrate e-commerce software*.

4 Conclusion

This work aimed to verify the effectiveness of our HCIC- M_{GK} tool in order to be able to analyze the computing curricula and the students abilities who pursue the computer course in the Anglo-Saxon model. The result of the analysis allows us to help scientists who are mainly studying computer science curriculum so that they can detect anomalies in the curriculum as well as to adapt and update the abilities of young graduates required by companies.

References

1. A. Totohasina, D. Feno, De la qualité des règles d'association: étude comparative des meures MGK et Confiance Actes du 9ème colloque Africain sur la recherche en Informatique et Mathématiques Appliquées, CARI-2008, pp. 561–568
2. R. Gras, J.-C. Régnier, C. Marinica, F. Guillet, L'Analyse Statistique Implicative- Méthode exploratoire et confirmatoire à la recherche de causalités, Cepaduès; édition : 2e édition revue et augmentée (2013). ISBN-13: 978-2364930568
3. H.F. Rakotomalala, A. Totohasina, J. Diatta, Extraction des règles d'associations M_{gk}-valides avec contribution de Support, Actes des 24èmes rencontres de la Société Francophone de Classification SFC 2017, Lyon, France, 2017, pp. 29–32
4. H.F. Rakotomalala, A. Totohasina, J. Diatta, Une mesure de cohésion basée sur la mesure de qualité des règles d'association M_{gk}, Actes des 24èmes rencontres de la Société Francophone de Classification SFC 2017, Lyon, France, 2017, pp. 21–24
5. H.F. Rakotomalala, A. Totohasina, An efficient new cohesion indice based on the quality measure of association rules M_{gk}, WorldS4 2018, in *2nd World Conference on Smart Trends in System* (Security & Sustainability, IEEE-UK, London, 2018)
6. H.F. Rakotomalala, B. Ralahady, A. Totohasina, A novel cohesitive implicative classification based on m_{gk} and application on diagnostic on informatics literacy of students of higher education in madagascar, in *3rd International Conference ICICT 2018-International Congress & Excellence Awards, London 2018. Advances in Intelligent Systems and Computing*, vol. 797 (Springer, 2018), pp. 161–174
7. R. Shackelford, J. Cross, G. Davies, J. Impagliazzo, R. Kamali, R. LeBlanc, B. Lunt, A. McGettrick, R. Sloan, H Topi The Overview Report covering undergraduate degree programs, in *CE-CS-IS-IT-SE, CC, 2005* (New York, 2005). ISBN 1-59593-359-X
8. H.F. Rakotomalala, A. Totohasina, J. Diatta, Classification des mesures des règles d'association selon CHIC-M_{gk}, Actes des 25èmes rencontres de la Société Francophone de Classification SFC 2018 (Paris Descartes, France, 2018)

A Modeling Environment for Dynamic and Adaptive Network Models Implemented in MATLAB



S. Sahand Mohammadi Ziabari and Jan Treur

Abstract In this paper, a software environment to support Network-Oriented Modeling is presented. The environment has been implemented in MATLAB. This code covers the principles of temporal-causal network models. The software environment has built-in options for network adaptation principles such as the Hebbian learning principle from neuroscience and the adaptation principle for bonding based on homophily from social science. The implementation is illustrated for an adaptive temporal-causal network model under acute stress for decision-making.

Keywords Network-oriented modeling · Temporal-causal network · Adaptive · Software environment · Hebbian learning · Bonding by homophily · MATLAB

1 Introduction

In this paper, a dedicated software environment to support Network-Oriented Modeling is presented. The Network-Oriented Modeling approach addressed uses temporal-causal network models. This means that any scientific field in which causal relations are used to explain hypotheses, findings, and theories can be used in Network-Oriented Modeling [1]. Such domains vary from mental processes in individuals to social processes. For example, the interactions among individuals can be modeled as a network taking into account a network adaption principle like bonding based on homophily principle [2, 3]. Individual mental processes can be modeled as an interaction between mental states taking into account a network adaption principle based on Hebbian learning [4]. The latter represents the notion of plasticity described in Neuroscience which means that the communications within the brain are often adaptive and change over time.

S. S. Mohammadi Ziabari (✉) · J. Treur
Behavioural Informatics Group, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
e-mail: sahandmohammadiziabari@gmail.com

J. Treur
e-mail: j.treur@vu.nl

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_8

There are two different representations of a temporal-causal network model named a conceptual representation (labeled graph or matrix representation) and a numerical representation (representation by difference or differential equations). Using the software environment presented here, a conceptual representation can be used as a basis. By the software, it is automatically translated into a numerical representation, which can be used for numerical simulation, mathematical analysis, validation by comparing to empirical data or properties, and tuning of parameters to characteristics of domain, person, or social context.

The example model presented in [5] has been used as an illustration. This model incorporates adaptation principles based on Hebbian learning and on suppression of connections due to acute stress. There are other implementations for other types of network-oriented modeling, some of which can be found in [6–11].

The sections of paper are as follows. In Sect. 2, the Network-Oriented Modeling approach based on temporal-causal networks is briefly described. In Sect. 3, modeling a temporal-causal network in **MATLAB** is introduced, and in Sect. 4 an Illustration for an example network model has been described. Finally, Sect. 5 is the discussion section.

2 The Network-Oriented Modeling Approach Addressed

This software environment covers the principles of Network-Oriented Modeling based on temporal-causal networks discussed in the book [12]. The Network-Oriented Modeling format used is based on a dynamic and adaptive variant of modeling, reasoning, and simulation in a causal way which is a topic with a long history in artificial intelligence [13]. In this respect, any scientific field of study in which causal relations are applied can be addressed on the basis of this Network-Oriented Modeling approach. Among the wide variety of application areas, there are two types of applications that in a sense are dominant: describing individual mental processes specifically and describing how individuals interact with each other [1]. Table 1 shows the overview of some combination functions. The following three notions are central

Table 1 Overview of some combination functions $c(V_1, \dots, V_k)$

Name	Description	Formula $c(V_1, \dots, V_k) =$
sum(...)	Sum	$V_1 + \dots + V_k$
ssum$_{\lambda}$(...)	Scaled sum function	$\frac{V_1 + \dots + V_k}{\lambda}$ with $\lambda > 0$
min(...) max(...)	Minimal value Maximal value	Min(V_1, \dots, V_k) Max(V_1, \dots, V_k)
slogistic$_{\sigma, \tau}$(...)	Simple logistic sum function	$\frac{1}{1 + e^{-\sigma(V_1 + \dots + V_k - \tau)}}$ with $\sigma, \tau \geq 0$
alogistic$_{\sigma, \tau}$(...)	Advanced logistic sum function	$\left[\frac{1}{1 + e^{-\sigma(V_1 + \dots + V_k - \tau)}} - \frac{1}{1 + e^{-\sigma \tau}} \right]$ $(1 + e^{\sigma \tau})$ with $\sigma, \tau \geq 0$

in the Network-Oriented Modeling approach and define a temporal-causal network model, and therefore are part of a conceptual representation of a temporal-causal network model [14]:

- **Connections strength** $\omega_{X,Y}$

The connection strength between a state X to a state Y is called weight value $\omega_{X,Y}$ which is normally between 0 and 1.

- **Aggregation of impacts of states** $\mathbf{c}_Y(\cdot)$

Each state needs a combination function $\mathbf{c}_Y(\cdot)$ to aggregate the impacts of other states on state Y .

- **Speed of change of a state** η_Y

There is a speed factor η_Y shows how fast a state changes over a period of time based on the impact.

A conceptual representation of a temporal-causal network model can be transformed in a systematic or automated manner into a numerical representation of the model as follows [1, 12]:

- $Y(t)$ represents the value of Y at time point t in the model which is in the interval $[0, 1]$.
- **impact** $_{X,Y}(t) = \omega_{X,Y}X(t)$ shows the influence of a state X connected to a state Y at time point t where $\omega_{X,Y}$ represents the weight of the connection.
- The *aggregated impact* of some states X_i on Y at t is calculated using a *combination function* $\mathbf{c}_Y(\cdot)$:

$$\begin{aligned} \mathbf{aggimpact}_Y(t) &= \mathbf{c}_Y(\mathbf{impact}_{X_1,Y}(t), \dots, \mathbf{impact}_{X_k,Y}(t)) \\ &= \mathbf{c}_Y(\omega_{X_1,Y}X_1(t), \dots, \omega_{X_k,Y}X_k(t)) \end{aligned}$$

- The impact of $\mathbf{aggimpact}_Y(t)$ on Y is applied over time gently, based on speed factor η_Y :

$$Y(t + \Delta t) = Y(t) + \eta_Y[\mathbf{aggimpact}_Y(t) - Y(t)]\Delta t$$

or

$$\mathbf{d}Y(t)/\mathbf{d}t = \eta_Y[\mathbf{aggimpact}_Y(t) - Y(t)]$$

- Therefore, the *difference* and *differential equations* for Y are achieved:

$$\begin{aligned} Y(t + \Delta t) &= Y(t) + \eta_Y[\mathbf{c}_Y(\omega_{X_1,Y}X_1(t), \dots, \omega_{X_k,Y}X_k(t)) - Y(t)]\Delta t \\ \mathbf{d}Y(t)/\mathbf{d}t &= \eta_Y[\mathbf{c}_Y(\omega_{X_1,Y}X_1(t), \dots, \omega_{X_k,Y}X_k(t)) - Y(t)] \end{aligned}$$

Adaptation principles covered

The following adaptation principles are covered.

Hebbian learning

For *Hebbian learning* of a connection from state X_i to state X_j , the following model is used

$$\omega(t + \Delta t) = \omega(t) + \eta_\omega[\mathbf{c}_\omega(X_i(t), X_j(t), \omega(t)) - \omega(t)]\Delta t$$

with

$$\mathbf{c}_\omega(V_1, V_2, W) = \mathbf{hebb}_\mu(V_1, V_2, W) = V_1 V_2(1-W) + \mu W$$

where μ is the persistence factor with 1 as full persistence.

State-connection modulation

For the adaptation principle for *state-connection modulation* with control state cs , the following model is used:

$$\omega(t + \Delta t) = \omega(t) + \eta_\omega[\mathbf{c}_\omega(cs_2(t), \omega(t)) - \omega(t)]\Delta t$$

with

$$\mathbf{c}_\omega(V, W) = \mathbf{scm}_\alpha(V, W) = W + \alpha V W(1-W)$$

where α is the adjustment parameter for ω from cs . In combination, these two adaptive combination functions can be used as a weighted average with $0 \leq \theta \leq 1$ as follows:

$$\mathbf{c}_\omega(V_1, V_2, V, W) = \theta \mathbf{hebb}_\mu(V_1, V_2, W) + (1 - \theta) \mathbf{scm}_\alpha(V, W)$$

$$\omega(t + \Delta t) = \omega(t) + \eta_\omega[\mathbf{c}_\omega(X_i(t), X_j(t), cs(t), \omega(t)) - \omega(t)] \Delta t$$

All these difference equations can be used for simulation.

This state-connection adaptation principle can also be applied in a social context. The hypothesis is based on that whenever a more intensive interplay between two persons occurs, the connection will become solid, e.g., [15].

Bonding based on homophily

Bonding based on homophily shows that the more look like the states of two connected states, the stronger their connection will become: ‘the more you are alike, the more you like (each other)’ [12]; see, for example, [2, 16, 17]. When also the states are assumed dynamic, this principle can be combined with contagion of states into a circular causal relation [18]: State \leftrightarrow Link. See also, for example [19–22]. The homophily principle can be as represented numerically by a combination function $\mathbf{c}_{A,B}(V_1, V_2, W)$ as follows:

$$\begin{aligned}\omega_{A,B}(t + \Delta t) &= \omega_{A,B}(t) + \eta_{A,B}[c_{A,B}(X_A(t), X_B(t), \omega_{A,B}(t)) - \omega_{A,B}(t)] \Delta t \\ dY(t)/dt &= \eta_{A,B}[c_{A,B}(X_A(t), X_B(t), \omega_{A,B}) - \omega_{A,B}]\end{aligned}$$

Three variants of models for the homophily axiom are the linear, quadratic, and logistic variants:

Linear

$$c(V_1, V_2, W) = \mathbf{slhomo}(V_1, V_2, W) = W + W(1 - W)(\tau - |V_1 - V_2|)$$

Quadratic

$$c(V_1, V_2, W) = \mathbf{sqhomo}(V_1, V_2, W) = W + W(1 - W)(\tau^2 - (V_1 - V_2)^2)$$

Logistic

$$\begin{aligned}c(V_1, V_2, W) &= \mathbf{sloghomo}(V_1, V_2, W) \\ &= W + W(1 - W)(0.5 - 1/(1 + e^{-\sigma(|V_1 - X_2| - \tau)}))\end{aligned}$$

Based on these options that can be chosen the following numerical differential and difference equations are generated

$$\begin{aligned}d\omega_{C,D}/dt &= \eta_{C,D} \omega_{C,D}(1 - \omega_{C,D})(\tau_{C,D} - |X_C - X_D|) \\ \omega_{C,D}(t + \Delta t) &= \omega_{C,D} + \eta_{C,D} \omega_{C,D}(t) (\tau_{C,D} - |X_C - X_D|) \Delta t\end{aligned}$$

$$\begin{aligned}d\omega_{C,D}/dt &= \eta_{C,D} \omega_{C,D}(1 - \omega_{C,D})(\tau_{C,D}^2 - |X_C - X_D|^2) \\ \omega_{C,D}(t + \Delta t) &= \omega_{C,D} + \eta_{C,D} \omega_{C,D}(t)(1 - \omega_{C,D})(\tau_{C,D}^2 - (X_A(t) - X_B(t))^2) \Delta t\end{aligned}$$

$$\begin{aligned}d\omega_{C,D}/dt &= \eta_{C,D} \omega_{C,D}(1 - \omega_{C,D})(0.5 - 1/(1 + e^{-\sigma(|X_C - X_D| - \tau_{C,D})})) \\ \omega_{C,D}(t + \Delta t) &= \omega_{C,D} + \eta_{A,B} \omega_{C,D}(t)(1 - \omega_{C,D})(0.5 - 1/(1 + e^{-\sigma(|X_C - X_D| - \tau_{C,D})})) \Delta t\end{aligned}$$

Here, X_C, X_D are the states of person C and D ; $\omega_{C,D}$ is the connection weight from person C to person D , $\eta_{C,D}$ the update speed factor for the connection from person C to person D , and $\tau_{C,D}$ the threshold or tipping point for connection adaption.

3 Modeling a Temporal-Causal Network in MATLAB

The advantage of using **MATLAB** for simulation is that (as the abbreviation of that says) ‘Matrix laboratory’ can easily work with matrices. The format is defined in Table 2, if there is a connection between X_1 and X_2 as states, the connection weight assigned to the matrix representation (w) of states between aforementioned states.

Table 2 Notions and MATLAB representations

Notions	MATLAB representations
Number of nodes (states)	N
The notion of states (X_1, X_2, \dots) and weights ω among states	W
The speed factors	Sp_f
Initial values	STDx
Combination function (identity)	First row of matrix 'O' id=O(1, :);
Combination function (sum)	Second row of matrix 'O' sum function
Combination function (scaled sum)	3rd and fourth rows of matrix 'O' Scaled sum, Scaling factor
Combination function (normalized sum)	5th and 6th rows of matrix 'O' normalised sum, normalizing factor
Combination function (adaptive normalized)	7th row of matrix 'O' adnorsum
Combination function (simple logistic)	8th, 10th and 11th rows of matrix 'O' slogistic(...)
Combination function (advance logistic)	9th and 10th and 11th rows of matrix 'O' Alogistic, steepness, threshold
Combination function (adaptive advanced logistic)	12th and 13th and 14th rows of matrix 'O' adaptive advanced logistic, steepness, steepness

(continued)

Table 2 (continued)

Notions	MATLAB representations
Speed factor for connection weight adaptation (used for Hebbian learning, homophily, and state-connection modulation)	eta
Hebbian learning principle	hebb, mu
Homophily principle (simple linear homophily)	slhomo, htau
Homophily principle (advanced linear homophily)	alhomo, htau
Homophily principle (simple quadratic homophily)	sqhomo, htau
Homophily principle (advanced quadratic homophily)	aqhomo, htau
State-connection modulation	Adcon, amp
Length of time for simulation	time=0:dt:398; L=length(time);
Δt	dt
Plotting (figures)	plot(time(1:230),STDX(1:230,i),'linewidth',3)
RMS (root mean square)	RMS = sqrt (nansum ((Output - emp_data) .^2)) / (col * row)

Due to simplicity of **MATLAB** and also providing many functions, the **MATLAB** software became one of the often-used software environments for engineers and computer science developers. It has been used in different aspects of sciences from image processing, due to providing many toolboxes, and also machine learning, analyze and simulates the behavioral dynamics of agents in cognitive science, social science, and in artificial intelligence.

The initialization of the matrices for doing actions, calculating based on the combination functions (identity, advance logistic, advance advanced logistic, scaled sum...), **MATLAB** representation of functions based on notations of states, relations, and all principles are shown in Table 4 in the appendix.

In Fig. 1, the functional view of the **MATLAB** code is shown. The process starts with the inputs named number of states (nodes), connection weights, speed factors, and initial values. In the next step, all parameters are allocated in matrices with $1 \times N$ dimensions, where N is the number of the states in the model.

The next phase is allocating the primitive values of the states in the first row of the matrix (STDX) and then specifying the time period for having simulation. If there is a Hebbian learning in the model, then parameters of Hebbian learning, η , hebb, and μ are allocated in three different matrices, similarly for the homophily principle, simple homophily (slhom), advanced homophily (alhom) simple quadratic homophily (sqhom) and advanced quadratic homophily (qhom), threshold (hthau), and finally for state-connection modulation (scm_a) and the number of the states which have modulation ability. Figure 1 illustrates the functional view of written **MATLAB** code.

Then in the next step, if there is any above-mentioned principle in the model they will multiply by the STDX matrix formed already with time. Meanwhile, the matrix called condy with the weights of states formed in a column-wise order and then make a new row based on any principle existed and then add the influence of them in the matrix. Finally, all matrices with STDX and condy (if there was any state-connection modulation) result in generating the simulation.

The human interaction flowchart is depicted in Fig. 2. As it can be seen, a first step is initialization as providing inputs, for instance, number of states, weights among states, speed factors, and combination functions. In the second step, it is needed to be decided in which system the user wants to work. In this phase, there are two systems, multi-agent system and single-agent system. The former offers simple linear, advanced linear, simple quadratic, advanced quadratic homophily principles, and for latter, there are Hebbian learning and state-connection suppression, and finally, the plotting of simulation occurs.

And for more specification, the flowchart of the code is presented in Fig. 3.

To enable easy use, one can also use a user interface between Excel and **MATLAB**. As **MATLAB** works with matrices it might be easier to use an Excel interface of matrices and just read the matrices from the Excel file and then do the execution in **MATLAB**. Such a matrix expresses the parameters, combination functions, identity function, sum function, scaled sum with scale factor, normalized with normalizing factor, adaptive normalized sum, simple logistic, advanced logistic, advanced logistic and adaptive advanced logistic function with steepness, and threshold and other

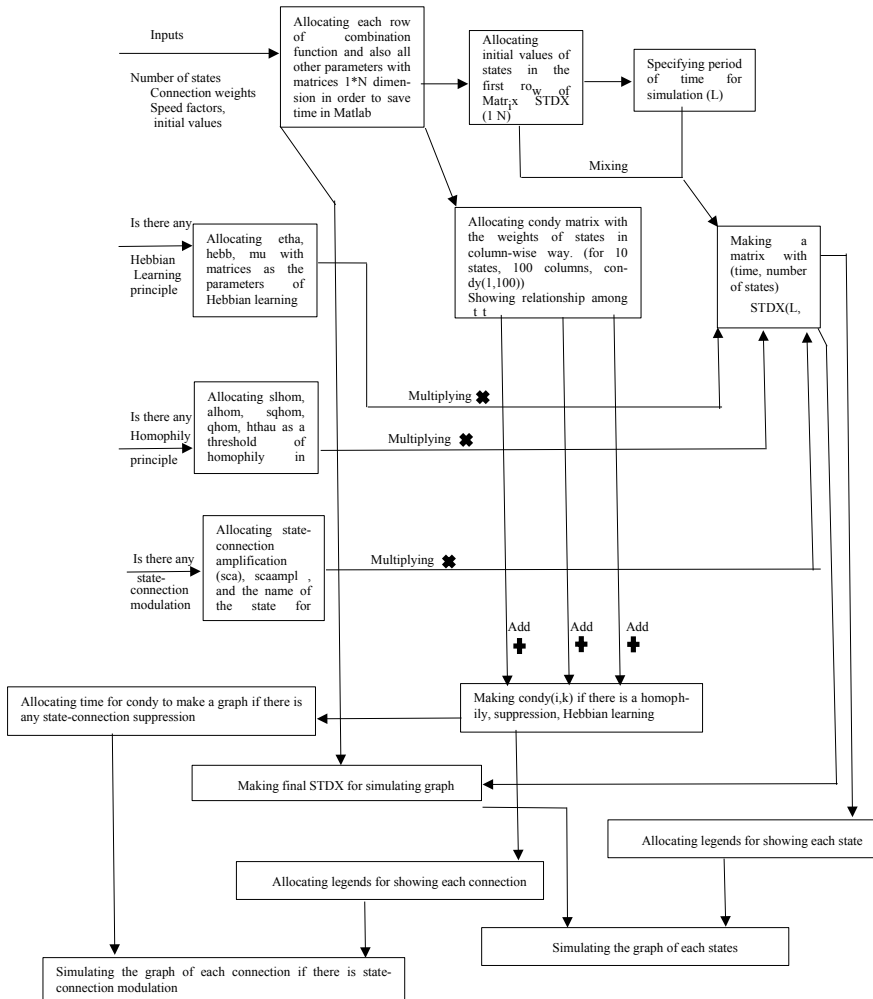


Fig. 1 Functional view of the MATLAB process

principles. As can be seen from Figs. 4 and 5, the MATLAB code reads the matrices based on the matrix representation in Excel; for instance, in these figures, it would be from column C1 to L32 to read matrix for all weights of states, speed factors, deltaT, maxt, combination functions, and finally initial values from the first sheet and if there is any principle combined with the model using the second sheet to read from Excel for Hebbain learning principle and homophily for their principles. Figures 4 and 5 show this option in Excel sheets.

Parameter tuning using the sum of squared residuals and root mean square

For comparison between empirical data and simulation results and optimization of

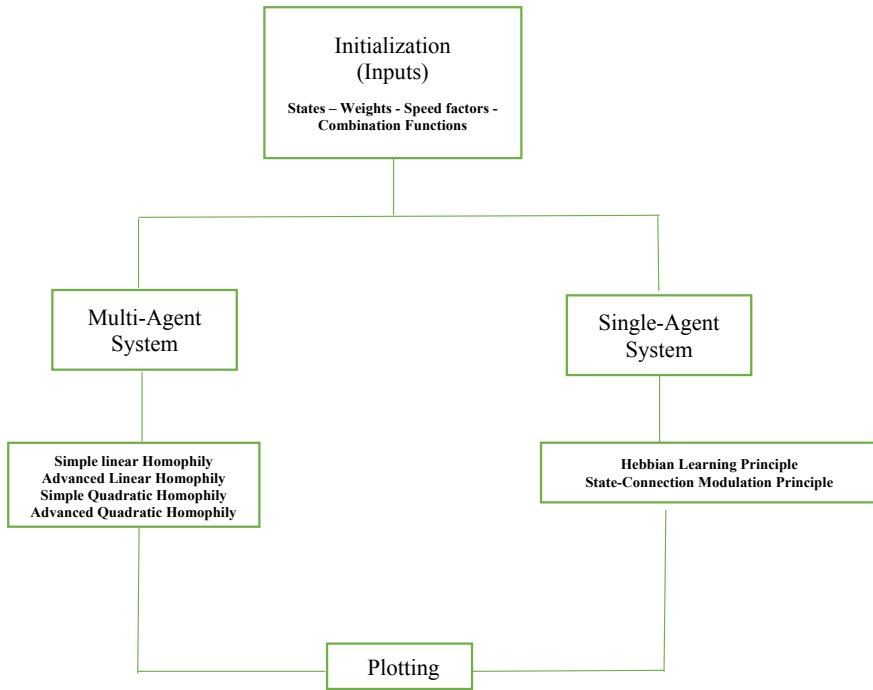


Fig. 2 Human interaction flowchart

parameters, MATLAB components are available; the sum of squared residuals (SSR) has been implemented to calculate the difference.

$$SSR = ((X(t_1) - Y(t_1))^2 + \dots + (X(t_N) - Y(t_N))^2)$$

$$RMS = \sqrt{\frac{SSR}{N}} = \sqrt{\frac{(X(t_1) - Y(t_1))^2 + \dots + (X(t_N) - Y(t_N))^2}{N}}$$

```

% loading empirical data
load('Data1.mat', 'Data1');
[row, col]=size(Data1);

% Calculating Root Mean Square
RMS = sqrt (nansum ((Output - emp_data).^2) / (col * row)
  
```

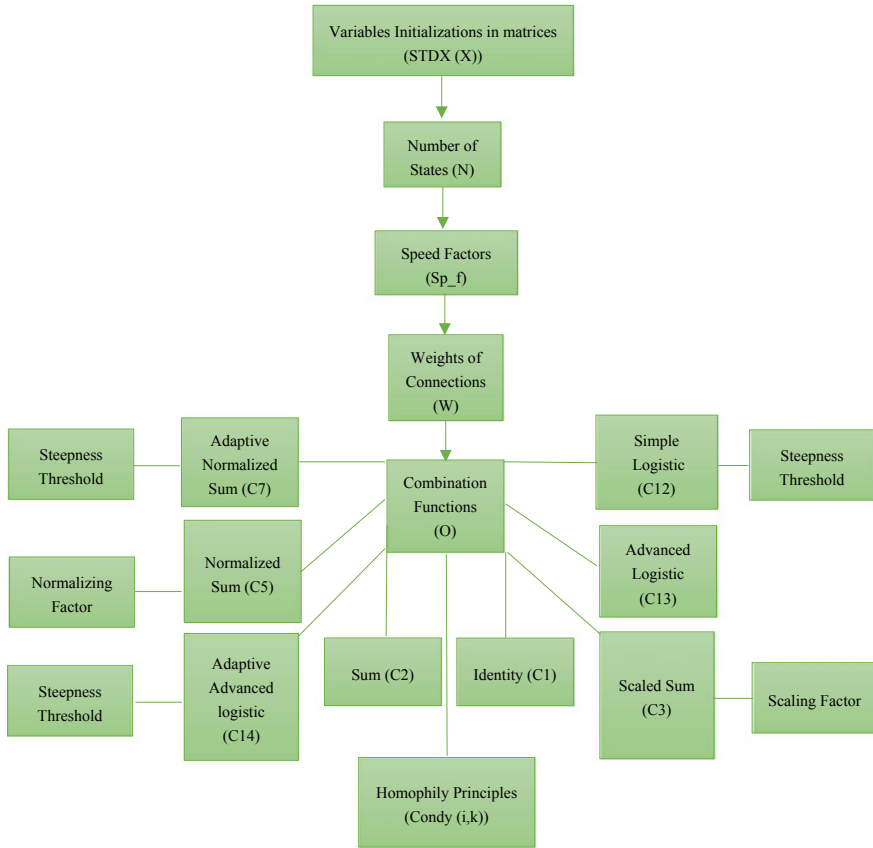


Fig. 3 Structural view of the code

4 Illustration for an Example Network Model

It is shown how the model presented in [5] can be executed in MATLAB. The conceptual representation of temporal-causal network of the model used in [5] is illustrated in Fig. 5, and the explanation of states has been shown in Table 3. As can be seen from Fig. 5, here, both the Hebbian learning principle and state-connection modulation were used. The Hebbian learning principle used between states srs_s and (ps_{a1} and ps_{a2}) and also state-connection suppression between state cs_2 and the connections with Hebbian learning principle. The number of states considered to be 10. Figure 7 shows simulation result of weights of states and Fig. 8 shows state-connection suppression (Fig. 6).

In Fig. 6 shows temporal-causal network model. An overview of explanation of the states is illustrated in Table 3.

	A	B	C	D	E	F	G	H	I	J	K	L
1			0.4									300
2		delta τ									max τ	
3		states and connections	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$	$X10$
4			SRSs	SRSs	SRSs	SRSs	FSee	PSa1	PSa2	PSee	CS2	CS1
5	$X1$	SRSs		1	-0.1	0.3		0.9	0.3			
6	$X2$	SRSs								1		
7	$X3$	SRSs						0.7				
8	$X4$	SRSs							0.7			
9	$X5$	FSee								1	1	
10	$X6$	PSa1			0.7							
11	$X7$	PSa2				0.7						
12	$X8$	PSee					1					
13	$X9$	CS1									-0.9	
14	$X10$	CS2										1
15		speed factors										
16			0	0.05	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.02
17		combination functions										
18	identity function	id(...)					1					1
19	sum function	sum(...)										
20	scaled sum	ssum(...)			1	1		1	1	1	1	
21	normalised sum	norsum(...)										
22	adaptive normalised sum	adnorsum(...)										
23	simple logistic	slogistic(...)										
24	advanced logistic	alogistic(...)										
25		steepness σ										
26		threshold τ										
27	adaptive advanced logistic	adalogistic(...)		1								
28		steepness σ		18								
29		threshold factor τ		0.2								
30		initial values										
31			1	0.1								
32												

Fig. 4 Excel interface to read for MATLAB programming (parameters, speed factor, combination function, and initial values)

	A	B	C	D	E	F	G	H	I
1		from	ws_1	ws_2	ws_3	ws_4	ws_5	ws_6	ws_7
2		to	ws_1	ss_1	srs_1	ps_1	srs_2	es_1	ws_8
3			$\Omega_{X1,X1}$	$\Omega_{X1,X2}$	$\Omega_{X1,X3}$	$\Omega_{X1,X4}$	$\Omega_{X1,X5}$	$\Omega_{X1,X6}$	$\Omega_{X1,X7}$
4		speed factor η							0.5
5	hebbian learning	hebb(...)							0.85
6		persistence μ							0.8
7	simple linear homophily	slhom $_{\tau, \sigma}$ (...)							
8	advanced linear homophily	alhom $_{\tau, \sigma}$ (...)							
9	simple quadratic homophily	sqhom $_{\tau, \sigma}$ (...)							
10	advanced quadratic homophily	aqhom $_{\tau, \sigma}$ (...)							
11		threshold τ							
12		amplification α							
13	state-connection amplification	sca $_{\sigma}$ (...)							0.15
14		scaamp α							0.5
15	i	X_1							
16	m	X_2							
17	p	X_3							
18	a	X_4							
19	c	X_5							
20	t	X_6							
21	f	X_7							
22	r	X_8							
23	o	X_9							-0.7
24	m	X_{10}							
25									

Fig. 5 Excel interface to read for MATLAB programming (Hebbian, homophily, and suppression principles)

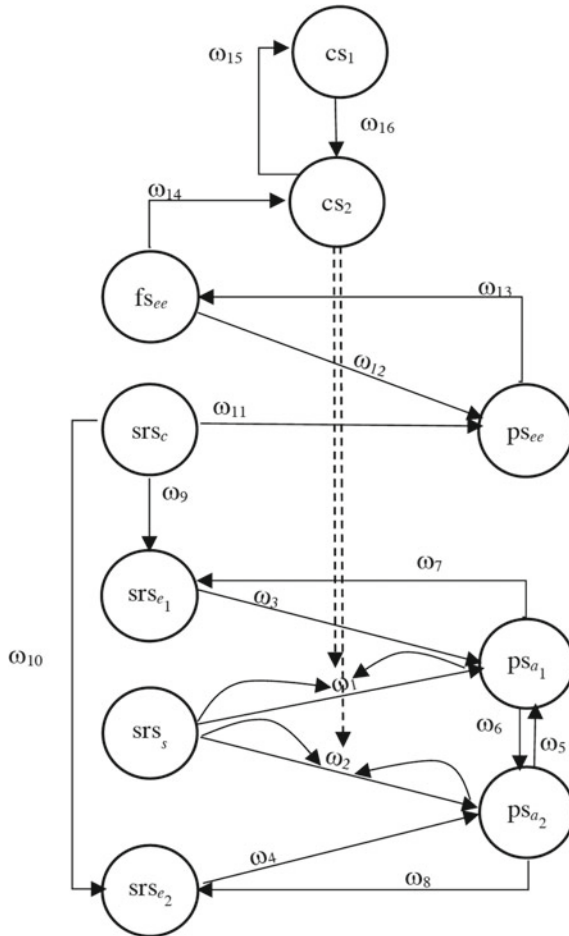


Fig. 6 Adaptive temporal-causal network model's conceptual representation [5]

Table 3 States explanations in the model [5]

X_1	Sensory representation of stimulus s
X_2	Sensory representation of context c
X_3	Sensory representation of action effect e_1
X_4	Sensory representation of action effect e_2
X_5	Feeling state for extreme emotion ee
X_6	Preparation state for action a_1
X_7	Preparation state for action a_2
X_8	Preparation state for response of extreme emotion ee
X_9	Control state for timing of suppression of connections
X_{10}	Control state for suppression of connections

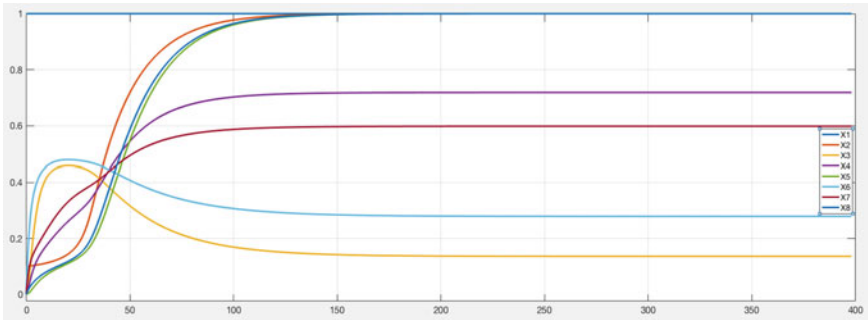


Fig. 7 Simulation outcome of presented model: states

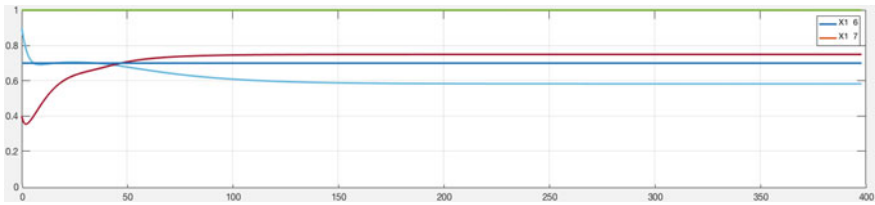


Fig. 8 Simulation outcome for suppression and Hebbian learning for ω_1 (connection X_1-X_6) and ω_2 (connection X_1-X_7)

```

% SRSs  SRSc  SRSe1  SRSe2  FSee  PSa1  PSa2  PSee  CS1  CS2
% X1  X2  X3  X4  X5  X6  X7  X8  X9  X10
W=[ 0  0  0  0  0  0.9  0.4  0  0  0  %X1  SRSs
    0  1  -0.1  0.3  0  0  0  1  0  0  %X2  SRSc
    0  0  0  0  0  0.7  0  0  0  0  %X3  SRSe1
    0  0  0  0  0  0  0.7  0  0  0  %X4  SRSe2
    0  0  0  0  0  0  0  1  0  0  %X5  PSee
    0  0  0.7  0  0  0  -0.2  0  0  0  %X6  PSa1
    0  0  0  0.7  0  -0.2  0  0  0  0  %X7  PSa2
    0  0  0  0  1  0  0  0  0  0  %X8  PSee
    0  0  0  0  0  0  0  0  0  0  %X9  CS1
    0  0  0  0  0  0  0  0  0  -0.9  %X10  CS2
];
    
```

```

Sp_f=[ 0    0.05   0.5   0.5   0.5   0.5   0.5   0.5   0.4   0.02   0.6];
O=[  0    0    0    0    1    0    0    0    0    1    0 % identity function id(.)
    0    0    0    0    0    0    0    0    0    0    0 % sum function sum (...)
    0    0    1    1    0    1    1    1    1    0    1 % Scaled sum sum (...)
    0    0    0.7  1    0    2    2    2    2    0    1 % Scaling factor
    0    0    0    0    0    0    0    0    0    0    0 % normalised sun norsum(...)
    0    0    0    0    0    0    0    0    0    0    0 % normalizing factor
    0    0    0    0    0    0    0    0    0    0    0 % adnorsum
    0    0    0    0    0    0    0    0    0    0    0 % slogistic(...)
    0    0    0    0    0    0    0    0    0    0    0 % alogistic(...)
    0    0    0    0    0    0    0    0    0    0    0 % steepness
    0    0    0    0    0    0    0    0    0    0    0 % threshold
    0    1    0    0    0    0    0    0    0    0    0 % adaptive advanced logistic (...)
    0    18   0    0    0    0    0    0    0    0    0 % steepness
    0    0.2  0    0    0    0    0    0    0    0    0 % threshold factor

% Suppression among connections in Hebbian learning
adcon(13,6)=0.15;
adcon(13,7)=0.15;
adcon(14,6)=0.5;
adcon(14,7)=0.5;
adcon(24,6)=-0.7;
adcon(24,7)=-0.7;

% Hebbian learning among states 1,6
eta(1,6)=0.5;
eta(1,7)=0.8;

hebb(1,6)=0.85;
hebb(1,7)=0.85;

mu(1,6)=0.8;
mu(1,7)=0.8;

% Assign time for plotting
dt=0.25;
time=0:dt:398;
L=length(time);
STDx=zeros(L,N);

% Assign the initialization for states 1 and 2 (can be for any states)
STDx(1,1)=1;
STDx(1,2)=0.1;

```

Figure 9 shows a difference between simulation result and empirical data provided. And the result is 0.01309.

5 Discussion

The implementation of a dedicated MATLAB-based software environment for Network-Oriented Modeling has been described. The modeling approach covered can be found in [12]; see also [1]. This implementation has been used in dynamic and adaptive network-oriented modeling. The environment was illustrated for the example model described in [5] for which previously only an Excel-based model was available.

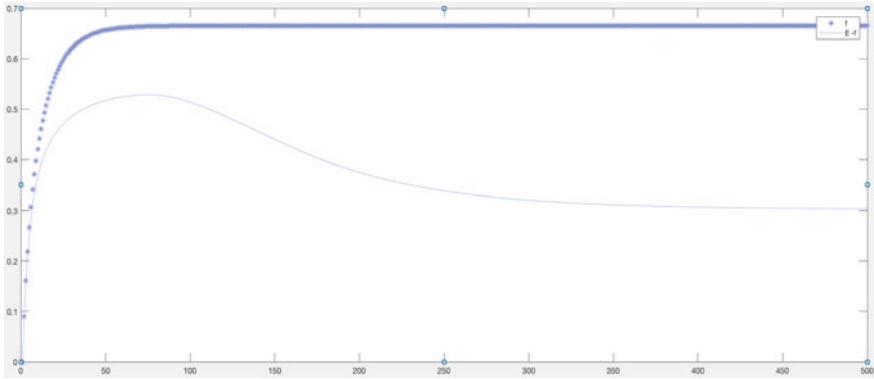


Fig. 9 Simulation result for empirical data (X_1) and simulation result (X_2)

An important advantage of the software environment is that modeling can take place at the level of conceptual representations expressed as labeled graphs or matrices. Therefore, it is suitable in a multidisciplinary context where different disciplines play a role, also disciplines where technical knowledge from computer science or AI is minimal. The more technical numerical representations and the actual execution are taken care of by the software environment, and therefore for users, no programming skills are needed.

Acknowledgements We would like to thank our colleague Fakhra Jabeen, Ph.D. candidate at Vrije Universiteit Amsterdam, for her assistance with making possible to have an Excel interface with current MATLAB code.

Appendix: Inputs Description of the MATLAB Code

```
% Initializing the matrices with zero in order to get less time in operating the main codes
id=zeros(1,N);
sum=zeros(1,N);
ssum=zeros(1,N);
lambda=zeros(1,N);
norsum=zeros(1,N);
norlambda=zeros(1,N);
adnorsum=zeros(1,N);
slog=zeros(1,N);
alog=zeros(1,N);
s=zeros(1,N);
t=zeros(1,N);
adalog=zeros(1,N);
```

```
adas=zeros(1,N);
adat=zeros(1,N);
id=0(1,:);
sum=0(2,:);
ssum=0(3,:);
lambda=0(4,:);
norsum=0(5,:);
norlambd=0(6,:);
adnorsum=0(7,:);
slog=0(8,:);
alog=0(9,:);
s=0(10,:);
t=0(11,:);
adalog=0(12,:);
adas=0(13,:);
adat=0(14,:);
eta=zeros(N);
hebb=zeros(N);
mu=zeros(N);
slhomo=zeros(N);
alhomo=zeros(N);
sqhomo=zeros(N);
aqhomo=zeros(N);
htau=zeros(N);
amp=zeros(N);
adcon=zeros(14+N, N^2);

clc
clear
close all
format long
N=10;
```

See Table 4.

Table 4 Notions of states and MATLAB representations of functions

Notions of states	MATLAB representations
Combination function (identity)	$C1 = \text{condy}(i-1, k);$
Combination function (sum)	$C3 = \text{htau}(ii, jj) - \text{abs}(\text{STDx}(i-1, ii) - \text{STDx}(i-1, jj));$
Combination function (scaled sum)	$C8 = \text{ssum}(j) * C7 / \text{lambda}(j);$
Combination function (normalized sum)	$C9 = \text{norsum}(j) * C7 / \text{norlambda}$
Combination function (adaptive normalized)	$C7 = C7 + \text{condy}(i-1, (jj-1) * N + j) * \text{STDx}(i-1, jj);$
Combination function (simple logistic)	$C12 = \text{slog}(j) * (1 / (1 + \exp(-s(j) * (C7 - t(j)))));$
Combination function (advance logistic)	$C13 = \text{alog}(j) * ((1 / (1 + \exp(-s(j) * (C7 - t(j)))))) - (1 / (1 + \exp(s(j) * t(j)))) * (1 + \exp(-s(j) * t(j))));$
Combination function (adaptive advance logistic)	$C14 = \text{adalog}(j) * ((1 / (1 + \exp(-\text{adas}(j) * (C7 - \text{adat}(j) * C11)))) - \text{condy}(i-1, k) + \text{eta}(ii, jj) * (\text{hebb}(ii, jj) * (\text{STDx}(i-1, ii) * \text{STDx}(i-1, jj) * C2 + \mu(ii, jj) * C1))$
Hebbian learning principle	$\text{slhomo}(ii, jj) * (C1 + \text{amp}(ii, jj) * C1 * C2 * C3)$
Homophily principles (simple linear homophily)	$\text{alhmo}(ii, jj) * (C1 + \text{amp}(ii, jj) * C2 * ((C4 + C3) / 2) + C1 * ((C4 - C3) / 2))$
Homophily principles (advanced linear homophily)	$\text{sqhomo}(ii, jj) * (C1 + \text{amp}(ii, jj) * C1 * C2 * C5)$
Homophily principles (simple quadratic homophily)	$\text{aqhmo}(ii, jj) * (C1 * \text{amp}(ii, jj) * C2 * ((C6 + C5) / 2) + C1 * ((C6 - C5) / 2)) - C1 * dt;$
Homophily principles (advanced quadratic homophily)	$\text{adcon}(13, k) * (C1 + \text{adcon}(14, k) * (CC) * (C2) * C1)$

(continued)

Table 4 (continued)

Notions of states	MATLAB representations
All values for states	<pre> STDX(i,j)=STDX(i-1,j)+Sp_f(j)*(aggimpact(i-1,j)-STDX(i-1,j))*dt; condy(i,k)=condy(i-1,k)+eta(ii,jj)*(hebb(ii,jj))* (... STDX(i-1,ii)*STDX(i-1,jj)*C2+mu(ii,jj)*C1)+... adcon(13,k)*(C1+adcon(14,k)*(CC)*(C2))*... C1)+slhomo(ii,jj)*(C1+amp(ii,jj)*C1*C2*... C3)+alhomo(ii,jj)*(C1+amp(ii,jj)*C2*((C4+C3)/2)+C1*... ((C4-C3)/2))+sqhomo(ii,jj)*(C1+amp(ii,jj)*C1*C2*C5)+... aqhomo(ii,jj)*(C1*amp(ii,jj)*C2*((C6+C5)/2)+C1*((C6-C5)/2))-... C1)*dt;</pre>
RMS (root mean square)	<pre> % loading empirical data load('Data1.mat','Data1'); [row, col]=size(Data1); % calculating root mean square RMS = sqrt (nansum ((Output - emp_data).^2)) / (col * row)</pre>

References

1. J. Treur, The Ins and Outs of Network-Oriented Modeling: from biological networks and mental networks to social networks and beyond. *Transactions on Computational Collective Intelligence*. Paper for Keynote Lecture at ICCCI'18 (2018)
2. M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather homophily in social networks. *Ann. Rev. Sociol.* **27**, 415–444 (2001)
3. A.L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999)
4. D. Hebb, *The Organization of Behavior* (Wiley, 1949)
5. J. Treur, S.S.M. Ziabari, An adaptive temporal-causal network model for decision making under acute stress, in *Proceedings of the 10th International Conference on Computational Collective Intelligence, ICCCI'18*, vol. 2, ed. by N.T. Nguyen. Lecture Notes in Computer Science, vol. 11056 (Springer, Berlin, 2018), pp. 13–25
6. S.S.M. Ziabari, J. Treur, Cognitive modelling of mindfulness therapy by autogenic training, in *Proceedings of the 5th International Conference on Information System Design and Intelligent Applications, INDIA'18*. Advances in Intelligent Systems and Computing (Springer, Berlin, 2018)
7. S.S.M. Ziabari, J. Treur, Integrative Biological, Cognitive and affective modeling of a drug-therapy for a post-traumatic stress disorder, in *Proceedings of the 7th International Conference on Theory and Practice of Natural Computing, TPNC'18* (Springer, Berlin, 2018)
8. S.S.M. Ziabari, J. Treur, Computational analysis of gender differences in coping with extreme stressful emotions, in *Proceedings of the 9th International Conference on Biologically Inspired Cognitive Architecture (BICA2018)* (Elsevier, Czech Republic, 2018)
9. S.S.M. Ziabari, Integrative cognitive and affective modeling of deep brain stimulation, in *Proceedings of the 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2019)* (submitted for publication)
10. S.S.M. Ziabari, J. Treur, An adaptive cognitive temporal-causal network model of a mindfulness therapy based on humor, in *International Conference on Computational Science (ICCS 2019)* (submitted for publication)
11. S.S.M. Ziabari, J. Treur, An adaptive cognitive temporal-causal network model of a mindfulness therapy based on music, in *Proceedings of the 10th International Conference on Intelligent Human Computer Interaction (IHCI2018)* (Springer, India, 2018)
12. J. Treur, *Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions* (Springer, 2016)
13. B.J. Kuipers, J.P. Kassier, How to discover a knowledge representation for causal reasoning by studying an expert physician, in *Proceedings English International Joint Conference on Artificial Intelligence, IJCAI'83*, ed. by F.R.G. Karlsruhe (William Kaufman, Los Altos, CA, 1983)
14. J. Treur, Verification of temporal-causal network models by mathematical analysis. *Vietnam J. Comput. Sci.* **3**, 207–221 (2016)
15. E. Pearce, J. Launay, R.I.M. Dunbar, The ice-breaker effect: singing together mediates fast social bonding. *R. Soc. Open Sci.* (2015). <https://doi.org/10.1098/rsos.150221>
16. D. Byrne, The attraction hypothesis: do similar attitudes affect anything? *J. Pers. Soc. Psychol.* **51**(6), 1167–1170 (1986)
17. A. Mislove, B. Viswanath, K.P. Gummadi, P. Druschel, You are who you know: inferring user profiles in online social networks, in *Proceedings of the WSDM'10*, New York City, New York, USA, 4–6 Feb 2010 (2010), pp. 251–260
18. M.E.J. Newman, The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
19. S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from Homophily driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* **106**(2), 1544–1549 (2009)

20. C.R. Shalizi, A.C. Thomas, Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* **40**(2), 211–239 (2011)
21. C.E.G. Steglich, T.A.B. Snijders, M. Pearson, Dynamic networks and behavior: separating selection from influence. *Sociol. Methodol.* **40**, 329–393 (2010)
22. M.P. Mundt, L. Mercken, L.I. Zakletskaia, Peer selection and influence effects on adolescent alcohol use: a stochastic actor-based model. *BMC Pediatr.* **12**, 115 (2012)

Face Authentication Using Image Signature Generated from Hyperspectral Inner Images



Guy Leshem and Menachem Domb

Abstract Face recognition technologies are commonly used in access control systems. It is done by extracting selected features from the face image, taken by a 2D camera. This technique lacks the case of a picture placed in front of the camera. The system will mistakenly recognize it as a real live person and so, allow the access of the picture holder, which may be an unauthorized person. A new generation of security systems uses a three-dimensional face recognition. Although it is better than 2D, it lacks a similar case, where a 3D image is generated from many 2D images. The system will assume it is a picture taken from a live person, and mistakenly, allow the access. We propose an enhancement to the existing authentication process given 2D face image. It is based on inner images extracted from a hyperspectral camera. These images represent inner layers of the person tissue structure, which in general are different from person to person and so, may be used to differentiate between two persons. We use these generated features to generate an authentication signature. The authentication signature is a composition of processed inner layers features. To prove that this signature is universally unique and can substitute the current use of 2D image recognition system, there is a need to conduct a comprehensive testing and apply other technologies to prove it. We are not at this stage. Therefore, at this stage, we propose adding to each image a unique signature generated from the corresponding hyperspectral inner layers. When a person is trying to access, the access control system, using a hyperspectral camera, captures its standard image features, and in addition, calculates the inner images to generate a relatively unique signature, and compares both elements to the identification table. Experiments show that this combination generates a relatively unique identification key. From the beginning of our initial experiments, it kept its attentiveness and uniqueness for all we tried to challenge it. Further experiments prove the significant contribution of inner features for strengthening the person authentication.

G. Leshem · M. Domb (✉)

Department of Computer Science, Ashqelon Academic College, Ashkelon, Israel
e-mail: dombmnc@edu.aac.ac.il

G. Leshem

e-mail: gjalsm@edu.aac.ac.il

Keywords Face recognition · Authentication · Hyperspectral image · Inner layers · Image verification algorithms · Access control

1 Introduction

Many data protection, security and access control systems, use various technologies, such as biometric fingerprints and face recognition. Current face recognition systems are based on photos taken by a standard camera. The naive face recognition algorithm uses one image to recognize a face. The problem with it is the ability to easily bypass it. Better algorithms use several images and build a three-dimensional model of the interior, but this approach can also be bypassed by building this model using several images of the person. To solve such issues, we propose a new identification process, which combines face recognition and image signature. It provides a reliable and hard to break identification system.

Spectroscopy is a valuable tool implemented in many applications. Spectral measurements from human tissue, is already used in biomedicine for classification and monitoring applications. Hyperspectral imaging captures distinctive unique patterns, which assists in composing unique signatures and codes.

2 Literature Review

There are four basic techniques for acquiring the three-dimensional multi-layer (x , y , λ) dataset of a hyperspectral cube. The use of hyperspectral in face recognition has been reported and led to an application of hyperspectral imaging to face recognition for a secure access control system, improving traditional face recognition. Pan et al. [1] explore measurements of hyperspectral artifacts to be used for face recognition in the near-infrared spectral range. Denes et al. [2] use three single visible bands (0.6, 0.7, and 0.8 μm) to test the spectral asymmetry. Chou and Bajcsy [3] use PCA for pre-processing hyperspectral, visible, and near-infrared images to extract initial principle information for face detection. Chang et al. [4] fused hyperspectral images in the visible spectrum (0.4–0.72 μm) into a single image and compared the result to the visible image, to validate the improvement of image fusion to face recognition. All the above, ignore the fact that each spectral band represents the specific reflectance/remittance information of the object. A certain type of tissue might have distinctive spectral reflectance features in some bands, but might not be significant in other bands [5, 6]. Cho et al. [7] present an automatic selection framework for the optimal alignment method to improve the performance of face recognition. Ghasemzadeh and Demirel [8] introduce a three-dimensional discrete wavelet transform (3D-DWT) based feature extraction for the classification hyperspectral. It is extracting the spatial and spectral information simultaneously. Sharma et al. [9] present a novel pipeline for discriminative band selection in hyperspectral images

for the task of image-level classification. It is based on a convolutional neural network for band selection. Chen et al. [10] reduce noise adaptively from each spectral band, crop each face according to its eye coordinates, conduct a log-polar transform to each cropped face image and extract 2D Fourier spectrum from them. They use the collaborative representation-based classifier with voting for hyperspectral face recognition.

3 The Proposed System

It has been proved that standard face recognition systems are vulnerable to hacking. To overcome it, we propose adding to it an image, hard to break, signature. The signature is generated from the inner image layers obtained by a hyperspectral camera.

Since we are working on monochrome (x, y) space, we prefer the use of spectral scanning, which generates multiple and separated layers of the same image, each in 2D. Each layer is analyzed separately to generate a unique binary string, which then is used to build the image, unique signature. Figure 1 depicts five image layers of the same person taken by a hyperspectral camera.

The following steps obtain the image signature:

- (1) Obtain the hyperspectral image and separate it into layers.
- (2) For each layer, select the “encryption” function to be activated.
- (3) Each layer is analyzed by the selected function and returns a binary string (partial signature).
- (4) The binary strings, generated by all analyzed layers are concatenated to build a single binary string.
- (5) The complete concatenated binary strings, is the image signature.

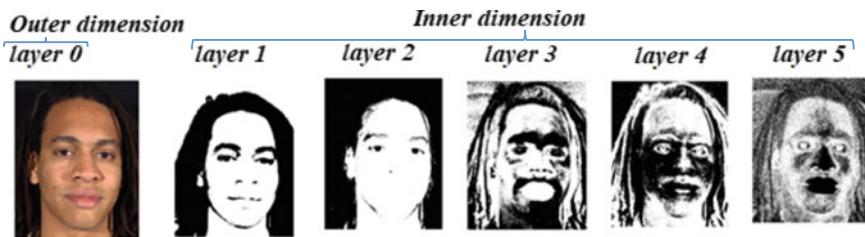


Fig. 1 Face image taken by a hyperspectral camera (with multiple layers)

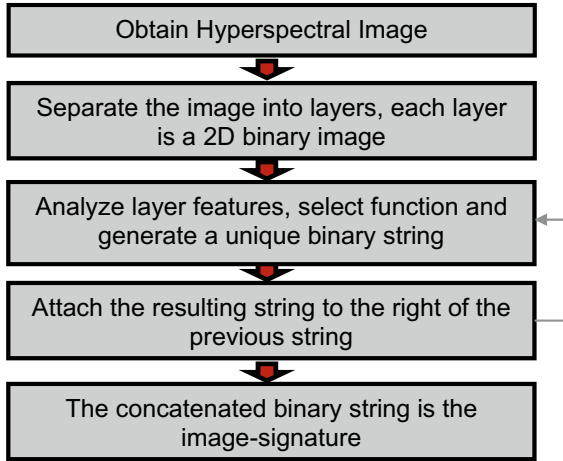


Fig. 2 Signature generation process

To reduce the probability of reconstructing the signature, we pair a function and an image layer. Assuming a hyperspectral image of l layers and an average number of functions per layer f , the probability of a specific combination of layer and function is $1: l * f$. To strengthen the signature uniqueness, we may apply each function several times for a given pair or change the concatenation sequence of the binary strings. Figure 2 describes the image-signature generation process.

4 Implementation of the Proposed Process

To better understand the implementation of the signature generation process, we use an example using five hyperspectral layers and five known “encryption” algorithms. In the experiment of the proposed system, we randomly selected one algorithm to perform the analysis. The numerical result of each algorithm is a binary string. The compound of these binary strings is the desired unique biometric signature, corresponding to the person behind the image. Following, we describe the implementation steps, using an example:

- I. **Obtaining the image layers:** Running the *converting* function (MATLAB software) to separate the different layers:

```
>> img1=converting(mcCOEF, 1);
>> img2=converting(mcCOEF, 2);
>> img3=converting(mcCOEF, 3);
>> img4=converting(mcCOEF, 4);
>> _img5=converting(mcCOEF, 5);
```

Figure 3 depicts the output of this stage:

II. Analyzing the inner layers to generate its associated bit string:

- a. First layer: we use the CascadeObjectDetector function, based on the Viola-Jones algorithm for finding features inside. (Nose, mouth, pair of eyes, right eye, left eye, and upper body.) The Viola-Jones algorithm for identifying facial features in a picture or identifying face features in a picture, requires a camera-facing image, without any side tilt. The algorithm has four main stages: use of rectangles for classification, creating an integral image, applying a learning algorithm, and classifying of cascade. Figure 4 depicts the output of this stage:
- b. **Second layer:** we use the local binary patterns (LBP) algorithm [11]. LBP is a type of visual descriptor used for classification. LBP is the case of the Texture Spectrum model. The feature vector can be processed using the support vector machine (SVM), extreme learning machines, or some other machine-learning algorithm for classifying images. Figure 5 presents examples of texture primitives.

Obtaining the image layers: Running the function (MATLAB software) to separate the different layers:

```
>> img1=converting(mcCOEF, 1);
>> img2=converting(mcCOEF, 2);
>> img3=converting(mcCOEF, 3);
>> img4=converting(mcCOEF, 4);
>> _img5=converting(mcCOEF, 5);
```

depicts the output of this stage:

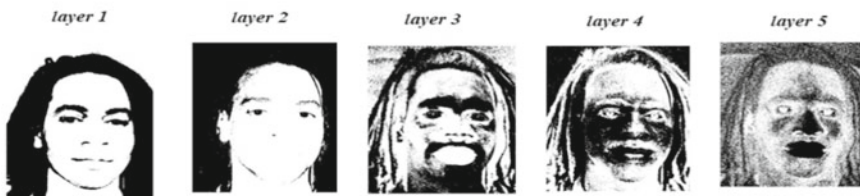


Fig. 3 Divide the hyperspectral image into five dimensions/layers

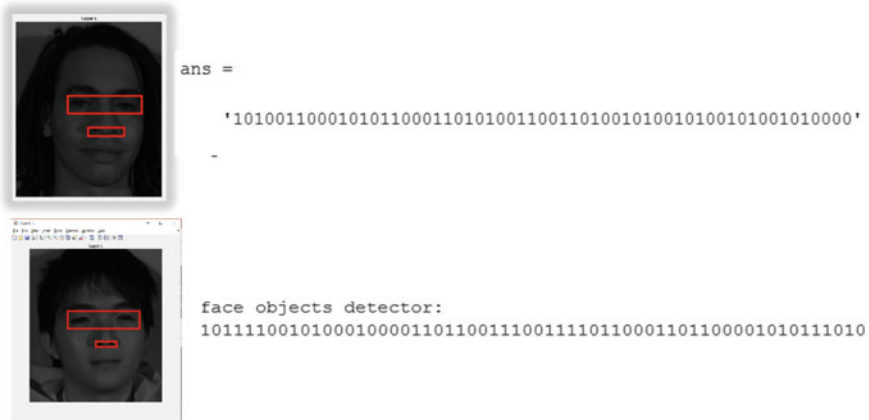


Fig. 4 Image analysis using the CascadeObjectDetector function

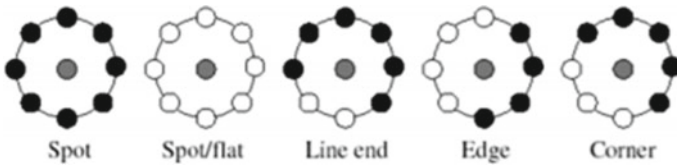


Fig. 5 Examples of texture primitives

The LBP algorithm starts with dividing the image into squares. For each square, it counts the number of appearances of each pixel in it and builds a binary string for each square and then combines these strings into a single string, which uniquely identifies a feature and face. The algorithm provides the pixels and their neighbors, allowing the discovery of a specific feature, such as the line of the lips or the eyes and its inner components. The output of this stage is demonstrated in Fig. 6.

- c. **Third layer:** In this layer, the *Imfindcircles* function is used. The function uses the circular Hough transform (CHT) algorithm to find circles in an image, and the CHT algorithm for identifying in an image, circles according to a given radius value range, and the color of the circle with respect to its surroundings. We guided the function to find just pupils, rather than other circles in the image. It is done by providing the min/max values determining the size of a pupil and finding the darkest circle in the eye. Once the pupil is identified, the output of the *Imfindcircles* function is a binary string. The output of this stage is demonstrated in Fig. 7.
- d. **Fourth Stage:** In this layer, we combine the two components related to the eyes: the angles between the eyes, and the distance between the two eyes. The eye detection is done separately using the Viola-Jones algorithm. The output of this stage is demonstrated in Fig. 8.



Fig. 6 Image analysis using LBP algorithm

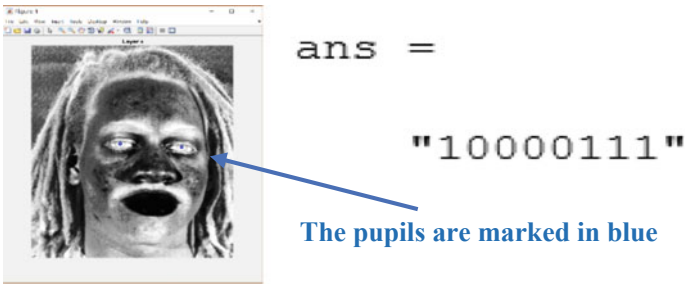




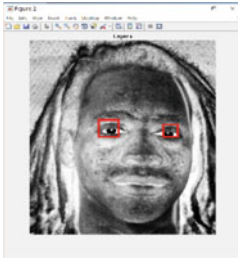


Fig. 7 Image analysis using *Imfindcircles* function

e. **Fifth layer:** Activating the *bwarea* function. The area of each pixel is determined by looking at its four neighbors. There are six different alternatives:

1. If there are zero white pixels, the area is 0. 
2. If there is only one white pixel, the area is 1/4. 
3. If there are 2 adjacent pixels, the surface brick is 1/2. 
4. If there are two diagonal pixels, the surface area is 3/4. 



ans =

'100110010101'

Fig. 8 Calculating the angles and the distance between eyes

5. If there are three white pixels, the surface is $7/8$.



6. If all four are white neighbors, the area is 1.



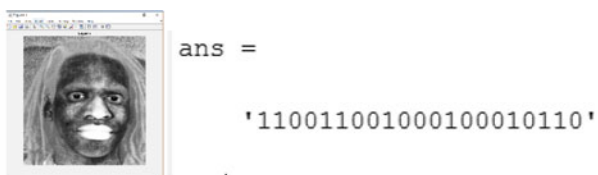
The output of this stage is demonstrated in Fig. 9:

f. **Final Stage:** The signature composition is done by combining the five-bit strings. Since we used an arbitrary algorithm for each hyperspectral layer, we can expect getting a unique biometric signature. This stage is demonstrated in Fig. 10.

5 Experiment

We used about 30 face hyperspectral images as input to produce a database containing just biometric signatures. We ran the system with a mix of new and existing images to check the level of miss detection and false alarms.

TEST #1 (Image #1 new, Image #2 database)



face area:

100101001010010101010011001101001010000

number of nonzero elements in the image:

1100110010001000101110

area of objects in the image:

100010100000011000101

euler number:

10010101110001010

number of positive elements in the image:

11010101110110010110

number of negative elements in the image:

11000011001110000000

ratio between the positive and the negative:

1

whole signature:

100101001010010101010011001101001010000110011001000100010111010001010000001100010110010101110

Fig. 9 Image analysis using the *bwarea* function

```

face area:
1001010010100101010011001101001010000
number of nonzero elements in the image:
110011001000100010110
area of objects in the image:
100010100000011000101
euler number:
10010101110001010
number of positive elements in the image:
11010101110110010110
number of negative elements in the image:
11000011001110000000
ratio between the positive and the negative:
1
whole signature:
100101001010010101001100110100101000011001100100010001011010001010000001100010110010101110

```

Fig. 10 Results for a typical figure



```

face objects detector:
110001110011101101
Local binary pattern:
10001101010101001
pupils detector:
Angle and distance between the eyes:
1
Black and white area:
10101001011011110000

```



```

face objects detector:
1111110100111100010111001001001011
Local binary pattern:
10010101101011100001
pupils detector:
1000
Angle and distance between the eyes:
1
Black and white area:
101111000010010011000

```


Experiment results**TEST #2 (Image #1 new, Image #2 database)**

FIRST SIGNATURE:
 11111101001111000101110010010010111001010110101110000110001101111000010010011000
 SECOND signature:
 11000111001110110110001101010101001110101001011011110000
 ACCESS DENIED



face objects detector:

11100000110101001001001110

Local binary pattern:

10010011000001000000

pupils detector:

Angle and distance between the eyes:

1

Black and white area:

101100111011100000111

whole signature:

111000000110101001001101100100110000010000001101100111011100000111

FIRST SIGNATURE:

111000000110101001001001110100100110000010000001101100111011100000111

SECOND signature:

10111100101000100001101100111001111011000110110000101011101011011001100110001110001100010000100100

ACCESS DENIED

face objects detector:

101111001010001000011011001110011110110001101100001010111010

Local binary pattern:

110110011001

pupils detector:

100

Angle and distance between the eyes:

1

Black and white area:

11000110001000100100

whole signature:

10111100101000100001101100111001111011000110110000101011010110011001110001100010000100100

TEST #3 (Image #1 new, Image #2 database)



face objects detector:
 11111101001111100010111001001001011
 Local binary pattern:
 10010101101011100001
 pupils detector:
 1000
 Angle and distance between the eyes:
 1
 Black and white area:
 101111000010010011000

face objects detector:
 11111101001111100010111001001001011
 Local binary pattern:
 10010101101011100001
 pupils detector:
 1000
 Angle and distance between the eyes:
 1
 Black and white area:
 101111000010010011000

Experimental results

```
FIRST SIGNATURE:
110001110011101101100011010101010011101010010110111110000
SECOND signature:
110001110011101101100011010101010011101010010110111110000
YOU GOT A PERMISSION TO ENTER INTO THE SYSTEM
.. |
```

6 Conclusions and Future Work

In this work, we cope with the reliability of standard image processing used for person identification. To enhance its reliability, we propose the use of image signature in addition to the existing standard image identification. Generating the image signature, we require a hyperspectral image, which is a multi-layer image. The proposed system, accepts the hyperspectral image, analyzes it, and generates a unique

signature. To identify and authenticate a person, we require the face image along with its signature. The probability to discover the pairs of layer and function is low, due to the wide number of pair combinations. However, to ensure the identification and authentication strength, we require the signature to be unique for the closed list of authorized persons. Hence, in a very rare case, if the same signature has already been generated, the system regenerates another signature.

In the future work, we plan to focus on further testing of various cases and applying this approach to other areas.

Acknowledgements The Hyperspectral images appear in this paper have been taken from Torbjørn Skaulia and Joyce Farrell paper [11]. We are also grateful to the subjects who have consented to publishing their image.

References

1. Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1552–1560 (2003)
2. L. Denes, P. Metes, Y. Liu, Hyperspectral face database. Technical Report CMU-RI-TR-02-25, 2002
3. Y. Chou, P. Bajcsy, Toward face detection, pose estimation and human recognition from hyperspectral imagery. Technical Report NCSA-ALG-04-0005, 2004. <http://isda.ncsa.uiuc.edu/peter/> [online]
4. H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi, M. Abidi, An indoor and outdoor, multimodal, multispectral and multi-illuminant database for face recognition, in *Proceedings of CVPR2006, Workshop on Multi-model Biometrics*, June 2006
5. B. Guo, S. Gunn, R. Damper, J. Nelson, Band selection for hyperspectral image classification using mutual information. *IEEE Geosci. Remote Sens. Lett.* **3**(4), 522–526 (2006)
6. R. Huang, M. He, Band selection based on feature weighting for classification of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2**(2), 156–159 (2005)
7. W. Cho, J. Jang, A. Koshan, M. Abidi, J. Paik, Hyperspectral face recognition using improved inter-channel alignment based on qualitative prediction models. *Opt. Express* **24**(24) (2016)
8. A. Ghasemzadeh, H. Demirel, Hyperspectral face recognition using 3D discrete wavelet transform, in *IEEE Xplore, Image Processing Theory Tools and Applications (IPTA)*, 2017. <https://doi.org/10.1109/ipta.2016.7821008>
9. V. Sharma, A. Diba, T. Tuytelaars, L. Van Gool, Hyperspectral CNN for image classification & band selection, with application to face recognition. Technical Report: KUL/ESAT/PSI/1604, 12/2016
10. G. Chen, W. Sun, W. Xie, Hyperspectral face recognition with log-polar Fourier features and collaborative representation based voting classifiers. *IET Digit. Libr.* **6**(1), 36–42 (2017). <https://doi.org/10.1049/iet-bmt.2015.0103>. Print ISSN 2047-4938
11. T. Skaulia, J. Farrell, A collection of hyperspectral images for imaging systems research. *Proceedings Volume 8660, Digital Photography IX; 86600C* (2013). <https://doi.org/10.1117/12.2007097>. IS&T/SPIE Electronic Imaging, 2013, Burlingame, California, United States. SPIE.digital.library 03-Feb-2013

Unsupervised Learning of Image Data Using Generative Adversarial Network



Rayner Alfred and Chew Ye Lun

Abstract Over the past few years, with the introduction of deep learning techniques such as convolution neural network (CNN), supervised learning with CNN had achieved a huge success in the computer vision area such as classifying digital images. However, supervised learning has a major drawback, in which it requires a large dataset for them to perform more effectively. As the data used in training grew bigger, the cost of labeling data for training becomes more expensive and impractical. In order to resolve this issue, unsupervised learning is encouraged to be used as it can draw inferences from datasets consisting of unlabeled input data. Generative adversarial network (GAN) is one of the unsupervised learning technique that has the ability to create natural-looking images, converting text description into images, recover resolution of images and last but not least, its power of representation learning from unlabeled data. Thus, this study attempts to evaluate the effectiveness of GAN algorithm in performing the supervised task and unsupervised task such as classification and clustering. Based on the results obtained, the GAN algorithm can learn the internal representation of data without labels and can act as good features extractor. Future works include applying GAN framework in other domains such as video, natural language processing and text to image synthesis.

Keywords Unsupervised learning · Supervised learning · Generative adversarial network · Feature extraction

R. Alfred (✉) · C. Y. Lun
Knowledge Technology Research Unit, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
e-mail: ralfred@ums.edu.my

C. Y. Lun
e-mail: c_hello94@hotmail.com

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_10

1 Introduction

Machine learning is a field of artificial intelligence that enables computers to act and make data-driven decisions rather than being explicitly programmed. The task of machine learning such as classification and recognition in computer vision often required input that is mathematically and computationally convenient to process. Thus, feature learning is important as it allows a system to discover the representation needed for those tasks from raw data. One of the deep learning techniques, convolution neural networks (CNN), is one of the top-performing algorithms and it had led a significant enhancement in computer vision [1–3]. However, almost all the CNNs require heavily labeled data and months of training. The annotation of a large dataset is very expensive which required human power, time, expensive annotation interfaces and usually subject to human biases. Other than that, the paucity of data in real-world problem such as rare diseases in the medical application made the process harder to complete. In such, unsupervised feature learning of image representation shows a promising direction as the resources are quasi-unlimited due to the high digital image data flows nowadays. Generative adversarial networks (GAN) consist of two models which are called the generative model G and the discriminative model D [4]. The whole idea of GAN can be thought of as the generative model, G trying to produce fake currency and use it without detection, while the discriminative model D trying to detect the fake currency. In such, GAN is one of the generative models that take advantage from unsupervised learning as it can learn features from the discriminator that could improve the performance of classifiers when limited labeled data is available.

Thus, the aim of this study is to evaluate the effectiveness of GAN algorithm in the representation learning power in the image domain. The rest of this paper is organized as follows. A brief background related to generative adversarial network will be described in Sect. 2. The experimental setup which is designed to evaluate the effectiveness of GAN algorithm in the representation learning power in the image domain will be outlined in Sect. 3. Section 4 presents the results and discussion related to the experiments conducted. Section 5 concludes the paper.

2 Generative Adversarial Network

GAN is composed of two models, a generative model and a discriminative model. The goal of GAN is to train a generator network (G) that produces samples from data distribution, by transforming vector noise. The training signal for G is provided by a discriminative network (D) that is trained to distinguish between samples that generated from G and real data. The generator network G trained to fool the discriminator into accepting its output as being real. G is a differentiable function represented by a multilayer perceptron where G can input noise variables to output the generator's distribution. D is trained to maximize the probability of assigning the correct label

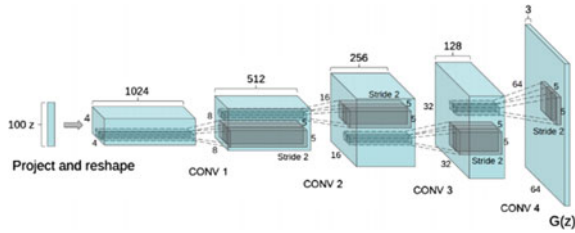


Fig. 1 Generator of DCGAN

to both training examples and samples from G . G is trained by receiving the update from D and G . In other words, D and G are playing the two-player minimax game and trained until they reach a point which both models cannot improve. One of the variants of GAN is called deep convolution generative adversarial network (DCGAN). The main difference between GAN and DCGAN is the replacement of any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator) which allow the network to learn its own spatial downsampling. It uses batch norm in both the generator and the discriminator in order to stabilize learning by normalizing the input to each unit to have zero mean and unit variance. It is very important as it prevents the generator from collapsing all samples to a single point. It also removes the fully connected hidden layers and uses rectified linear unit (ReLU) activation in the generator for all layers except the output, which uses Tanh activation function. In order to test the quality of the representation learned by DCGANs, they use the convolutional features of the discriminator to form a dimensional vector and a regularized linear L2-support vector machine (SVM) classifier is trained on top of them. Figure 1 shows the architecture of the generator in DCGAN.

Currently, there are many different variations of GAN frameworks performing testing on different dataset [5–9]. This paper will implement the mentioned DCGAN framework to perform clustering and classification on the STL-10 dataset. One of the motivations to perform mentioned tasks on the STL-10 dataset is that the dataset never been tested using DCGAN framework. Since every dataset has different parameter settings for a machine learning model to achieve the best performances, this paper intends to determine the parameter setting of DCGAN which produces the best performance when using the STL-10 dataset with k-means clustering method and SVM classification method. Besides that, due to the difficulty of convergence in DCGAN framework reported in past research, this study aims to improve the performance of overall framework by tuning the hyper-parameters of the discriminator.

3 Experimental Setup

In this experimental study, DCGAN algorithm is implemented using the Python platform in order to cluster and classify image data from STL-10 into correct categories. Four experiments were conducted by using different learning rate settings to determine which setting is the best for STL-10 dataset.

Data Acquisition. The STL-10 dataset [10] is a set of digital images which best used for determining the performances of the unsupervised learning, deep learning, and self-taught learning algorithms. The dataset is inspired by CIFAR-10 dataset which contains pictures of objects belonging to ten classes. The ten classes are airplane, bird, cat, car, deer, dog, horse, monkey, ship and truck. There are 500 training images and 800 test images per class. Additionally, the dataset contains 100,000 unlabeled images for unsupervised learning which extracted from a similar but broader distribution of images. As an example, it contains other types of animals (bear, rabbit, etc.) as well as different types of vehicles (trains, buses, etc.) in addition to the ones in the labeled set. The size of each image is roughly 96×96 pixels with color. One of the reasons to use this dataset is due to finding the best parameters settings of the mentioned dataset in GAN framework as there is not any past research conducted using STL-10 dataset. Other than that, this dataset is also convenient to investigate the relationship of the unsupervised learning of GAN framework toward the performance of classification and clustering task because it contains 10,000 unlabeled images for training as well as 5000 labeled images for testing. Although the dataset consists of annotation files, the dataset will be trained in unsupervised fashion. In other words, the model will be trained with the 100,000 unlabeled images. After that, 5000 test images will be used in the testing of k-means feature extraction clustering. The same batch of test images also will be used in the testing of SVM classification. The images will be center cropped to 64×64 height and width before fed into DCGAN.

Network Training. Both the generator and discriminator models are trained with stochastic gradient descent. Stochastic gradient descent is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions. Here, the training will maximize the D of x for every image from the true data distribution while minimizing the D of x for every image not from the true data distribution until a fixed number of iterations. Adam optimizing is used and the learning rate will be fixed to the settings of each different experiment and momentum term will be fixed in 0.5. Batch size of 64 is used in training. All the weights are initialized with zero Gaussian noise with a standard deviation of 0.02. In the LeakyReLU, the slope of the leak was set to 0.2 in all models. For every time step, the data is chosen from both distributions, then updated discriminator using the ascending gradients. After that, G is updated using the descending gradients.

Feature Extraction. Once the training is convergence, 5000 test images are used for the feature extraction. The features of those image data are extracted from the final fully connected layer in discriminator. Then, the features are then flattened and

fit into a k-means method to cluster these features into categorical distribution. The same features also fit into an SVM classifier.

Clustering using k-means. The aim of k-means algorithm is to partition M points in N dimensions into k clusters in which each observation belongs to cluster with the nearest mean. K-means algorithm is known as one of the most common and simplest clustering methods. The extracted features from the fully connected layer of discriminator will be clustered using k-means into discrete set of k -categories. K will be the number of categories that available in the dataset. After all the features are clustered into several categories, a statistical measurement will be conducted manually.

Support Vector Machine (SVM). Support vector machine (SVM) is widely used for pattern classification problems. The advantages of SVMs are high generalization ability especially when the number of training data is small, adaptability to various classification problems by changing kernel functions. Support vector machines with a square sum of slack variables are called L2-SVMs [11]. The objective of L2-SVM is defined in Eq. 1.

$$\text{minimize } \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^M \varepsilon_i^2 \quad (1)$$

Comparing to other classifiers, L2-SVM is much more stable due to the L2-SVM are less susceptible to outliers. Other than that, L2-SVM is differentiable and impose a bigger (quadratic vs. linear) loss points which violate the margin over L1-SVM [12]. As all the deep neural networks will ultimately output a final feature vector representation of the input, which must then be classified or performed some other task. This is generally done using a simple linear classifier.

Evaluation. After all the features are clustered into categories, purity of each cluster (PUR) shown in Eq. 2 will be calculated to evaluate whether the categories' predictions match the actual classes. In addition to that, the adjusted Rand index (ARI) and normalized mutual information (NMI) are also used as shown in Eqs. 3 and 4.

$$\text{purity} = \frac{1}{N} \sum_{i=1}^k \max_j |C_i \cap t_j| \quad (2)$$

$$\text{NMI}(Y, C) = \frac{2 \times I(Y, C)}{[H(Y) + H(C)]} \quad (3)$$

$$\text{ARI} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \frac{a_i}{2} \sum_j \frac{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \frac{a_i}{2} \sum_j \frac{b_j}{2} \right] / \binom{n}{2}} \quad (4)$$

Confusion matrix will also be used for the evaluation of the SVM classifier. Based on the number of true positive, true negative, false positive and false negative, accuracy, precision and recall can be computed. GAN and its variant frameworks are notorious for unstable training in which the instability of training occurred due to the one network overpowers the other one and it could be discriminator network or generator network. For example, if the discriminator behaves badly, the generator could not receive accurate feedback and the loss function cannot be used to represent the ability of the generator to generate image look closely to the real images. However, if the discriminator performs well, the learning becomes slow as the generator struggle to create something new to trick the generator. In both cases, the training is considered as a failure. Thus, in order to investigate the relationship between the successful training of DCGAN toward the clustering and classification task, the learning rate of DCGAN is tuned with various settings. There are four settings of DCGAN with different learning rates of both networks (discriminator learning rate (DLR) and generator learning rate (GLR)). The four settings are DLR 0.0002 and GLR 0.0002, DLR 0.0004 and GLR 0.0004, DLR 0.0006 and GLR 0.0002 and DLR 0.0002 and GLR 0.0006. The reason for using the learning rate setting of the first experiment is to validate whether the learning rate setting used in previously published research involving other datasets can be used again in the STL-10 dataset. In the next experiment, higher learning rates on generator and discriminator are used in order to investigate the effect of these learning rate settings toward the performance of clustering and classification compare to the lower learning rate settings. The third experiment has a high learning rate which is 0.0006 applied only on the discriminator in order to investigate the effects toward the performance of clustering and classification due to the discriminator are used as a feature extractor for those mentioned tasks. Last but not least, to find out whether a strong learning generator could prevent instability of training and thus increase the performance of classification and clustering, the fourth experiment applies a high learning rate only on the generator.

4 Results and Discussion

Based on the experiments conducted, the results are shown in Fig. 1. The results of clustering and classification tasks are shown in Table 1. The result of training on

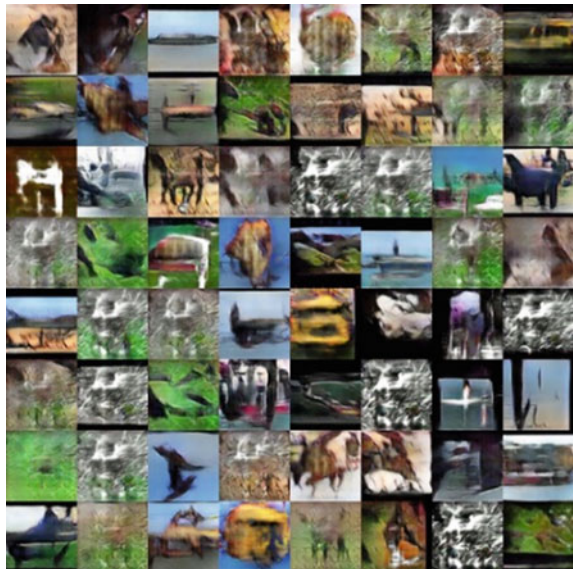
Table 1 Results of clustering and classification

Model	ARI	NMI	PUR	ACC
InfoGAN [5]	0.093	0.203	–	–
DCGAN DLR (0.0002), GLR(0.0002)	0.080	0.240	0.2804	0.569
DCGAN DLR (0.0004), GLR (0.0004)	0.074	0.233	0.2664	0.555
DCGAN DLR (0.0006), GLR (0.0002)	0.084	0.251	0.2565	0.588
DCGAN DLR (0.0002), GLR (0.0006)	0.112	0.267	0.2906	0.613

STL-10 dataset shows the generator can generate images, and however, the generated images are not similar to the real images. Since the objective of the discriminator is to make $D(G(z))$ near 0, while the generator needs to make the same quantity near 1, and the training of GAN is to find the Nash equilibrium of both networks, the ideal loss of both discriminator and generator are 0.5. However, it is difficult to achieve due to updating the gradient of both models cannot guarantee the occurrence of convergence [13]. During the start of the training, the loss of generator is near to 0 and the loss of discriminator is very high. This is because the discriminator cannot discriminate between the actual sample and the fake sample. Once the training goes on, the loss of discriminator decreases steadily because it can learn and classify the actual and fake images. During the training for epoch 10–25, the discriminator loss value is near 1 which indicates it is learning through the process. However, the discriminator can overpower the generator and achieving loss near to 0. When the discriminator is too powerful, it is almost guaranteed that the $D(x) = 1$ where the x is from real data pool while the $D(x) = 0$ where the x is from the generated data pool. In such, the loss function falls to 0 and there is no gradient to update the loss during learning iterations. This indirectly leads to “the Helvetica scenario,” in which G collapses too many values of z to the same value of x to have enough diversity to model P_{data} when G trained too much without updating D. Partial mode collapse can be observed from Fig. 2 where the generator collapses to a setting that it produced multiple images that contain same color or texture themes. Nevertheless, the generator can generate some images that are related to real images such as bus and airplane.

Overall, the result of all clustering using DCGAN with various learning rate exceed the result of InfoGAN when using the same dataset to perform clustering. During the experiments, the discriminator is often become too strong for the epoch

Fig. 2 Result of training



over 30 and caused discriminator to learn things slowly as there is no gradient to improve the generator. Generator is unable to produce realistic images and tends to have partial mode collapse. In such, the features extracted from discriminator are not distinct enough to cluster into own class by k-means method. From the collected sample, the discriminator can somehow extract the feature of background color and cluster images with the identical background color. However, the detail feature inside an image such as a curve line is not able being extracted from discriminator. Thus, k-means method is not able to cluster images effectively with the same similarities (e.g., cat and dog). This proves that the discriminator can learn the features obtained from training data since the generated images are abstract and mostly consist of distinct background colors. The generator with higher learning rate can catch up with the overpowered discriminator and allowing the discriminator to learn more during the training. Still, the performance of the clustering with higher generator's learning rate only exceeds the others by little margin due to the mode collapse of the DCGAN. The experiment with the higher discriminator's learning rate generally performed worse than other models as this only speeds up the mode collapse and the generator was not able to trick discriminator thus slowing the learning of discriminator. Even discriminator that uses 0.0004 learning pair with the generator that uses the same learning rate, it produces the lowest result in term of ARI and NMI. Thus, this indicates that higher the learning rate of a discriminator, faster the generator overpowered the discriminator, sooner the mode collapse occurred due to the instability training.

Overall, all experiments have produced accuracy performances of more than 50% compared to the clustering result due to the cross-validation as it fits 80% of the extracted features into the SVM classifier for training. The generator with the highest learning rate scored the best, and this indicates that the mode collapse does affect the classification task. The lowest score of precision of all models is obtained for categories 4 (cat) and 6 (dog), respectively. Once again, the SVM classifier is unable to perfectly classify the class having almost same features (e.g., dog and cat) due to the mode collapse of the DCGAN. Although it does learn the hierarchy of features, it cannot identify the unique features of the images. However, unlike clustering task, the discriminator with the highest learning rate can score 0.596 which is the second best among the other models.

5 Conclusion

In this paper, DCGAN was implemented to perform the unsupervised learning of STL-10 dataset. The training stage of the DCGAN was difficult because partial mode collapse always occurs after 30th or longer epoch even with different sets of learning rate used during the training because the discriminator managed to always accurately classify fake images and real images. Thus, this indirectly causes generator to stop learning and produce only certain images that are confident to trick discriminator. After the training reached 50th epoch, the discriminator of DCGAN is used as feature

extractor of the test images data for clustering and classification. The highest purity score of clustering is 0.2906 which is achieved by using generator with learning rate of 0.0006 and discriminator's learning of 0.0002. The ARI and NMI scores of the clustering are also the highest and exceed past research with the mentioned learning rate setting. Besides that, the highest accuracy score, 0.613 is achieved using the same learning rate setting. This indicates that higher the learning rate of the generator, the later the mode collapse occurred and thus higher the performance of clustering and classification. Since partial mode collapse happened during the training of all the experiments, the future works include tuning of parameters which could not be done in this project due to time constraint. The results obtained are encouraging, but the combinations of values for the discriminator learning rate (DLR) and the generator learning rate (GLR) are not exhaustively investigated. Other than that, it is also interesting to apply GAN framework in other domains such as video, natural language processing and text to image synthesis.

References

1. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
2. E.A. Hay, R. Parthasarathy, Performance of convolutional neural networks for identification of bacteria in 3D microscopy datasets. *PLoS Comput. Biol.* **14**(12), e1006628 (2018). <https://doi.org/10.1371/journal.pcbi.1006628>
3. W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* **29**, 2352–2449 (2017)
4. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 2672–2680 (2014)
5. V. Premachandran, A.L. Yuille, Unsupervised learning using generative adversarial training and clustering, in *Proceedings of the International Conference on Learning Representations (ICLR)*, Toulon, France, 24–26 April 2017
6. X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: interpretable representation learning by information maximizing generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 2172–2180 (2016)
7. X. Mao, et al., Least squares generative adversarial networks, in 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE, 2017)
8. J.T. Springenberg, Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv preprint [arXiv:1511.06390](https://arxiv.org/abs/1511.06390) (2016)
9. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN. arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875) (2017)
10. A. Coates, H. Lee, A.Y. Ng, An analysis of single layer networks in unsupervised feature learning, in *AISTATS*, 2011
11. Y. Koshiha, S. Abe, Comparison of L1 and L2 support vector machines, in *Proceedings of the International Joint Conference on Neural Networks*, 2003, vol. 3 (IEEE, 2003), pp. 2054–2059
12. Y. Tang, Deep learning using support vector machines. *CoRR*, abs/1306.0239, 2013
13. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs, in *Advances in Neural Information Processing Systems* (2016), pp. 2234–2242

Adaptive Message Embedding in Raw Images



Tamer Shanableh

Abstract In this paper, we propose an adaptive approach to message embedding in raw images. The cover image is divided into blocks where the top left pixels are dedicated as control pixels used to identify blocks carrying message information. Message bits are embedded while restricting the number of pixel changes in a cover segment to one change only. This is achieved through a concept of overloading the embedding indices; therefore, a maximum of one pixel value is changed per block for message embedding. The proposed solution is assessed in terms of percentage of changed pixels, PSNR, SSIM, histogram changes, and blind steganalysis. Comparison with existing work reveals that the proposed solution reduces both the image distortions and pixel change rates. It is also shown that the proposed solution is less detectable when tested with blind steganalysis in comparison to existing solutions.

Keywords Data embedding · Image processing · Steganography

1 Introduction

Multimedia data embedding has a number of applications including copyright protection, quality assessment and error concealment of transmitted multimedia content, medical applications like hiding patient records in medical images and convert conversations.

In general, the aim of data embedding techniques is to hide messages while minimizing detectability of the embedding and minimizing distortion caused to the cover medium like images, videos, and audio files. If the cover medium is compressed, then data embedding techniques has to minimize the increase in file size as well.

It was reported in [1] that data can be embedded successfully in motion vectors of compressed video. This was further extended by [2] to include both motion vectors of multilayer or scalable video coding and transcoding with the aim of increasing the overall data embedding rate while minimizing distortion.

T. Shanableh (✉)
American University of Sharjah, Sharjah, UAE
e-mail: tshanableh@aus.edu

It was also reported that the quantization scales of compressed video can be used for data embedding with minimal detectability [3]. Coding modes of compressed videos are also used for data embedding by altering the partitioning modes and split decisions of HEVC videos as reported in [4].

Data embedding in images, on the other hand, can be implemented in compressed JPEG images [5] or compressed JPEG2000 images [6]. It can also be implemented in uncompressed or raw images as well.

In this work, we focus on the spatial domain of natural images where data is embedded by modifying raw pixel values. Recent work focusing on this research domain is reported in [7] where a comprehensive solution was proposed based on a novel edge detection approach in which data is embedded adaptively in edge-based areas. The purpose is to minimize visual distortion while achieving an embedding rate up to 40%. In the spatial domain, raw images were divided into 3×3 block of pixels where corner pixels are used for edge detection and therefore facilitating content-adaptive data embedding. A similar approach was also reported in [8] for hiding encrypted patient records in MRI images.

Other pixel-domain data embedding solutions exist for both binary and grayscale images. For instance, Feng et al. [9] proposed a binary image steganographic solution where syndrome-trellis code is used to minimize the embedding distortion. As for the grayscale images, the work in [10] used the values of a pixel pair to search for a coordinate in the neighborhood set according to a given message digit. Message embedding is then achieved by replacing the pixel pair by the searched coordinate.

The solution proposed in our paper is inspired by the work of [7]. We have two contributions in this work, namely the paper proposes a fixed-size block-based solution that uses only one control pixel for adaptive data embedding.

The rest of the paper is organized as follows. Section 2 starts by explaining how message bits are embedded and extracted; it then introduces the proposed fixed-size block-based embedding solution. Section 3 discusses security considerations using blind steganalysis. Section 4 introduces the experimental results and Sect. 5 concludes the paper.

2 Data Embedding and Extraction Solution

This section explains how message bits are embedded and extracted from pixels. It also proposes a data embedding solution based on fixed block sizes.

2.1 Message Embedding and Extracting

The basic concept of message embedding is to hide a sequence of message bits while restricting the number of changes in the cover image to one change only. This is achieved through a concept of overloading the embedding indices or locations.

Assume that the message is divided into a sequence of n bits, typical values for n are 2, 3, and 4. The n bits can be embedded in k pixel values with a maximum of 1 pixel change given that:

$$k = {}_n C_1 + {}_n C_2 + \dots + {}_n C_n \tag{1}$$

For example, if $n = 3$, then k is equal to the total number of ways in which 1 out of 3 locations can be selected plus, the total number of ways in which 2 out of 3 locations can be selected plus, the total number of ways in which 3 out of 3 locations can be selected. That is, $k = {}_3 C_1 + {}_3 C_2 + {}_3 C_3 = 7$. The relation between the 7 pixel locations and the 3 message bits is illustrated in Table 1.

To embed 3 message bits in 7 pixel locations, we start with the first message bit, m_1 . If the parities of m_1 and the summation of the pixels at locations corresponding to m_1 are different, then we set the flag `m1_modification_required` to true. From the table, the pixels at locations corresponding to m_1 are 1, 4, 5, and 7. In general:

$$\text{modify_}m_i = \begin{cases} \text{T, } |\text{parity}(m_i) - \text{parity}(p_1 + \dots + p_j)| = 0, & i = 1, \dots, n \\ \text{F, } |\text{parity}(m_i) - \text{parity}(p_1 + \dots + p_j)| = 1, & i = 1, \dots, n \end{cases} \tag{2}$$

where p_i indicates the value of a pixel at index i . The subscript j is the total number of pixel locations to examine. The modification flag of Eq. 2 is computed for each of the 3 message bits; m_1 , m_2 , and m_3 . To embed the 3 message bits in the 7 pixel locations, we use the following table which is based on the relation between the message bits and pixel locations as illustrated in Table 2.

According to the truth table, if only the modification flag for m_1 , m_2 , or m_3 is true, then the LSB of p_1 , p_2 , or p_3 , is modified, respectively, by a random ± 1 . If 2 out of three message modification flags are set to true, then the LSB of p_4 , p_5 , or p_6 is modified, respectively. Lastly, if all message modification flags are set to true, then only p_7 is modified.

It is seen from the above example that at most one out of 7 pixels is modified to insert a sequence of 3 bits. If by chance, all the message modification flags are false, then no pixel modification is required.

Table 1 Relation between pixel locations and message bits

Pixel's location/index	Corresponding msg bits
1	m_1
2	m_2
3	m_3
4	m_1, m_2
5	m_1, m_3
6	m_2, m_3
7	m_1, m_2, m_3

Table 2 Table for identifying location of pixels to be modified

m_1 flag	m_2 flag	m_3 flag	Modify
F	F	F	–
T	F	F	p_1
F	T	F	p_2
F	F	T	p_3
T	T	F	p_4
T	F	T	p_5
F	T	T	p_6
T	T	T	p_7

Table 3 Number of required pixels needed to guarantee a maximum modification of 1 pixel value to embed n message bits

Message segment (n)	Number of required pixels, k
2	$2C_1 + 2C_2 = 3$
3	$3C_1 + 3C_2 + 3C_3 = 7$
4	$4C_1 + 4C_2 + 4C_3 + 4C_4 = 15$
5	$5C_1 + 5C_2 + 5C_3 + 5C_4 + 5C_5 = 31$

The extension of the above example to other values of n is straightforward, namely two message bits ($n = 2$) can be embedded in 3 pixels locations ($k = 3$) with a maximum of 1 pixel change. Likewise, four message bits ($n = 4$) can be embedded in 15 pixel locations ($k = 15$) with a maximum of 1 pixel change. Typical n values and corresponding k values are summarized in Table 3.

The message extraction procedure is based on examining the pixel parities according to the values in Table 1. More specifically, m_1 is listed with the corresponding pixel locations/indices {1, 4, 5, 7}, m_2 is listed with the corresponding pixel locations of {2, 4, 6, 7}, and lastly, m_3 is listed with the corresponding pixel locations of {3, 5, 6, 7}. To extract m_1 , simply compute the parity of the sum of the pixel locations {1, 4, 5, 7} and so forth for m_2 and m_3 .

Clearly, this message extraction procedure assumes that the length of the message segment, n , is known to the extractor. Otherwise, the value of n can be embedded in the cover image. The aforementioned message embedding and extraction algorithms are original; interestingly, they are similar to matrix encoding [11].

2.2 Selective Pixel Blocks for Message Embedding

In this solution, we vary the pixel block size according to the length of the message segment, n , namely, for $n = 2$, we use a pixel block size of 2×2 , for $n = 3$, we use 3×3 , and for $n = 4$, we use 4×4 . For control pixels, we use only one pixel located at the top left corner of each block. The pixel block sizes are summarized in Table 4.

Table 4 Pixel block sizes for difference lengths of message segments, n

Message segment (n)	Number of required pixels, k	Block size	Excessive pixels
2	3	2×2	$2 \times 2 - 3 = 1$
3	7	3×3	$3 \times 3 - 7 = 2$
4	15	4×4	$4 \times 4 - 15 = 1$

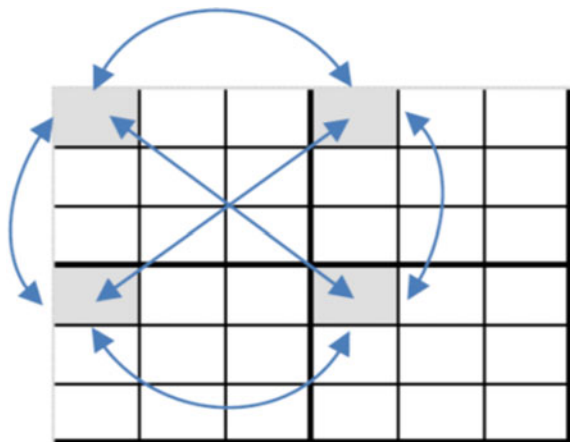
In the table, the last column indicates that for $n = 2$ and $n = 4$, each pixel block will contain one extra pixel that is not used for data embedded. This will be used as a control pixel and it is located at the top left corner of the block. For the case of $n = 3$ and $k = 7$, two extra pixels are available in the 3×3 pixel block. The top left pixel will be used as a control pixel as the case for $n = 2$ and $n = 4$. All remaining pixel values shall be grouped into vectors of 7 pixels ($k = 7$) and used for data embedding.

The control pixels are used for computing the sum of absolute differences in all directions similar to the work of [7] with the difference of having one control pixel per block instead of 4. The proposed arrangement of control blocks is illustrated in Fig. 1.

In this work, we propose to compute the summation of absolute differences (SAD) in the three directions (i.e., horizontal, vertical, and diagonal) and select the maximum value as mentioned above. The maximum SADs are then sorted in ascending order and the percentiles are found. If the message payload is say, 70% of the available pixel count then the SAD at the (100-70%)th percentile is selected as a global TH. Once the global TH is computed, message embedding will take place in pixel blocks with a max SAD greater than or equal to the global TH. Other pixel blocks will remain as is.

The selected global TH needs to be available for message extraction as well. Therefore, the global TH is embedded into the cover image as well.

Fig. 1 Location of control pixels for 3×3 blocks and arrows indicate the summation of absolute differences calculated in three directions



3 Steganalysis Consideration

In some scenarios, it is important to prevent detecting the existence of message embedding using blind steganalysis. Examples of such scenarios are hiding copyright information for digital rights management and hiding medical records of patients in medical images with privacy taken into account. Therefore, we assess the detectability of proposed solution using a well-known spatial domain steganalysis method called Li-11D [12] which uses normalized histograms of linear transformations as feature vectors. In the experimental results section, we show that blind steganalysis does not distinguish between stego images and clear images based on the proposed message embedding solutions.

4 Experimental Results

In all the experiments to follow, the message to be embedded is generated with a uniform random number generator.

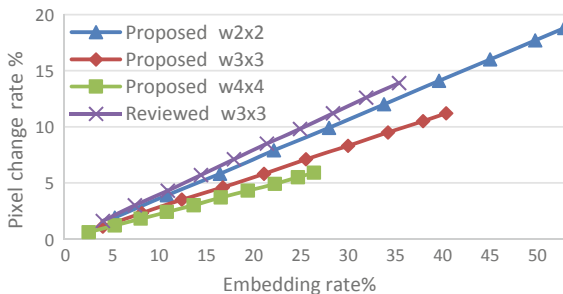
For data embedding in grayscale images, we compare our work against the recent work proposed by [7], and therefore, we use similar experimental setup. The image dataset used is BOWS2 [13] which contains 10,000 natural images with a spatial resolution of 512×512 . Since the number of image is very large, we use systematic sampling in which every third image is used in our experiments. Therefore, a total of 3334 images are used in our experiments.

The results to follow apply the proposed embedding solution on pixel blocks of sizes 2×2 , 3×3 , and 4×4 . The reviewed work of [7] presented results for 3×3 blocks only.

In all the figures to follow, the embedding rate for the solutions using a fixed block size is changed by varying the global embedding TH from the 90th SAD percentile all the way to 0th percentile where all pixels are used for data embedding. The computation of the TH based on SAD percentiles was introduced in Sect. 2.2.

We start by plotting the pixel change rate against the embedding capacity as shown in Fig. 2.

Fig. 2 Data embedding rate versus pixel change rate



As mentioned previously, in the proposed fixed block size solution, two message bits are embedding in a 2×2 block of pixels, three message bits are embedded in a 3×3 block of pixels, and four message bits are embedded in a 4×4 block of pixels. Therefore, the embedding rate in 2×2 pixels is the highest as shown in Fig. 2. However, higher embedding rates imply higher percentage of pixel changes as shown in the figure as well. Clearly, embedding 4 bits in 4×4 pixel blocks results in the lowest embedding rate and lowest pixel change rate.

It is also shown that the proposed embedding solutions have lower rate of pixel change in comparison to the reviewed solution. This result indicates that the image distortion caused by the proposed solution is lower than that of the reviewed work. This is so because the proposed solution changes a maximum of one pixel value in each embedding block.

The average PSNRs of the proposed and reviewed solutions are plotted in Fig. 3.

As expected, the proposed fixed block size solution results in lower image distortions and therefore higher PSNR in comparison to existing work. When compared to the proposed 3×3 solution, it is interesting to see that the difference is around 2.5 dB, which is a noticeable improvement. The PSNR results are consistent with the pixel change rate of Fig. 2.

The figure presents another important conclusion; data embedding in pixel blocks of 4×4 results in the high PSNR and low change rate (Fig. 2). However, this solution is restricted to a maximum data embedding rate of 25%. In conclusion, for data embedding with rates less than 25%, a fixed pixel block size of 4×4 is preferred, again as illustrated in Table 4, and in this case, four message bits are embedded in 15 pixels. Additionally, we can conclude from Figs. 2 and 4 that for data embedding

Fig. 3 Average PSNR as a function of the embedding rate

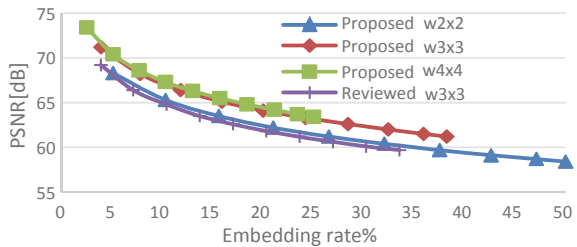
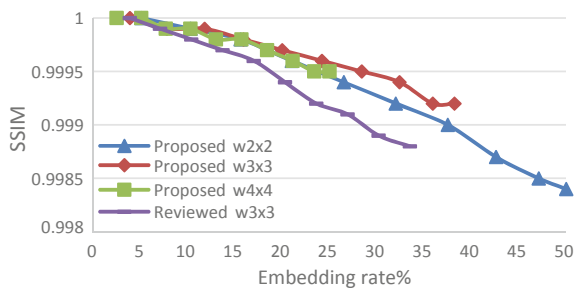


Fig. 4 Average SSIM as a function of the embedding rate



with rates between 25 and 40%, the proposed data embedding of three message bits in 7 pixels (a block size of 3×3) is preferred.

The structural similarity index, SSIM, is a method for predicting the perceived image quality. It is considered an enhancement over PSNR as it considers changes in structural information while taking image contrast into consideration. The average SSIM results are reported in Fig. 4.

As shown in the figure, at low embedding rates, nearly all solutions result in similar SSIM. As the embedding rate increases, the reviewed solution starts to embed more message bits into high-frequency regions which affects the structure of the image, therefore affecting the average SSIM values. The proposed 2×2 block size solution has a similar behavior as small block sizes capture the structure of the image and store message bits in high-frequency areas.

The normalized sum of absolute histogram differences between the cover and stego images are presented in Fig. 5.

The lower the histogram SAD, the better is the overall result. The results in Fig. 5 are consistent with the pixel change rate of Fig. 2 and the average PSNR values of Fig. 3 as well.

Table 5 presents the results of blind steganalysis with comparison to existing work. For the proposed solution, we used a fixed block size of 3×3 which is similar to the reviewed work. The aim is to train a classifier to distinguish between clean and stego images based on features extracted using various techniques.

Fig. 5 Normalized sum of absolute histogram differences between the cover and stego images

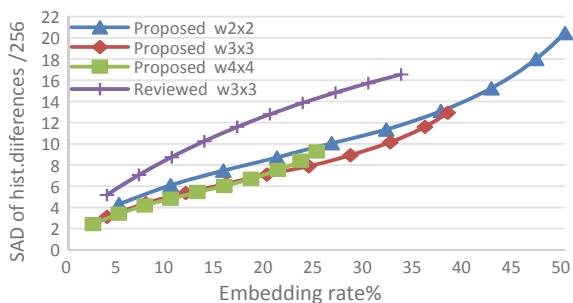


Table 5 Classification results using Li-110D steganalysis

Embedding rate (%)	Reviewed (%)	Proposed (%)
5	49.58	49.66
10	51.27	50.19
20	54.37	49.94
25	55.69	50.9
30	57.69	52.04
40	60.23	54.8
Avg.	54.78	51.25

As mentioned previously, we extract feature using the Li-110D technique which is designed for spatial domain steganalysis [12]. Half the data is used for training and the other half for testing. The model generation is done through SVM with a linear kernel. Tenfold round robin is used for training and testing and the average classification results are reported in Table 5. The classification results in the table indicate that the proposed solutions are less detectible than the reviewed work, and they also indicate that the use of Li-110D blind steganalysis cannot be used to detect stego images using the proposed solutions.

5 Conclusion

The paper proposed data embedding solutions in grayscale images. Pixel blocks with high-frequency content are adaptively selected for data embedding. This is achieved by examining the differences between control pixels in each block, where only one control pixel is used per block. Experimental results revealed that when the proposed solution is used with data embedding rates less than 25%, a block size of 4×4 is preferred. Whereas, for data embedding rates between 25 and 40%, a block size of 3×3 is preferred. The proposed solution also resulted in competitive image quality and provided a trade-off between embedding rates and image quality. It was also shown that by using two different blind steganalysis approaches, the classification accuracy was around 51%, which makes it difficult to detect data embedding using the proposed solution.

References

1. H. Aly, Data hiding in motion vectors of compressed video based on their associated prediction error. *IEEE Trans. Inf. Forensics Secur.* **6**(1) (2011)
2. T. Shanableh, Matrix encoding for data hiding using multilayer video coding and transcoding solutions. *Signal Process.: Image Commun.* **27**(9), 1025–1034 (2012)
3. T. Shanableh, Data hiding in MPEG video files using multivariate regression and flexible macroblock ordering. *IEEE Trans. Inf. Forensics Secur.* **7**(2) (2012)
4. T. Shanableh, Altering split decisions of coding units for message embedding in HEVC. *Multimed. Tools Appl.* (2017). <https://doi.org/10.1007/s11042-017-4787-6>
5. J. Guo, T. Le, Secret communication using JPEG double compression. *IEEE Signal Process. Lett.* **17**(10) (2010)
6. L. Zhang, H. Wang, R. Wu, A High-capacity steganography scheme for JPEG2000 baseline system. *IEEE Trans. Image Process.* **18**(8) (2009)
7. H. Al-Dmour, A. Al-Ani, A steganography embedding method based on edge identification and XOR coding. *Expert Syst. Appl.* **46** (2016). <http://dx.doi.org/10.1016/j.eswa.2015.10.024>
8. H. Al-Dmour, A. Al-Ani, Quality optimized medical image information hiding algorithm that employs edge detection and data coding. *J. Comput. Methods Progr. Biomed.* **127** (2016)
9. B. Feng, W. Lu, W. Sun, Secure binary image steganography based on minimizing the distortion on the texture. *IEEE Trans. Inf. Forensics Secur.* **10**(2) (2015)
10. W. Hong, T. Chen, A novel data embedding method using adaptive pixel pair matching. *IEEE Trans. Inf. Forensics Secur.* **7**(1) (2012)

11. R. Crandall, Some notes on steganography (1998). Available at: http://dde.binghamton.edu/download/Crandall_matrix.pdf. Accessed Dec 2018
12. B. Li, J. Huang, Y. Shi, Textural features based universal steganalysis, in Electronic Imaging, International Society for Optics and Photonics, 2008
13. P. Bas, T. Furon, Bows-2 (2007)

Megapolis Tourism Development Strategic Planning with Cognitive Modelling Support



Alexander Raikov 

Abstract The strategic planning of the megapolis tourism development is the process that has to take into account hundreds of factors. Some of the factors cannot be calculated or do not have statistic history. Some of the factors are latent or incorrect. The long-term strategic planning usually includes a short-term action planning. In this case, experts are creating cognitive models that take into consideration non-formalized cognitive semantics. The modelling shows that small change in resource allocation or some mistakes in decision-making can be the reason not to achieve the goals and can replace the optimistic scenario of tourism development by a pessimistic one. The cognitive models could be improved by mapping on the relevant Big Data. The special author's approach to make the process of tourism strategic decision-making convergent was applied. This paper addresses the issue of using convergent approach to megapolis tourism development strategic planning with a lot of focus groups and cognitive modelling. The inverse problem solving with genetic algorithm helped to find effective strategic decisions and reduce risks of decision-making. For taking into account non-quantitative factors, it is suggested using networked expertise. The convergent approach with cognitive modelling was applied for creating the megapolis tourism development strategic plan for the megapolis. It helps to find the multiplying events and prioritize strategic directions for tourism development.

Keywords Artificial intelligence · Big data · Cognitive modelling · Convergent approach · Megapolis · Strategic planning · Tourism development

A. Raikov (✉)

Laboratory of Modular Information-management System, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznaya street, Moscow 117997, Russia
e-mail: Alexander.N.Raikov@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_12

147

1 Introduction

The strategic planning is well-known process that is realized with expert participation. The process is ill-defined and characterized by high risks' level as usual. The long-term strategic planning usually includes a short-term action planning for understanding priority actions to be undertaken to ensure sustainable tourism development [1]. The different methods of computer modelling are applying. But the models are created by experts and may have many mistakes, which give entail excessive risks.

The strategic planning is the poorly formalized process. It is unique and original as usual, described by a lot of conceptual characteristics that cannot be proved with statistics. The information is unreliable or simply missing. In the strategic decision-making process, it is necessary to take into account non-formalized factors, such as institutional strengthening, human resource development, socio-cultural environment, as emotions, feelings, and meditative states of consciousness of the participants.

There is a problem of model building in the absence of the opportunity to present the problem logically, with an indication of the causes and consequences of the occurrence of events. In such conditions, cognitive modelling [2] comes to the rescue with its potential for describing situations at a qualitative level, in the form of concepts and indicating the mutual influences (relations) between them. A cognitive model is a schematic conceptual model of the problem that takes into consideration not only the real situation, but also the specifics of the processes occurring under conditions of instability and uncertainty. Such a model allows getting answers to questions like: "What will happen if ...?" or "What should be done to ...?". This provides an opportunity to consider the dynamics of changes in the state of the studied subject area (tourism) and to evaluate potentially possible scenarios for its development.

The main emphasis in the traditional construction of strategies is made on assessing existing experience, regression and statistical styles of analysis, and extrapolating current trends. In contrast, cognitive modelling relies on the analysis of the strategic future and points the way to the future with the inverse problem-solving approach. The inverse problem solving is incorrect. To ensure its correctness, the special author's convergent approach to decision-making is applied [3].

This work is devoted to the convergent strategic planning of the tourism development of large megapolis. The Artificial Intelligence (AI), cognitive modelling, and verifications of cognitive models with Big Data analysis were applied.

2 Cognitive Semantic

The difficulties of cognitive modelling connect with requirements of taking into account cognitive semantics. These semantics are not formalized, cannot be represented in a logical way, and have to be considered indirectly.

Classical semantic approaches of AI are based on the fact that the semantics of texts is described by other texts or ontologies (frame-like constructions). These are

denotative semantics, which do not take into account non-formalized mentality, emotions, and feelings. Here, as in quantum physics, if one only starts to measure the cognitive semantic with logical constructions, so the cognitive semantic is immediately switched to denotative ones. The continuum power of denotative semantics is many orders of magnitude lower than the cognitive one [4, 5].

To enhance the inclusion of coverage of cognitive semantics in a weakly structured cognitive modelling of tourism sphere, the concepts of category theory and monads are used, and the idea of “convergent monad” (CO) was introduced [3]. The classical monads can be used for representing divergent and formalized decision-making processes. The classical monad does not provide the necessary conditions for ensure decision-making convergence. The CO prevents a divergence of decision-making processes. It is obtained to develop the classical monad by applying the author’s approach to convergent support of decision-making. The axioms of “convergent monad” \mathcal{E} were introduced [3]:

- $D: \mathbf{Set} \rightarrow \mathbf{Set}$; the number of elements in the system of the sets \mathbf{Set} is infinite, and the maps of the objects are the maps with a closed graph;
- \mathcal{B} is a non-empty finite subcover of the monad \mathcal{E} (*Compact space*—a topological space if each of its open covers has a finite subcover);
- Every neighbourhood of the point $e \in \mathcal{E}$ can be associated with some neighbourhood (in the topological sense, every open set contains this point) such that for each pair of points, there are always disjoint neighbourhoods (*Hausdorff space*).

These axioms help to make the tourism strategic-making processes fast and convergent. It should be noted that connecting external experts and even civil society to the strategic decision-making process makes the process almost infinite-dimensional. The convergent structuring of this process accelerates it [6].

3 Cognitive Modelling

The order of conducting expert procedures and cognitive modelling for tourism strategic planning is represented in Fig. 1.

With the SWOT-analysis method and focus groups, the substantial 15 factors were selected and the importance of the weight of their interaction was evaluated. To answer the question “What should be done to...?” (inverse problem solving) the genetic algorithm was used. With its help, the optimal ratio of controlling factors to ensure the implementation of the optimistic scenario of development of strategic tourism activities in megapolis was calculated. The cognitive modelling helps to kick-start stable tourism development. The following list of factors was formed:

1. The quality (including filling the budget) of tourism activity (*goal*).
2. Digitization of tourism services (*Digitization*).
3. Holding major international events: forums, congresses, etc. (*Major events*).
4. Development of event and business tourism (*Event and Business*).

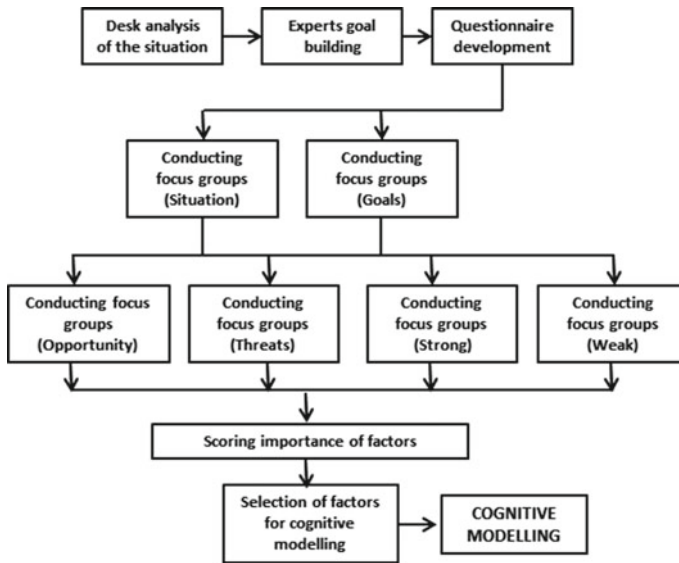


Fig. 1 Order of expert procedures for cognitive modelling

5. Promotion of the image of the megalopolis (*Image*).
6. Globalization of the market of tourism and hotel services (*Globalization*).
7. Investing in tourism infrastructure by foreign competitors (*Competitors*).
8. Investment in infrastructure related to inbound tourism (*Infrastructure*).
9. There is a reserve for the active development of tourism (*Rating*).
10. Objects of cultural heritage are world famous (*Heritage*).
11. Security level (*Security*).
12. People who have visited the megalopolis become promoters (*Promoters*).
13. Visa restrictions (*Visas*).
14. Improving transport accessibility (*Transport*).
15. The expansion of non-cash payments (*Non-cash*).

Then, the interaction between factors was assessed, and about 70 mutual factors' influences were defined. The weight of influence was estimated with the scale $[-1, 1]$. The factors that can be exercised as controlling factors was: 2, 4, 8, 11, and 14. The result of cognitive modelling in the author's software is shown in Fig. 2.

The blocks of factors and mutual factors' influences are responsible for editing models (Fig. 2). The direct and inverse problem-solving blocks are responsible for evaluating strategic scenarios. The block of model displays the cognitive model (factors and mutual factors' influences) as a directed graph.

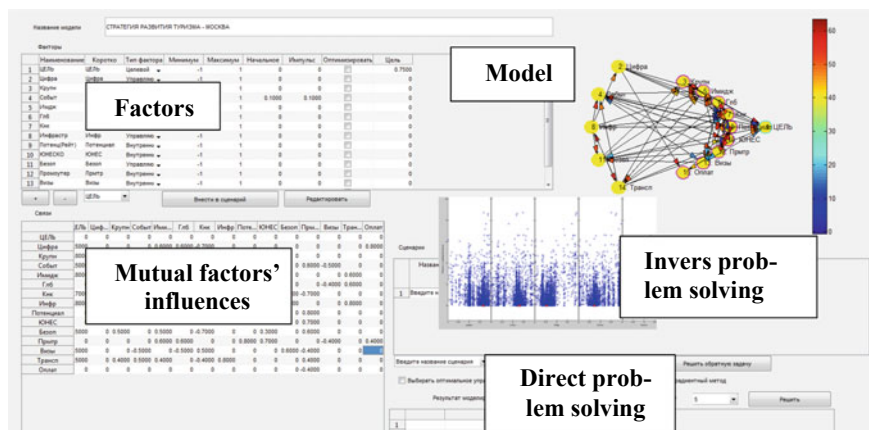


Fig. 2 Cognitive model in the special software

4 Big Data Verification of the Model

The verification of the cognitive model was carried out through its mapping on Big Data with tourism activity documents. The mapping was made for every individual factor or relation. The cognitive model verification software allows displaying verification results with superimposing it on a model built by experts. The verification process was conducted on a block of data of Russian-language news sources of the Internet (e.g. “Yandex.News” media database).

The search and collection of information were carried out using a shareware program for processing data from Internet sites (the crawler program). Large-scale studies of the news background can be conducted using the services that allow sending requests to the databases of Internet aggregators (Yandex, Google, etc.) and receiving answers in the format required for cognitive model verification (Fig. 3).

In Fig. 3, the arcs of the computer simulation scheme (cognitive model) are colour-coded in accordance with the verification results:

- Green—expert hypothesis about the relationship between a pair of factors was confirmed (high probability of a connection);
- Yellow—the hypothesis of experts about the relationship between a pair of factors was confirmed (low probability of a link);
- Blue—the possibility of a connection for expert verification of its inclusion in a computer simulation scheme.

The result of the verification of the cognitive model showed a high degree of compliance of the constructed cognitive scheme with the data of the subject area, collected from the media (78% confirmed hypotheses about the presence of a relationship between factors).

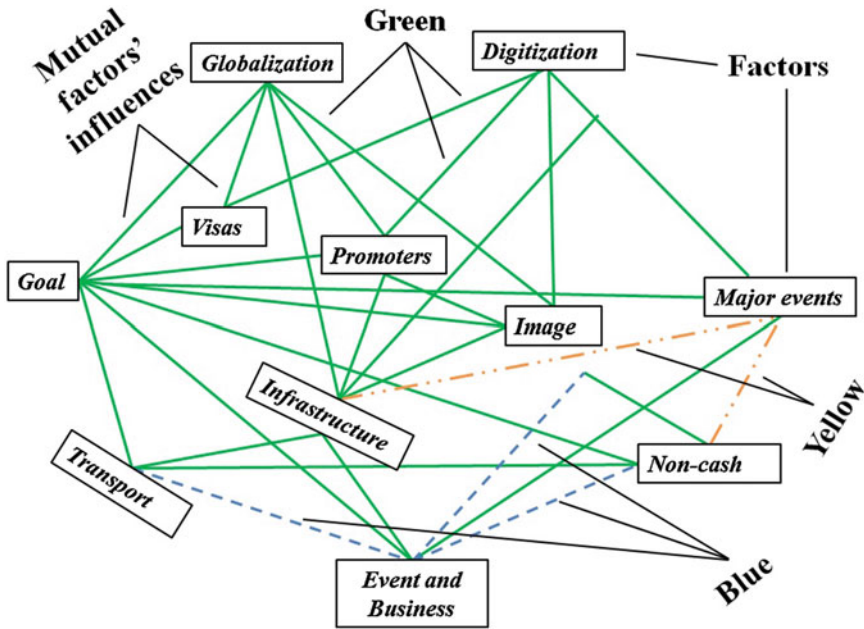


Fig. 3 Result of verification of a fragment of the cognitive model (there are the number of factors in the nodes of the graph, see Sect. 3)

5 The Results of Modelling

The optimization modelling was carried out for determining the values of control factors required to achieve the strategic goals of tourism development in the context of getting the pessimistic and optimistic scenarios (Table 1).

Table 1 Results of optimization modelling (genetic algorithm)

№, Scenario	Goal Value	Factor: value		Recommendation
		First decision	Second decision	
Pessimistic scenarios	-0.87	2: -0.25 4: 0.6 8: -0.6 11: -0.1 14: 0.2	2: 0.7 4: 0.3 8: -0.5 11: -0.3 14: -0.2	Strengthen factors 2 and 4, while weakening factor 8
Optimistic scenarios	+1	2: 0.7 4: 0.7 8: 0.3 11: -0.7 14: -0.2	2: 0.5 4: -0.9 8: 0.7 11: 0.3 14: -0.4	Strengthen factors 2, 4, and 8

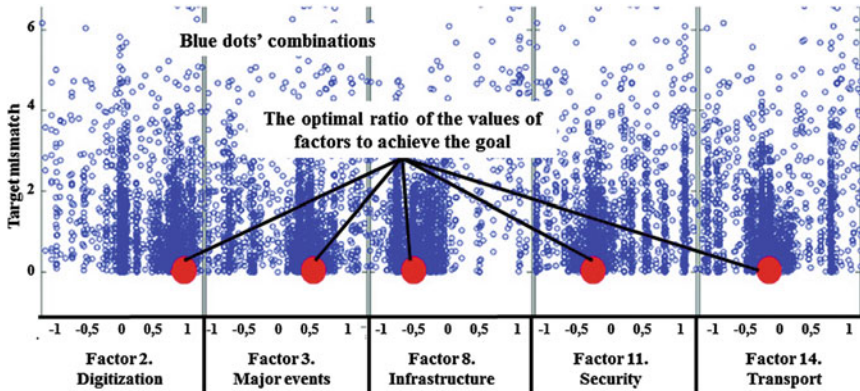


Fig. 4 Optimistic management scenario for 5 controlling factors

The method of inverse problem solving with the genetic algorithm works as follows. The goal value is setting, and the combination of controlling factors values is automatically calculating. Selection of a set of values of controlling factors in computer simulation, in which the goal value “+1” will be achieved, is considered as a complete achievement of goals. Setting the goal value “+1” in this simulation was considered as a search for the optimistic scenario. Setting a negative target value on the interval from “-1” to “-0.5”, for example, “-0.87”—corresponds to the pessimistic scenario decisions, that is, a scenario in which the combination of the values of the controlling factors will worsen the situation. Assigning the value of the target factor in the interval from “+0.6” to “+1”, for example, “+0.75”, corresponds to the construction of the baseline scenario.

The launch of the optimization modelling mode demonstrates the results, example of which is shown in Fig. 4. These are bubble charts that reflect the result of the cognitive modelling with genetic algorithm. Each of the five vertical columns corresponds to one of the controlling factors. Each combination of blue dots that are on the same horizontal level corresponds to one solution of the direct problem, which does not give a required value of the goal. There may be several thousand such decisions. The genetic algorithm for solving the inverse problem is constructed in such a way that each subsequent solution gives a better approximation to the target value, that is, the genetic algorithm builds converging solutions to the goal. The final decision—a combination of the values of the five controlling factors—is indicated by a red (big on Fig. 4) dot. With the values of the controlling factors corresponding to the red points, the target value of the goal factor is reached. For example, in the implementation of the optimistic scenario, the strongest impacts should be carried out by the same factors as in the pessimistic scenario, but in an effort ratio of 0.7:0.7:0.3.

The modelling can also help to assess risks of implementation of various tourism development scenarios. For example, let us illustrate this on comparative evaluation of two risks: “Banning of monetary transactions” and “Strengthening the visa regime”. To do this, the risks’ factors were defined as controlling factors, then an

extremely pessimistic scenario was set (the goal value is “-1”), and the inverse problem was solved. The behaviours of the controlling factors were received. It helps to get various conclusions, the main one of which states that even if anybody tries to improve the meaning of the controlling factors, the result will depend on values’ changes in other factors.

For example, the modelling shows that insufficient attention to risk “Strengthening the visa regime” while simultaneously improving the factors associated with the development of event tourism and payment for services, with insufficient attention to the development of the infrastructure factor, will lead to a pessimistic scenario. Additional modelling shows that improving the infrastructure factor can greatly reduce strategic risks.

6 Conclusion

One of the most important conclusions of modelling can be attributed to the fact that different impacts on the same controlling factors can lead to both an optimistic and a pessimistic scenarios. The correct influence on the controlling factors can either be guessed, but it threatens with big risks. The correct impact can be determined fairly accurately with computer cognitive and genetic modelling.

The indicated values of continuum power of cognitive semantics show that the traditionally used logical and even artificial neural network constructions are insufficient in power to cover cognitive semantics. The convergent approach can help to make the process of megalopolis tourism strategic planning much more efficient.

According to the results of convergent cognitive modelling, the following recommendations can be given in terms of the strategic management of tourism activities in the megalopolis:

- It is advisable to conduct regular convergent cognitive modelling of strategic decisions to reduce risks of decisions and increase the likelihood of an optimistic scenario for the development of tourism activities;
- Create a distributed expert-analytical decision support system using cognitive modelling and verification of cognitive models mapping its components on relevant Big Data;
- Provide storage and end-to-end access to goal setting, forecasting, planning and programming documents, as well as documents reflecting the results of the implementation of decision-making with the cognitive modelling;
- Ensure the reengineering of the cognitive model when the external conditions for the development of tourism activities are changed;
- Use the AI with an emphasis on cognitive semantic, cognitive modelling, convergent approach, and genetic algorithm;
- Create a system for maintaining a knowledge base of the best practices of strategic decision-making in the field of tourism development;

- Ensure the maintenance of a register of typical factors and cognitive models, which reflects the possibilities and limitations of tourism activities, compliance with the megapolis development priorities, a summary and detailed description of the cognitive modelling, etc.

Acknowledgements Real practice implementation and finance support: Boston Consulting Group, Russian Foundation for Basic Research, Grant 18-29-03086.

References

1. UNWTO, <http://cooperation.unwto.org/technical-product/tourism-development-master-plans-and-strategic-development-plans>. Last accessed 13 Dec 18
2. Z. Avdeeva, S. Kovriga, Cognitive approach in simulation and control, in *Proceedings of the 17th World Congress "The International Federation of Automatic Control"*, Seoul, Korea, 6–11 July 2008, pp. 1613–1620. <https://doi.org/10.3182/20080706-5-kr-1001.00275>
3. A. Raikov, A. Ermakov, A. Merkulov, Self-organizing cognitive model synthesis with deep learning support. *Int. J. Eng. Technol. (IJET)* 7(2), 168–173 (2018). <https://doi.org/10.14419/ijet.v7i2.28.12904>. (Special Issue on Computing)
4. H. Atmanspacher, Quantum approaches to brain and mind. An overview with representative examples, in *The Blackwell Companion to Consciousness*, ed. by S. Schneider, M. Velmans (Wiley, London, 2017), pp. 298–313. <https://doi.org/10.1002/9781119132363.ch21>
5. A. Raikov, Cognitive modelling quality rising by applying quantum and optical semantic approaches, in *18th IFAC Conference on Technology, Culture and International Stability*, Baku, Azerbaidshchan, 13–15 Sept 2018, pp. 492–497. <https://doi.org/10.1016/j.ifacol.2018.11.309> (PapersOnLine 51–30)
6. A. Raikov, Accelerating technology for self-organising networked democracy. *Futures* 103, 17–26 (2018). <https://doi.org/10.1016/j.futures.2018.03.015>

Design and Implementation of Ancient and Modern Cryptography Program for Documents Security



Samuel Sangkon Lee

Abstract Encryption technology is to hide information in a cyberspace built using a computer and to prevent third parties from changing it. If a malicious user accesses unauthorized device or application services on the Internet of objects, it may be exposed to various security threats such as data leakage, denial of service, and privacy violation. One way to deal with these security threats is to encrypt and deliver the data generated by a user. Encrypting data must be referred to a technique of changing data using a complicated algorithm so that no one else knows the content except for those with special knowledge. As computers process computations that can be done at a very high speed, current cryptographic techniques are vulnerable to the future computer performance improvements. We designed and implemented a new encryption program that combines ancient and modern cryptography so that the user never knows about data management and transmission. The significance of this paper is that it is the safest method to combine various kinds of encryption methods to secure the weaknesses of the used cryptographic algorithms.

Keywords Ancient cryptography · Modern cryptography · Shift encryption · Polyalphabetic substitution · Transposition cipher · Nihilist encryption · DES and AES encryption · MVVM

1 Introduction

For a long time, cryptography has played an important role in safely storing and transmitting information. From the history of ancient cryptography, it can be observed that people have been discussing the importance of cryptography and creating systems for transmitting secret information since antiquity. These encryption technologies led to the creation of modern cryptography systems, and as modern computer technology is developing at an accelerated pace, the amount of information that must be stored by creating specialized systems is increasing rapidly. In the past, encryption

S. S. Lee (✉)
Jeonju University, Jeonju, Chonbuk 55069, South Korea
e-mail: samuel@jj.ac.kr

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_13

was mainly used to store information within a particular organization or transmit information from one organization to another, but the spread of Internet technology has created the need for encryption technology for transmitting information between individuals. However, so far, programs that allow individual citizens to easily create their own cipher texts using a cryptography algorithm have not become widespread.

The goal of encryption is to prevent unauthorized third parties from acquiring or reading one's information. However, general users do not know how to encrypt their data, or they are not aware that they must use certain encryption methods to be safe and do not know how to use these methods; hence, it is not easy to encrypt documents for personal purposes. Therefore, even mid-level users store documents that they intend to safeguard from others without encryption. In addition, the cryptography used by current users employs encryption programs that are recommended and supported by the government or verified organizations. The encryption techniques provided this way mostly use published algorithms, and they can be decoded by using a high-performance computer. The goal of this study is to design and implement a new encryption system that combines ancient and modern cryptography so that third parties can never decode a user's data when it is transmitted and to make it possible for anybody to easily use the encryption program.

To say that modern encryption technology has evolved implies that it was once possible for encryption technology to be decoded by third parties. Therefore, if a vulnerability is discovered in an encryption method currently being used, a better encryption technique will be developed. Through these efforts, existing encryption methods have continuously evolved from antiquity to the present. Initially, cryptography began with the substitution encryption method, which simply substituted one letter for another letter. Subsequently, transposition methods, which changed the order and position of letters simultaneously by moving them to different positions, were developed. Currently, ancient cryptography can easily be decrypted by using the impressive power of computers. To improve upon these methods, cryptography methods that use mathematical algorithms were created. As the performance of computers has improved, encryption methods have become more difficult, and decryption technology has also been developed. The developed encryption methods have evolved in a direction that depends upon the power of computing. Computers can perform calculations that humans can perform but at a very fast speed, and cryptography technologies that use this advantage are impressive. However, this has led to the conclusion that, as the performance of computers improves, it is inevitable that cryptography technologies will become vulnerable. This fact can be considered a limitation of encryption algorithms that depend on computer performance. As the algorithms used in encryption become more difficult, better performance of computers is required, and computers with higher specifications are required to perform decryption.

Most current encryption techniques depend upon mathematical encryption algorithms, but most of these algorithms are public. If an encryption technique is public, anybody can create a decryption algorithm if they decide to. Nevertheless, even though the algorithms are public, their stability has been confirmed to the extent that they cannot be easily decrypted, and hence, these encryption algorithms continue to

be used. However, even though they are verified algorithms, if the performance of computers continues to improve in the future, it cannot be known whether they will be safe forever. One method for resolving this problem is to increase the length of the encryption key. However, even if an encryption algorithm with a 128-bit key length is used, decryption is not impossible if a quantum computer is used. Quantum computing has not yet become common among average users, but as the performance of computers improves, we still cannot ignore the possibility that encryption algorithms that have been used in the past will be decrypted [1].

In addition, authentication methods that use biometric recognition (which is an encryption technology currently attracting attention) have a vulnerability in that camouflage authentication may succeed if a sufficient number of attackers join together and attempt to pass the biometric authentication [2]. As such, encryption that uses biometric authentication is likely to be combined with other traditional encryption methods. Methods that combine several types of encryption techniques are the safest way to compensate for the vulnerabilities of cryptography methods currently being used. This idea is the main concept behind this study.

This study uses existing algorithms, but it does not depend solely on existing algorithms, and it focuses on reducing the possibility of unauthorized decryption in the future. It does so by combining classical and modern encryption techniques. The best method for verifying an algorithm's performance is to make the algorithm public so that it can be tested. Algorithms that have been verified this way can be selected according to the user's needs, and documents can be stored using a multiple encryption method that combines ancient and modern methods. To perform decryption, the maximum combination of ten encryption methods selected by the user must be known in inverse order, and each key value must be known. This method significantly increases the strength of the encryption compared with existing methods in that it performs multiple encryption through cryptography algorithms selected by the user from among encryption methods that depend on existing algorithms. The following chapter describes each of the encryption methods.

2 Theoretical Background of Each Encryption Method

2.1 Ancient Cryptography

The shift encryption method, which is a part of ancient cryptography, is a relatively simple method for creating cipher texts by shifting the 26 letters in the English alphabet either left or right by the amount of the shift value (numerical), which is the key. For example, the 26 letters of the English alphabet have been shifted by two places. If this method is to be used for encryption, the key is applied to all the input text sequentially, and the shift encryption is complete.

In this study, the shift encryption method described above was expanded to include the 11,172 characters of the complete form of the complete Hangeul text used in

Korea, as well as special characters (44) and English uppercase and lowercase letters (26 + 26) for a total of 96 additional characters (44 + 26 + 26). Related published books use only English letters and describe the implementation principles for such a system, but this study includes Hangeul so that Korean citizens can explore the encryption and decryption processes. The special characters, English uppercase and lowercase letters, and complete Hangeul are converted to numerical ASCII codes and Unicode values. The spaces between the number values are removed, and the shift encryption method is applied. Subsequently, an operation is performed to insert the same number of spaces that were removed from between the number values before encryption, so that only characters within the possible input range are outputted. A unique feature of the algorithm in this study is that it was designed so that, when a positive integer is entered as the key value, a shift to the right is performed, and when a negative integer is entered, a shift to the left is performed.

Vigenere cipher is the most typical form of “polyalphabetic substitution (PS),” as referred to by researchers. This encryption method is also a classical encryption method like the previously described existing method. Usually a 26×26 grid for the 26 letters of the English alphabet is filled out so that the letters are moved by one position for each row and column. The key used for encryption and the input text, respectively, forms the horizontal and vertical coordinates to encrypt the document [3].

Encryption with PS is usually based on the 26 letters of the English alphabet, but similar to the shift encryption above, the 11,172 characters of complete Hangeul and the 96 (= 44 + 52) special characters and English uppercase and lowercase characters were added. The characters of the input text are each converted to Unicode values and the code values not used in the program are excluded to form a continuous numerical string. The same operation is applied to the key value, the value is added to the calculated input text, and this is restored back to characters. The restored text forms the cipher text. To summarize, this study uses the concept of a cipher text from the existing PS method, which is created with only English letters, but it implements an algorithm expanded to include Korean, English, and special characters to create the cipher texts. A unique feature of this study is that it has developed a program that properly encrypts all characters when encrypting all text in a document written in Korean.

PS is a standard substitution encryption method that is a part of classical encryption, and it can be classified according to the block encryption techniques [4], which use blocks. The input text to be encrypted is entered (by row units) in the block, which has as many columns as the key value amount (numerical), and this is outputted by column units to create the cipher text. The key can create completely different cipher texts according to its block size. This method has the advantage of increasing the strength of the encryption by creating completely different cipher texts when the size of the key is changed. In other words, when the key value is 4, a $4 \times Y$ block is created, and the input text to be encrypted is entered. Subsequently, the output text is created by column units (vertical columns) in the block to complete the cipher text.

This method is called a “standard transposition cipher (STC),” and the program in this study was designed so that the aforementioned style of transposition cipher can

be used. The method in this study was designed so that the STC, can be combined with English characters, complete Hangeul, and special characters.

The key transposition cipher (KTC) is similar to the previous method in that it uses a block encryption method like the STC. However, the program designed in this study provides an improved encryption method that allows character key values, which increases its level of analysis strength compared with the existing standard transposition method, which limits the entered key values to numbers. The advantage of the KTC is that numbers, letters, and special characters can be used as key values. In this encryption method, a block with as many columns as the length of the entered key value (character) is created, and similar to the STC, the input text is entered. The key value (text) is substituted with numbers that correspond to the alphabetical order of its characters, and the output is created starting at the column with the smallest number to complete the cipher text.

2.2 *Modern Cryptography*

In 1975, the United States' National Bureau of Standards established a type of block encryption [4] called the data encryption standard (DES) as the standard for encrypting data. This is the encryption selected as the national standard by the United States' National Bureau of Standards (NBS, currently the NIST). This technology is a symmetric key encryption method, which uses a 56-bit key; in the past, it was used without issue, but it has problems in that its key length is too short and it can be defeated with special techniques if it includes a backdoor. As such, the vulnerabilities of DES became known after a short time, and currently, it has been replaced by advanced encryption standard (AES).

The original name of AES-256 was Rijndael encryption, and it is a type of modern encryption method. As the United States' national standard encryption method, AES-256 was designed to compensate for the vulnerabilities of DES, which was previously used as the national standard encryption method, as explained previously. AES-256 is known as one of the strongest forms of encryption currently being used [5]. AES is a 128-bit block encryption method newly developed to resolve the problems of DES encryption technology. AES still uses a symmetrical key technique in which the same key is used in the encryption and decryption processes, but its main difference from DES is that it can process 128-bit blocks with 128-, 196-, and 256-bit keys.

AES is used as one of the encryption methods for Wi-Fi, the most popular short-distance wireless communications technology. In the case of Bluetooth, the SAFER+ encryption technology was used until version 3.0, but since Bluetooth 4.0, the AES standard has been used. This study has implemented the modern cryptography methods of DES and AES, and it used the System.Security.Cryptography C# library to develop a more effective and accurate encryption program and to implement it easily. This library not only performs tasks such as encoding/decoding hashes, random number generation, and message authentication, but also provides methods for easily

implementing encryption techniques that include data. In addition, this program uses AES-256 to use 256-bit key values.

3 Design of Encryption Method

The purpose of this system is to implement functions that encrypt documents with a combination of ancient and modern cryptography and save the documents and, conversely, decrypt saved documents and save them in their original plain text. Figure 1

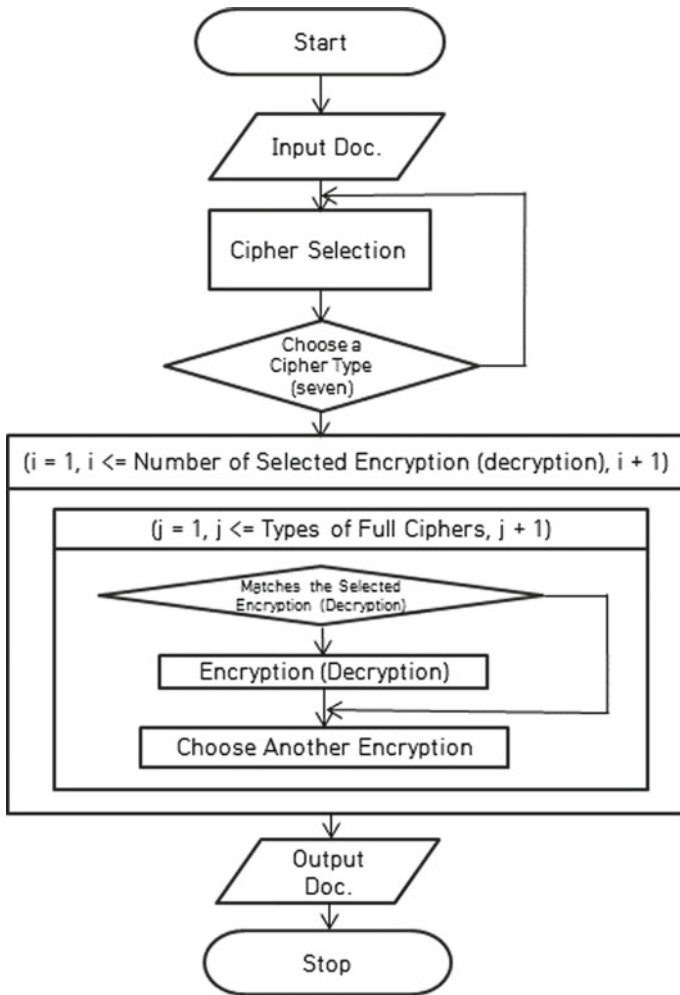


Fig. 1 Overall control

shows the structure of the entire system. The requirements for the functions are as follows: The document-opening function allows text files to be imported and displayed on the screen. The encryption type selection function allows up to ten encryption methods to be repeatedly selected to create a hybrid encryption method. This hybrid encryption method allows for seven types of ancient and modern cryptography methods to be selected repeatedly up to 10 times for mixed encryption. This paper distinguishes itself from other papers in that several types of encryption can be combined in various ways.

When entering the key values that will be used in the encryption, it must be possible to enter each of the key values to be used according to the type of cryptography. Here, the cryptography algorithm types and key lengths are important factors in ensuring the level of system safety. The Korea Internet Security Agency (KISA) recommends using an individual key length of at least 160 bits and a valid usage period of maximum two years. The seven types of encryption supported are Shift, PS, STC, KTC, Nihilist, DES, and AES-256. It is possible to use these seven types in combination, and their order of use can be changed. The resistance of the encryption varies significantly according to the order in which the methods are used. The encryption output (displaying the encrypted text on the screen), cipher text storage (saving the cipher text as a text file), and document decryption (decrypting the cipher text and restoring the original document) functions are all integrated into the program.

From a software engineering perspective, the quality requirements of the program are as follows. It must perform correct encryption according to the input and output (it must have correct functions for importing input text and cipher texts, and there must be no data loss) as well as the encryption types. It must be designed so that, if the original text is modified even slightly, decryption is impossible. When decrypting the cipher text, the decryption function must perform correct decryption. Figure 1 shows a flowchart of the overall encryption/decryption process of the program. A description of the use-case actor details and the outline details are provided as follows. The program has an actor known as the “user,” as shown in Table 1. Use-cases diagram is used to easily explain the scope and functions of the system from the user’s perspective. Use-cases gather together the purposes for which the system is used to create a multi-purpose system. By gathering together use-cases and connecting them as a system, the user can easily understand the development process. The program’s use-cases include Open, Selection, Key Registration, Encryption, Save, and Decryption. Table 2 presents use-case scenarios.

The sequence diagram was created using Web Sequence Diagrams. The sequence diagram describes the sequence of message exchanges between interacting objects,

Table 1 Specification of the use-case “actor”

Actor	Description
User	<ul style="list-style-type: none"> • The user is the main agent who uses the program • The user can use the program to open and save text • The user can use the program to select or add document encryption methods • The user can use the program to encrypt documents

Table 2 Scenario of the use-case

3.3.1 Open			3.3.2 Selection		
A	Overview	The user can open document files from within the program	A	Overview	The user can select and add document encryption methods
B	Relation	<ul style="list-style-type: none"> – Initiator: User – Supporters: None – Pre-condition: None – Post-condition: The text of the file selected by the user is displayed on the screen 	B	Relation	<ul style="list-style-type: none"> – Initiator: User – Supporters: None – Pre-condition: None – Post-condition: The encryption method selected by the user is added
C	Basic flow	C-1. Select Open File in the program C-2. Select the file to be opened C-3. The text of the document file is displayed on the screen	C	Basic Flow	C-1. Select the encryption type to be used in an encryption list combo box C-2. Click the select button to add the encryption
D	Alternative flow	None	D	Alternative flow	None
E	Exception flow	When a file is not selected 1. Cancel file selection and return to the initial program	E	Exception flow	When the maximum number of selections is exceeded, an error message is shown
3.3.3 Key registration			3.3.4 Encryption		
A	Overview	The user can enter the key to be used according to the selected encryption method	A	Overview	The user can encrypt the document according to the selected encryption method
B	Relation	<ul style="list-style-type: none"> – Initiator: user – Supporters: – Pre-condition: Select/add encryption – Post-condition: The key value to be used in the encryption is registered 	B	Relation	<ul style="list-style-type: none"> – Initiator: User – Supporters: – Pre-condition: Import document, Select/add encryption – Post-condition: The encrypted document is displayed on the screen

(continued)

Table 2 (continued)

C	Basic flow	C.1. Click the added encryption button C.2. Enter the key value	C	Basic flow	C1. Click the Convert button to convert the cipher text. There are no alternative or exception flows
D	Alternative flow	None	D	Alternative flow	If the key value is not entered, an input window appears
E	Exception flow	If a key value that violates the rules is entered, an error message is displayed	E	Exception flow	If the encryption to be used is not selected or the key value is not entered, an error message appears
3.3.5 Save			3.3.6 Decryption		
A	Overview	The user can save encrypted/decrypted documents as files	A	Overview	The user can decrypt documents according to the selected encryption method
B	Relation	<ul style="list-style-type: none"> - Initiator: User - Supporters: - Pre-condition: Encryption/decryption - Post-condition: The document is saved as a file 	B	Relation	<ul style="list-style-type: none"> - 0 Initiator: User - Supporters: - Pre-condition: Open document, Select/add encryption - Post-condition: The decrypted document is displayed on the screen
C	Basic Flow	C.1. Click Save C.2. Enter the file name at the desired path C.3. Click the Save button	C	Basic flow	C.1. Enter a key value according to the encryption method C.2. Click the Restore button to convert the cipher text
D	Alternative flow	None	D	Alternative flow	None
E	Exception flow	E1. If the save is canceled 1. Return to the initial screen	E	Exception flow	If the cipher text has been changed or the key value is incorrect, an error message is displayed

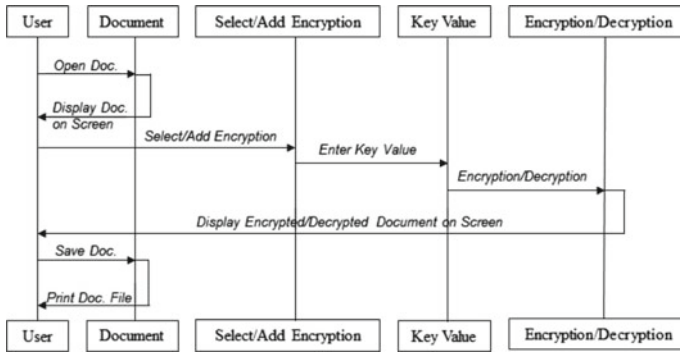


Fig. 2 Sequence diagram

and it presents a unified modeling language (UML) diagram. UML is used in models that establish words and rules to describe systems conceptually and physically in order to create software. This helps in resolving structural problems in a system, problems with decision-making within the project team, and software structure reuse problems. The design content used in the program in this study is presented as a sequence diagram in Fig. 2.

4 Implementation

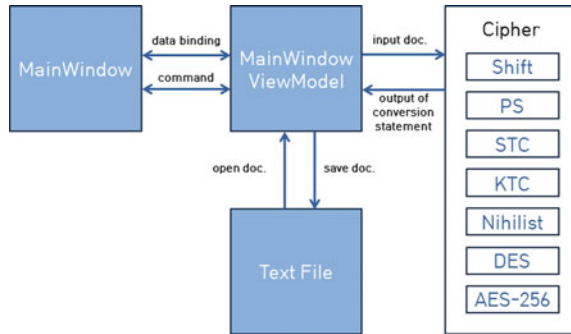
4.1 Development Environment and Implementation Tools

The development environment is shown in Table 3. The program is an application program that runs in the Windows environment, and hence, a language and tools optimized for Windows applications were used, and the program was run in the Windows 10 environment. It was developed in the C# language, which is optimized for Windows applications, and the design was developed via eXtensible Application Markup Language (XAML) using Windows Presentation Foundation (WPF). Visual Studio 2017 Community RC was used as the development tool. As for the execution environment, the program was designed to operate correctly on a virtual machine with the .NET Framework 4.6 or higher.

Table 3 Development environment

Environment	MS-Windows 10
Language	C#, WPF (XAML)
Tools	Visual Studio 2017 Community RC
Execution environment	.NET Framework 4.6

Fig. 3 System architecture



4.2 User Interface

The screen design was created to be a familiar design that users can use easily. A menu is located at the top, and below this, a toolbar is used to quickly access menu commands. The center of the program screen is divided into the window that displays the input text and the window that displays the cipher text. A log Window is located at the bottom so the user can check if the program is running, as shown in Fig. 3.

The menu bar contains two menus: File and Info. The file menu contains three items: Open, Save, and Close. The first toolbar contains a collection of functions related to file input and output. The second toolbar uses a combo box to display a list of encryption methods. It also implements functions by which the encryption type can be selected or the list can be reset by pressing the Select and Clear buttons. The third toolbar contains functions related to encryption conversion. It contains five buttons: Encrypt, Input ← Output, Decrypt, Reset Input, and Reset Output.

When the program was developed, the important design patterns were developed using a Model-View-ViewModel (MVVM) model. Compared with the conventional Model-View-Controller (MVC) or Model-View-Presenter (MVP) patterns, the user interface (UI) in this model is efficient in dealing with UI accessibility or analyzing program source code. In the conventional MVC pattern, the controller directly accesses and controls the views and models, and the views can be indirectly accessed from the models. In the MVP pattern, the views and models are accessed by the presenter. However, the MVVM pattern used in this program is designed with three structures—Views, Models, and ViewModels. The views lay out the design, the models process the data, and the ViewModels use the information of the models to control views through data binding. When the view on the left side of the figure was designed, the code (which is composed in a language useful for development and UIs composed in XAML) was designed in code behind to create the UI logic. In the notifications at the center of the image, the view is operated by the notifications of ViewModels, and if this kind of operation occurs, a data binding information exchange is achieved between the views and ViewModels; furthermore, the commands are executed simultaneously. Speaking of efficiency, the ViewModels corresponding to the presentation logic control the entire part that displays the views. The model on the right side of

the figure, which corresponds to the business logic and data, manages the saving, modification, and deletion of information.

The MVVM pattern is more modern than the previous two design patterns, and it provides an evident distinction between domain logic and the presentation layer so that the views from the user's perspective and the models can be separated to achieve efficiency. In addition, MVVM provides clean separation of code and has good maintainability. The separation of core logic parts by external and internal dependencies (i.e., unit testing of core logic) is easy, and program testing is easy. MVVM is good for code reuse, code exchange, and adding code to areas suitable for the software architecture. Simultaneously, it can perform abstraction by hiding code. As mentioned above, the MVVM model was used in POWINZ_CC6 (the program name) to design a more efficient program.

Figure 3 shows the structure of the system. First, the main window, which is used as a view, was created, and WPF's XAML6 was used to create the source code. The main window view model (which becomes a ViewModel) was created, and data binding with the main window was performed so that, if the data in the main window view model changes, it can be immediately reflected in the view. In addition, commands were used to transmit user commands between views and ViewModels so that each function can operate properly. In the main window view model, the text file that corresponds to the model can be opened or saved. In addition, the system was implemented so that encryption algorithm files that are implemented in the Cipher on the right side of the image can be accessed to use the encryption functions.

The program employs regular expressions for processing text strings and string internal operations. Regular expression is effective at preprocessing the input text. The regular expressions only allow the input of complete Hangeul and English letters (including special characters). In the case of the transposition encryption method, the text string internal operations are not required in the input text, but they are required for the key value. The substitution encryption method requires string internal operations for both the input text and key values. The shift encryption method and the STC allow only numbers as the key value regardless of the format of the input text; hence, the entered key value must be converted into an integer type, which is a number type, to perform operations. As input, the key value of the shift cipher takes plus or minus values, which indicate difference directions. The input text is converted to ASCII code or Unicode, and internal operations are performed. These internal operations were implemented to substitute the numbers with characters in reverse fashion after the encryption process is finished and the cipher text is to be displayed. This is performed in the same way in not only the encryption process but also the decryption process.

5 Conclusions

When using this program to encrypt text, it is expected that there will be considerable differences in the encryption strength when using the AES-256 encryption algorithm for one round of encryption, when using it for two rounds of encryption, and when also combining it with the various ancient ciphers for encryption. This indicates that it is possible to create ciphers that are safer and stronger by using a combination of ancient and modern cryptography, unlike existing encryption methods, which rely on the performance of computers. Data encrypted in this way through a combination of several ciphers cannot be decrypted to plain text through the decryption algorithm used when decrypting AES-256. It can only be decrypted by applying all the encryption methods used during encryption in reverse order. However, when the user combines several types of encryption methods in various orders to create a custom encryption algorithm, the number of cases that must be considered when decrypting the data is exponentially increased, which indicates that the strength of the data security can be considerably increased. This indicates that this method cannot be easily decrypted even as the performance of computers improves in the future. The encryption method presented in this paper does not use a single algorithm but combines several types of methods that have already been verified to make safe encryption possible.

References

1. Wikipedia (DES), [https://ko.wikipedia.org/wiki/DES_\(encryption\)](https://ko.wikipedia.org/wiki/DES_(encryption))
2. WebSequenceDiagrams, <https://www.websequencediagrams.com>
3. MSDN, MVVM examples, <https://msdn.microsoft.com>
4. Regular Expressions, RegExr v2.1, Regular Expression: Learn, Build, and Test Reg. Ex., <http://regexr.com>
5. MSDN, WPF examples, <https://msdn.microsoft.com>

Methodology for Selecting Cameras and Its Positions for Surround Camera System in Large Vehicles



S. Makarov Aleksei and V. Bolsunovskaya Marina

Abstract This article describes the methodology for selecting cameras and its location for a surround camera system in large vehicles. The methodology includes selecting a camera using a database that stores information about vehicle model, its dimensions, angle of camera view, angle of camera tilt, area of blind zones, and image sharpness. In addition, examples of cameras location and images from them are shown.

Keywords Surround camera system · Trucks · Wide-angle cameras

1 Introduction

The surround camera system in the cars is an auxiliary to driver and displays inaccessible to the driver through the rear-view mirrors and windows areas around the vehicle on the monitor located in passenger compartment. We are developing a surround camera system, which is a hardware–software system consisting of several (4–6) cameras, a computer system and a monitor, which shows information from cameras as a bird’s-eye view. In large-sized vehicles, dead zones are much larger. Therefore, when designing such systems, there is a question of correct location of cameras on vehicle body. There are three parameters required to select correct camera position:

1. Position on vehicle body.
2. Height of location.
3. Angle of camera tilt to surface of earth.

S. Makarov Aleksei (✉) · V. Bolsunovskaya Marina
Peter the Great St. Petersburg Polytechnic University (SPbPU), Polytechnicheskaya 29,
195251 St. Petersburg, Russia
e-mail: lyohamakarov@yandex.ru

V. Bolsunovskaya Marina
e-mail: bolsun_hht@mail.ru

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_14

On the first two points, the question is solved depending on the model. Cameras are usually located: in front (on radiator grille), rear (above license plate), and on sides (on rear cab racks). From the choice of angle of tilt in combination with first points, two output parameters depend: a camera's area of view and a quality of output image in pixels/m. In addition, area view of camera depends on its viewing angle. That is, should choose the right camera. Then a question arises, how to make this choice correctly. We propose a methodology for choosing it.

2 Input Parameters

To solve the issue of camera selection and its position relative to earth's surface, we offer a database containing the following data:

1. Model of vehicle. Therefore, its dimensions are stored: length, width, height, and cameras position on vehicle body.
2. Angle of camera view.
3. Angle of camera tilt.

2.1 *Model and Dimensions of Vehicle*

In our methodology, all cargo vehicles we divide into three categories depending on body type:

1. Vans.
2. Flatbed/dump trucks.
3. Saddle tractors.

Number of cameras and its position depends on type of vehicle. So, for vans and flatbed/dump trucks, four cameras are required: front (above windshield), rear (on top of van, and on board, respectively) and on the sides of cab at rear of it. An example of location of cameras is shown in Figs. 1 and 2 [1]. Red color indicates cameras located in plane of picture and blue color indicates cameras located in perpendicular plane.

For surround camera system, six cameras should be used in saddle tractors (Fig. 3) [2]. Three of them are located on front (number 1 in figure) and on sides of cabin (2–3). Two are on trailer—rear (6) and on left side in direction of traffic (5). Side camera on trailer serves to increase area of view and improve picture when cornering (one of problems with introduction of surround camera systems in trucks is turning parts). Camera from right side is considered unnecessary. In addition, another camera (4) is located on rear of truck and is used when a trailer is absence.



Fig. 1 Location of cameras on flatbed truck



Fig. 2 Location of cameras on van

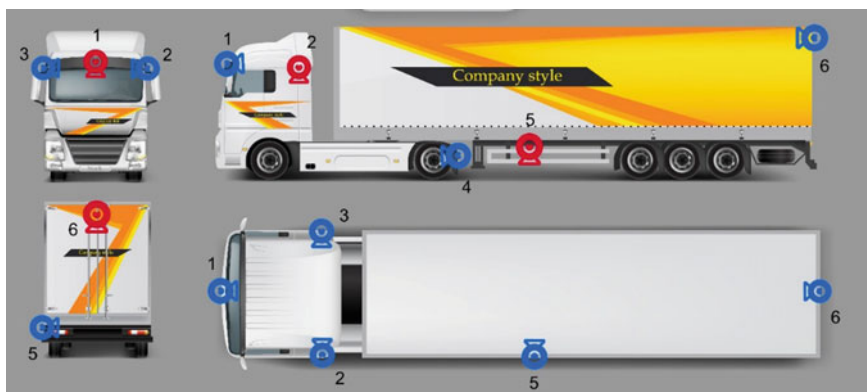


Fig. 3 Location of cameras on saddle tractor

The above location of cameras is recommendatory, and final position depends on dimensions of vehicle and camera model. The dependence on these parameters is described later in this article.

2.2 Angle of Camera View

Angle of camera view is a value that determines the area that falls within field of view of camera. Angle of camera view depends on such criteria as focal length of camera lens and size of matrix. Because cargo vehicles are preferably large, then to obtain the largest area of vehicle environment, wide-angle cameras, including fish-eye type, should be used. From this parameter, depends on the completeness of picture of bird's-eye view from cameras. Using wide-angle cameras allow to minimize number of blind areas. But to get rid of barrel distortion, additional transformations are required.

2.3 Angle of Camera Tilt

Angle of camera tilt to surface of earth is also an important parameter. The quality of image depends on it in combination with viewing angle. This will be written in the next section.

Since surround camera system can be used in the dark, problem of appearance of glare from light sources may appear on image, which will affect its quality and area of camera's view. We have made experiments to determine effect of camera tilt angle to light source on appearance of glare in the image. Results showed that serious distortion is practically impossible, because glare occurs only at small distances to light source, and camera's direction to plane of propagation of light at an angle close to zero. Results were obtained for a camera with a viewing angle of 150° , and its schematic image is shown in Fig. 4.

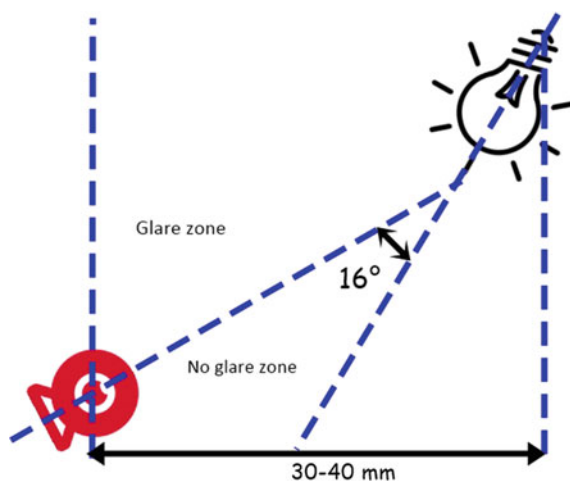
3 Methodology

For correct selection of camera and its position on vehicle body, we offer a database that stores the above parameters. It is a model of vehicle, a camera number on vehicle body, an angle of camera view, and an angle of camera tilt. Last two parameters are stored separately for each camera, because they can be located at different heights.

For each combination of input parameters, two outputs are stored:

1. The percentage of area covered by cameras, from total required for visibility.
2. Sharpness of image, pixel/m.

Fig. 4 Glare and no glare zones, depending on angle of camera tilt to light source



The data was collected using the IP Video System Design Tool v.8 [3].

For example, for KamAZ-5320, table for front camera with viewing angle of 150° will look like Table 1. First column contains minimum value of angle of camera tilt and last column is maximum for a particular set of parameters.

In Fig. 5 two graphs are presented: dependence of viewing area on angle of camera tilt and sharpness of image on angle of camera tilt. Optimal angle is an angle corresponding to intersection point of graphs. But in database, not only optimal value is stored, but also rest, so that there is a choice if area of view is more important than image sharpness or vice versa.

In Figs. 6 and 7 scheme of location of camera's viewing areas in environment of cargo vehicle and an example of location of cameras and view that is obtained from its is presented.

4 Conclusions and Further Researches

We have developed a methodology and database scheme that can be used to select cameras and its position on vehicle body for surround camera system in large-sized cargo vehicles. This methodology will be tested by us when installing on a real vehicle. After the installation of cameras, it is planned to collect data from cameras for further analysis and experiments on its.

Table 1 Example of dependency table of camera's area of view and sharpness of image from angle of camera tilt

Angle of camera tilt	59.8	61.0	61.9	62.7	63.3	64.1	64.9	65.8	66.7	67.6
Area of view, %/sharpness, pixel/m	54/21.4	53/28.6	51/32.1	49/35.7	46/42.9	43/50.6	39/59.3	35/71.7	31/82.6	26/97.1

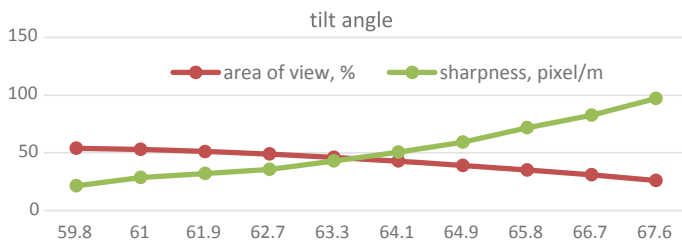


Fig. 5 Graphs of dependence of viewing area on angle of camera tilt and sharpness of image on angle of camera tilt

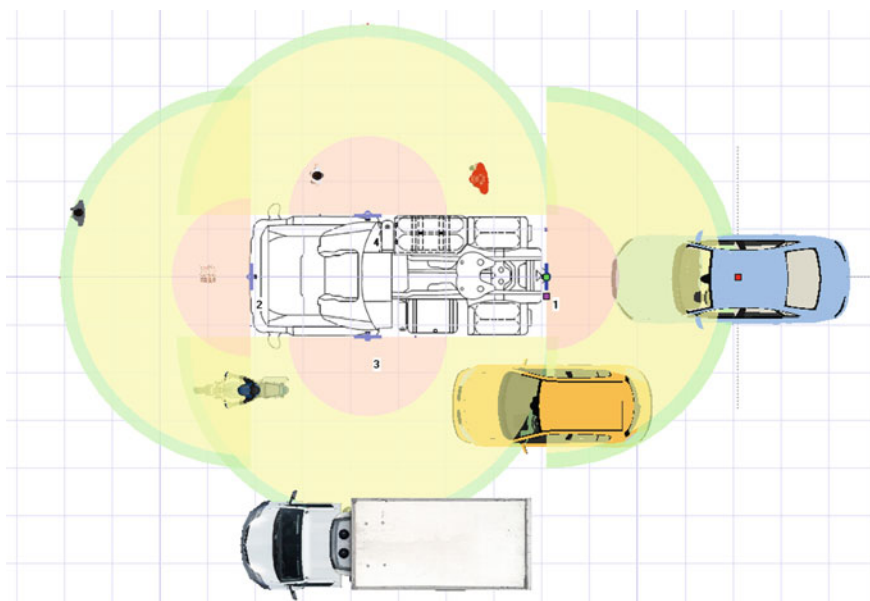


Fig. 6 Example of location of cameras on a cargo vehicle

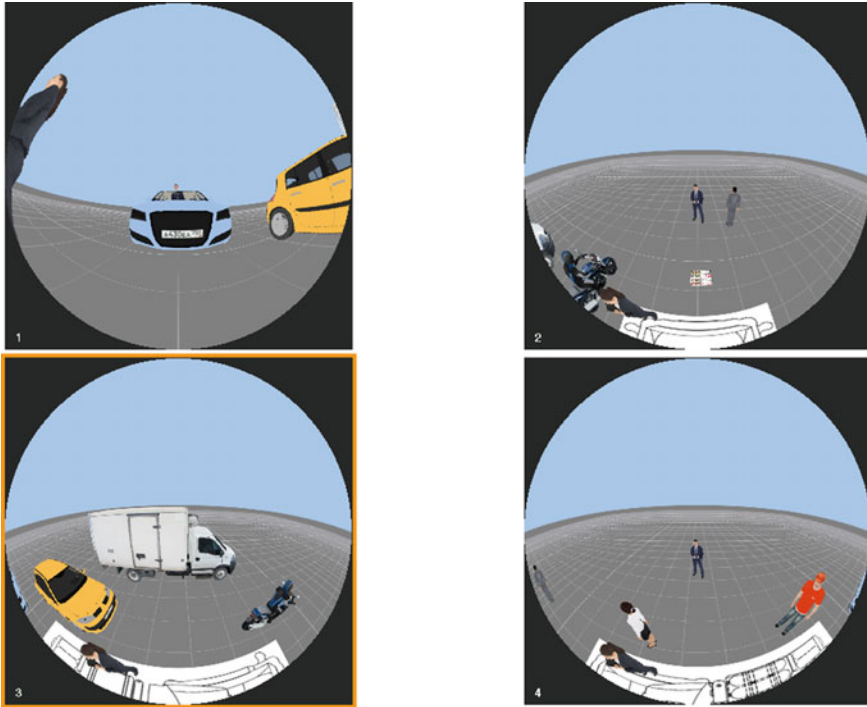


Fig. 7 Views from cameras in surround camera system

References

1. Trucks vector set [Gruzoviki vektornyj nabor], https://ru.freepik.com/free-vector/transportation-trucks-set_717957.htm#term=truck&page=6&position=37, last accessed 2018/09/21
2. Mock-up, brand design for a truck [Maket, dizajn trgovoj marki dlya gruzovika], https://ru.freepik.com/free-vector/mock-up-template-brand-design-for-truck_1472097.htm#term=truck&page=1&position=40, last accessed 2018/09/21
3. IP Video System Design Tool, <http://www.jvsg.com/>. Last accessed 8 Sept 2018

Fuzzy Models and System Technical Condition Estimation Criteria



Gennady Korshunov, Vladimir Smirnov, Elena Frolova
and Stanislav Nazarevich

Abstract The problem of the limited hardware capability of the parametric tolerance control process of the state of technical systems is considered. A more complete assessment of the technical condition of a workable product is necessary to support decision making and reduce risks. An approach to estimating the parameters of systems based on the theory of fuzzy sets to determine the state characterized by considerable uncertainty and incompleteness of information for its modeling by traditional methods is proposed. This approach is applicable to the organization of tolerance control at different stages of the life cycle. This approach uses an additional fuzzy classification of parameter values to increase the reliability of control results, taking into account uncertainty factors. It is proposed to use the working capacity criterion, the criterion for the steadiness of the tendency of the dynamics, the criterion of the rate of change of the parameter, and the complex criterion for working capacity level in addition to the criterion of belonging to tolerance zones. Four fuzzy classifiers have been developed, which allow to take into account the inaccuracy and approximation of the initial information, operate with linguistic criteria and include qualitative variables in the analysis. The procedure for estimating the value of the parameter according to the complex criterion for working capacity level is considered.

Keywords Working capacity · Parametric control · Tolerance field · Estimation criterion · Fuzzy classifiers · Linguistic variable · Membership function

G. Korshunov · E. Frolova · S. Nazarevich

State Autonomous Educational Institution of Higher Education, “Saint-Petersburg State University of Aerospace Instrumentation” (SUAI), ul. Bolshaya Morskaya, 67, lit. A, St. Petersburg 190000, Russian Federation
e-mail: kgi@pantes.ru

G. Korshunov

Ciberphisc Systems and Control High School, Peter the Great St. Petersburg Polytechnic University (SPB STU), 29, Polytechnicheskaya St., St. Petersburg 195251, Russian Federation

V. Smirnov (✉)

Closed Joint Stock Company, Scientific-Production Center “Akvarin”, Tallinskaya St., 7, St. Petersburg 195196, Russian Federation
e-mail: vlad.sm2010@yandex.ru

© Springer Nature Singapore Pte Ltd. 2020

X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_15

1 Introduction

The limited capabilities of the hardware to control the technical condition of the systems do not automatically allow a deeper assessment of the technical condition of the product and cause a lack of necessary information for decision making. The paper presents the results of research on the development and improvement of the method of tolerance control of the state of systems. An additional process of intellectual processing of parametric control results using four fuzzy classifiers is proposed. The proposed model of a parametric assessment of the state of the system based on fuzzy classifiers and the matrix method of data aggregation allows one to take into account the closeness of the parameter values to the limits of tolerance fields, the dynamics of their changes within the limits of the tolerance fields, to obtain an estimate of the parameter state according to the complex criterion for working capacity level, and to increase the reliability of monitoring results' uncertainties. The software and hardware implementation of the model can be integrated into the intelligent decision support system (*IDSS*). Such a system is designed to uncover the uncertainty of the technical state on the basis of an updated knowledge base, which allows for automatic analysis of the initial data and excludes the adoption of erroneous decisions on the results of control with the maximum elimination of the human factor from this process [1].

The system state up at any fixed point in time is characterized by a set of quantitative values of parameters set by technical documentation that are subject to change during production and operation. Parametric instability is due to the continuous change in the properties of the elements of the system due to their aging, wear, and the action of various destabilizing environmental factors. These reasons lead to a change in the working capacity level. A traditional method for evaluating the performance of the product, based on the criteria for the belonging of the parameter values to the corresponding tolerance zones, is implemented in the known systems of automated parametric control [2]. When controlling responsible and expensive systems, such a binary approach does not take into account qualitative assessments of intermediate states of a parameter of a workable system and the individual nature of their dynamics.

2 Fuzzy Parametric Control Models

The binary approach to estimating the monitored parameters does not take into account finding the values of the parameters of a workable product near the border of the tolerance zone and their abnormal behavior, which is a sign of the development of hidden defects or the influence of unaccounted factors. For the timely detection of pre-failure states of the system and the detection of susceptibility to the instability of the working state at the early stages of the development of the defect, it is necessary to record and analyze such changes, which are the forerunner of system

failure. A promising direction for solving the problem of improving the accuracy of assessments of the results of parametric control, the quality of recognition, and evaluation of the technical state of the system is the development of new models, criteria, and algorithms based on the theory of fuzzy sets [3–6].

To improve the reliability of estimates of the results of parametric control, a model of parametric estimation of the system has been developed on the basis of four fuzzy classifiers and the matrix method of data aggregation. Three fuzzy classifiers make it possible to obtain a qualitative assessment of the state of a parameter according to the criteria of working capacity, steadiness of the tendency, of the rate of change of the parameter within the limits of tolerance fields. The results of the classification are the input data for the fourth fuzzy classifier, which evaluates the state of the parameter according to the complex criterion for working capacity level.

To reduce the large dimension of the data set, which are subject to more accurate estimation, it is necessary to choose the most informative set of parameters, taking into account the specifics of the problem area. It is proposed to use parameters which values are beyond the tolerances and a critical failure of the product could occur. The criterion for the selection of such parameters is the permissible level of significance of possible effects.

The results of measurements of the parameter values obtained during various tests come from an automated control system to the *IDSS*. Measured values, which are time series, are converted and stored in a knowledge base in a case-based format. On their basis, the numerical values of the input variables are determined, which are fed to the inputs of fuzzy classifiers.

The use of a fuzzy-set approach for solving the problem of classifying parameters is due to the presence of a high degree of uncertainty in the initial information and the need to mathematically formalize fuzzy expert information [7]. The vagueness of information is related to the uncertainty of the expert that occurs during the classification. The construction of classifiers is based on the results of processing expert opinions and analyzing the existing database of previous test results. The key model formalism is the membership function of a fuzzy subset of a linguistic variable defined on the corresponding real media, allowing you to more fully take into account the qualitative aspects that do not have an exact numerical evaluation.

Since the parameters to be classified have a different range of possible values, a single segment of the real axis $[0,1]$ is chosen as the carrier of linguistic variables. The finite length segments of the real axis can be reduced to the $[0,1]$ segment by a simple linear transformation (1). The selected segment of unit length (relative scale) is universal. Thus, the condition of proportionality of elementary parameters and classifiers constructed on their basis is satisfied.

$$z_n = \frac{z - z_{\min}}{z_{\max} - z_{\min}}, \quad (1)$$

where z_n —normalized variable value z , having a range of possible values from z_{\min} till z_{\max} .

Table 1 Setting term sets for variables A_1 , A_2 , A_3 , and B

Designation	Linguistic variable	Term set
A_1	Estimation of the parameter state by the criterion of dynamic tendency steadiness	EC ₁ —"excellent condition" GC ₁ —"good condition" SC ₁ —"satisfactory condition" DC ₁ —"dangerous condition" PS ₁ —"pre-failure status"
A_2	Estimation of the parameter state by the criterion of working capacity	EC ₂ —"excellent condition" GC ₂ —"good condition" SC ₂ —"satisfactory condition" DC ₂ —"dangerous condition" PS ₂ —"pre-failure status"
A_3	Estimation of the parameter state by the criterion of the rate of change of the parameter	EC ₃ —"excellent condition" GC ₃ —"good condition" SC ₃ —"satisfactory condition" DC ₃ —"dangerous condition" PS ₃ —"pre-failure status"
B	Estimation of the parameter state by the complex criterion for working capacity level	EC ₄ —"excellent condition" SC ₄ —"satisfactory condition" PS ₄ —"pre-failure status"

When constructing fuzzy classifiers, first the names of linguistic variables were defined, term sets were defined and ordered, and membership functions were constructed for each term of the linguistic variable.

The linguistic variables of the model for obtaining comprehensive assessments of parameters are A_1 —"Assessment of the state of the parameter by the criterion of dynamics tendency steadiness," A_2 —"Assessment of the state of the parameter according to the criterion of working capacity," A_3 —"Assessment of the state of the parameter by the criterion of the rate of change of the parameter" and B —"Assessment of the state of the parameter according to the complex criterion for working capacity level." The combination of linguistic values A_1 , A_2 , A_3 , and B presented in Table 1.

The choice of the number of terms of the linguistic variable and the names of the fuzzy variables corresponding to them are made on the basis of the analysis of verbal scales used in practice to measure the degree of confidence and taking into account the need to ensure the minimum degree of difficulty in using the classifier in the control process and maximum consistency of expert judgments in its creation.

The use of membership functions depends on the expert's opinion and is not formalized. In [8], recommendations are given on the areas of application of membership functions. Following [9] to specify fuzzy sets characterizing the uncertainty of the type "located in the interval" corresponding to the tolerance parametric control with lower and upper tolerances, piecewise linear functions are used. Z-shaped or S-shaped membership functions are used to define fuzzy sets characterizing uncertainty of the "lower tolerance" or "upper tolerance" type, respectively. It is possible to build bell-shaped functions. With a lack of information, it is recommended to

use piecewise linear membership functions. In the course of operation, the type and parameters of the membership functions, the number of terms of linguistic variables and their names can be adjusted upon receipt of clarifying information and new empirical data.

To describe subsets of a term set (Table 1), a system of membership functions is used, the basis of which is constituted by triangular membership functions. The choice of the type of triangular membership function is justified by the following considerations:

- when estimating the parameters of the membership functions, only interval constraints and the most acceptable values of the parameters are known. If the researcher does not have more information, then the only acceptable approximation is linear. The triangular membership function is defined by the minimum number of parameters: minimum value, modal value, and maximum value;
- triangular membership functions have a low computational complexity, which allows them to be used in situations with a limited time for making management decisions. In addition, they are widely used in existing fuzzy logic applications; their reliability and efficiency have been tested in practice.

The numerical value of the input variable x_1 is determined based on the time series of the parameter values and the rank correlation coefficient of C . Spearman R_S using the following formulas:

$$x_1 = |1 - R_S|, \tag{2}$$

$$R_S = 1 - \frac{6 \sum_{h=1}^n d_h^2}{n^3 - n}, \tag{3}$$

$$R_S = 1 - \frac{6 \sum_{h=1}^n d_h^2 - A}{n^3 - n - 12A}, \tag{4}$$

$$A = \frac{\sum_{t=1}^m (A_t^3 - A_t)}{12}, \tag{5}$$

where d —the difference of the ranks of the parameters and the ranks of the moments of time in the series, n is the number of moments of time in the series, A —amendment to the Spearman formula in the presence of fractional ranks, t —number of groups of the same rank in order, A_t —the number of identical ranks in the t th group. The numerical value x_1 from the interval $[0,1]$ characterizes the steadiness of the directional change of the parameter values and the steadiness of the parameter values. A violation of a strictly ranked sequence of parameter values indicates the incomplete stability of their directional change, and a deviation from the tendency in one direction or the other direction indicates an incomplete stability of parameter values.

In the fuzzy classifier of the parameter x_2 , the interval $[0, 1]$ on the relative scale corresponds to the width of the tolerance field. The initial data for determining the

numerical value of x_2 are x_i is the current value of the parameter, x_{\min} and x_{\max} are the parameter values that specify the lower and upper limits of the tolerance field, and R_S . It should be noted that with $R_S > 0$, the main tendency of the dynamics is the growth of parameter values (upward trend), with $R_S < 0$ —a decrease, and with $R_S = 0$ —there is no tendency. The parameter x_2 , which characterizes the working capacity, allows us to estimate the distance of the current value from the boundary of the tolerance field, taking into account the tendency of dynamics. The numerical value x_2 is determined using a system of three production rules:

Rule 1: If “ $R_S = 0$ ”, then “Fulfill Rule 2”; otherwise, “Fulfill Rule 3”;

Rule 2: If “ $x_{\max} - x_i > x_i - x_{\min}$ ”, then “ $x_2 = x_{\max} - x_i$ ”, otherwise “ $x_2 = x_i - x_{\min}$ ”;

Rule 3: If “ $R_S > 0$ ”, then “ $x_2 = x_{\max} - x_i$ ”, otherwise “ $x_2 = x_i - x_{\min}$ ”.

The value of the input parameter x_3 expresses the speed of increasing or decreasing the current value of the parameter for the time interval between the current and the previous measurement. It is determined by the formula:

$$x_3 = \frac{x_i - x_{i-1}}{t_i - t_{i-1}}. \quad (6)$$

The value of y is a numerical estimate of the state of the “fit” parameter by the complex criterion for working capacity level, obtained on the basis of a fuzzy classifier and a matrix method of data aggregation [10].

As weights for the aggregation of data at the level of their qualitative states, a set of nodal points was introduced:

$\alpha_1 = \alpha_{EC} = 0$, $\alpha_2 = \alpha_{GC} = 0.3$, $\alpha_3 = \alpha_{SC} = 0.5$, $\alpha_4 = \alpha_{DC} = 0.7$, $\alpha_5 = \alpha_{PS} = 1$. The nodal points are symmetrical with respect to node 0.5 and are the abscissas of the maxima of the corresponding membership functions, which makes it possible to unambiguously recognize the values of the input and output variables with one hundred percent expert confidence.

Thus, the set of linguistic variable A_j , defined on a single segment of the real axis $[0, 1]$, the term set of its values and a set of node points is a five-level fuzzy classifier of the parameter x_j . The term set of values is described by triangular, z -shaped, and s -shaped membership functions. Graphically, the set of membership functions of a term set of values of the linguistic variable A_j is presented in Fig. 1, B in Fig. 2.

It should be noted that the sum of all values of the membership functions for any parameter x_j and y is equal to one, which indicates the consistency of the classifier.

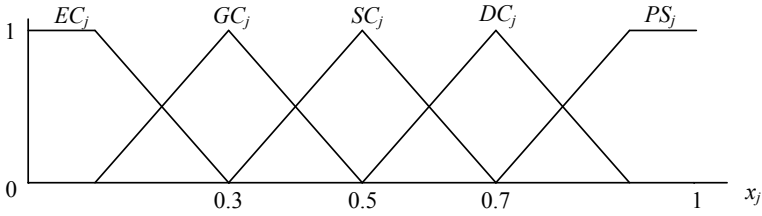


Fig. 1 Five-level fuzzy parameter classifier x_j

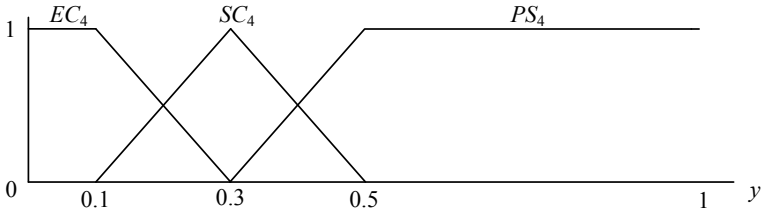


Fig. 2 Three-level fuzzy parameter classifier y

3 Criteria for Assessing the Technical Condition

Recognition of values x_1 , x_2 , and x_3 produced by the criteria of classification tables. Table 2 is a classification of parameter values x_j .

Table 2 Classification of parameter x_j

The interval of values x_j	Classification of the parameter value	The degree of evaluation confidence
$0 \leq x_j \leq 0.1$	EC_j	$\mu_{EC_j}(x_j) = 1$
$0.1 < x_j < 0.3$	EC_j	$\mu_{EC_j}(x_j) = \frac{0.3-x_j}{0.2}$
	GC_j	$\mu_{GC_j}(x_j) = 1 - \mu_{EC_j}(x_j)$
$x_j = 0.3$	GC_j	$\mu_{GC_j}(x_j) = 1$
$0.3 < x_j < 0.5$	GC_j	$\mu_{GC_j}(x_j) = \frac{0.5-x_j}{0.2}$
	SC_j	$\mu_{SC_j}(x_j) = 1 - \mu_{GC_j}(x_j)$
$x_j = 0.5$	SC_j	$\mu_{SC_j}(x_j) = 1$
$0.5 < x_j < 0.7$	SC_j	$\mu_{SC_j}(x_j) = \frac{0.7-x_j}{0.2}$
	DC_j	$\mu_{DC_j}(x_j) = 1 - \mu_{SC_j}(x_j)$
$x_j = 0.7$	DC_j	$\mu_{DC_j}(x_j) = 1$
$0.7 < x_j < 0.9$	DC_j	$\mu_{DC_j}(x_j) = \frac{0.9-x_j}{0.2}$
	PS_j	$\mu_{PS_j}(x_j) = 1 - \mu_{DC_j}(x_j)$
$0.9 \leq x_j \leq 1.0$	PS_j	$\mu_{PS_j}(x_j) = 1$

All the necessary data for the calculation of the estimated state of the parameter by the complex criterion are collected in a matrix (Table 3).

To obtain a numerical estimate of the state of the “fit” parameter using the complex criterion of working capacity level, the double convolution formula is used:

$$y = \sum_{j=1}^3 p_j \sum_{i=1}^5 \alpha_i \mu_{ji}(x_j), \quad (7)$$

where p_j are the weights of the input parameters x_j , α_i are the node points of the classifiers, $\mu_{ji}(x_j)$ are the values of the membership functions of fuzzy sets corresponding to the node points relative to the current values of the input parameters x_j . The values of weights for different stages of control may vary. The values can be determined by the Fishburn rule [11], and with sufficient justification, they are taken equal and then $p_j = 1/3$.

Recognition of the obtained value of assessing the state of the “fit” parameter according to the complex criterion of working capacity level is made according to the classification Table 4.

The result of the classification is a linguistic description of the state of the parameter and the degree of confidence of the expert in this recognition result.

4 Conclusion

The results of the evaluation of all the considered parameters are the initial data for subsequent analysis and decision making. It should be noted that the proposed approach can be used in monitoring and diagnostic systems for solving the following applied and research tasks:

- increase the reliability of control;
- increase of control productivity by reducing the time for analyzing the results of test checks;
- reducing the cost of labor for control;
- detection of pre-failure states of technical systems at the earliest stages of defect development;
- evaluation of the effectiveness of various control actions and prediction of their consequences;
- prediction of the technical condition of the product, its component parts, defects, and pre-failure states in various operating conditions;
- determination of the degree of perfection of the design, production technology, the correctness of the choice of the denominations of the elements, circuit solutions, and operating modes;
- making more informed decisions on the suitability of the product to perform its functions, on proactively setting up, adjusting the working capacity of its components, and debugging their interaction;

Table 3 Matrix for estimating the state of a parameter by a complex criterion
 Membership functions of fuzzy sets of terms A_1, A_2 and A_3

Parameters	Weightcoefficient	Membership functions of fuzzy sets of terms A_1, A_2 and A_3					
		Excellent condition	Good condition	Satisfactory condition	Dangerous condition	Pre-failure status	
x_1	p_1	$\mu_{EC_1}(x_1)$	$\mu_{GC_1}(x_1)$	$\mu_{SC_1}(x_1)$	$\mu_{DC_1}(x_1)$	$\mu_{PS_1}(x_1)$	
x_2	p_2	$\mu_{EC_2}(x_2)$	$\mu_{GC_2}(x_2)$	$\mu_{SC_2}(x_2)$	$\mu_{DC_2}(x_2)$	$\mu_{PS_2}(x_2)$	
x_3	p_3	$\mu_{EC_3}(x_3)$	$\mu_{GC_3}(x_3)$	$\mu_{SC_3}(x_3)$	$\mu_{DC_3}(x_3)$	$\mu_{PS_3}(x_3)$	
Nodal points		0	0.3	0.5	0.7	1	

Table 4 Classification of parameter y

The interval of values y	Classification of the parameter value	The degree of evaluation confidence
$0 \leq y \leq 0.1$	EC ₄	$\mu_{EC_4}(y) = 1$
$0.1 < y < 0.3$	EC ₄	$\mu_{EC_4}(y) = \frac{0.3-y}{0.2}$
	SC ₄	$\mu_{SC_4}(y) = 1 - \mu_{EC_4}(y)$
$y = 0.3$	SC ₄	$\mu_{SC_4}(y) = 1$
$0.3 < y < 0.5$	SC ₄	$\mu_{SC_4}(y) = \frac{0.5-y}{0.2}$
	PS ₄	$\mu_{PS_4}(y) = 1 - \mu_{SC_4}(y)$
$0.5 \leq y \leq 1.0$	PS ₄	$\mu_{PS_4}(y) = 1$

- assessment of the wear of complex systems with a significant service life, to extend their service life with the possible replacement of individual units and assemblies;
- transition from the system of preventive maintenance and servicing to the strategy of managing the operational reliability of the system according to its actual technical condition;
- a retrospective analysis of the causes of failures of the technical system;
- optimization of the program, scheme, and scope of control, taking into account the specifics of each specific system, its components in the specific conditions of production and operation;
- monitoring the development of dangerous tendency in the parameters to identify potentially defective components of the system;
- optimization of the nomenclature and the number of spare parts, assemblies, and materials;
- addition and training of the knowledge base in *IDSS*.

The proposed approach provides more correct system parameters estimating for tolerance control processes. It is characterized by the presence of an additional original procedure using four fuzzy classifiers that take into account the proximity of parameter values to the limits of tolerance fields and the dynamics of their change. This approach allows to increase the reliability of control results under conditions of uncertainty and risk.

The estimates obtained make it possible to make informed decisions aimed at preventing potential failures in order to prevent abnormal and emergency situations in operational conditions under complex technical system, to take timely remedial and preventive measures aimed at increasing its working capacity level and ensuring sustainable operation, i.e., increasing its actual resource.

References

1. G.I. Korshunov, S.A. Nazarevich, V.A. Smirnov, Fuzzy classification of technical condition at life cycle stages of responsible appointment systems, in *Proceedings of the II International Scientific and Practical Conference "Fuzzy Technologies in the Industry—FTI 2018"*, Ulyanovsk, Russia, 23–25 Oct 2018, vol. 2258. CEUR Workshop Proceedings, pp. 427–437
2. V.A. Smirnov, Malfunction searching in onboard control systems during acceptance control. *Informatsionno-upravlyayushchie sistemy [Inf. Manage. Syst.]* **2**, 24–28 (2013) (in Russian)
3. L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1**, 3–28 (1978)
4. S. Nahmias, Fuzzy variables. *Fuzzy Sets Syst.* **1**, 97–110 (1978)
5. K.M. Passino, S. Yurkovich, *Fuzzy Control* (Addison Wesley Longman, Boston, USA, 1998), p. 522
6. M. Friedman, M. Ming, A. Kandel, Fuzzy linear systems. *Fuzzy Sets Syst.* **96**, 201–209 (1998)
7. R. Brachman, P. Sefridge, Knowledge representation support for data archeology. *Intell. Cooper. Inform. Syst.* **2**, 113–120 (1993)
8. R.E. Bellman, L.A. Zadeh, Decision making in a fuzzy environment. *Manage. Sci.* **17**, 141–164 (1970)
9. A.V. Leonenkov, *Nechetkoe modelirovanie v srede MATLAB i fuzzy TECH [Fuzzy Modeling in MATLAB and fuzzyTECH]* (BKHV-Petersburg, St. Petersburg, 2005) 736p (in Russian)
10. A.O. Nedosekin, *Nechetko-mnozhestvennyj analiz riska fondovyh investicij [Fuzzy Multiple Risk Analysis of Stock Investment]* (Printing House "Sesame", St. Petersburg, 2002) 181p
11. P.C. Fishburn, *Utility Theory for Decision Making* (Wiley, New York, 1970), p. 234

Software Tools and Techniques for the Expert Systems Building



Arslan I. Enikeev, Rustam A. Burnashev and Galim Z. Vakhitov

Abstract The report presents the results of research on the creation of CASE tools that provide the ability to effectively build expert systems. As a part of the creation of CASE tools, we focus on building an integrated development environment that includes a combination of SWI-Prolog, Java, Python programming languages, PostgreSQL database management system (DBMS) as well as telemetry tools. On the basis of the created integrated development environment, an experimental version of the expert system was built. This expert system is mainly focused on automating the analysis processes and forming requirements for the software applications and hardware being developed using the built-in telemetry tools and taking into account the specifics of the corresponding subject area. The expert system is performed using the logical rules concerning the characteristics of workstations and corresponding software systems. As a result, the expert system forms requirements and recommendations to the properties of the software and hardware products being developed.

Keywords Expert system · Database management systems · CASE tools · Continuous integration · Telemetry

1 Introduction

Currently, when developing software applications, the process of communication between the customer, the developer, and user is often limited only by the functionality of the application being developed. The practice of a software development shows that this is not enough because, first of all, it is necessary to clarify why the program is needed and who will use it. The developer of a software product, in order to make quick money, is often taken to design an application without even getting an answer to a question from a customer—why? As a result, sometime after the exploitation of the embedded application, the developer is informed that the application will be used for example not only by the personnel department but also by

A. I. Enikeev (✉) · R. A. Burnashev · G. Z. Vakhitov
Kazan Federal University, Kazan, Tatarstan, Russia
e-mail: r.burnashev@inbox.ru

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_16

the accounting department and the director. Even worse, this application needs for its work the different operating systems. Thus, the developer spends a lot of effort, time, and finance to rebuild this application.

For the correct and effective implementation of a software application, the developer has to precisely specify the properties of the application with the customer and only then begin to collect and analyze the requirements.

The analysis of requirements for a software product is the foundation of a future software product, which requires a lot of time and knowledge in the development of software products for subsequent integration into the subject area processes. In order to formulate and analyze software product requirements, one can use the specialized expert systems based on the dynamically updated knowledge base and the respective telemetry tools. Depending on the operating environment, the hardware used, and the subject area, it can be various specialized expert systems. Therefore, for the effective creation of such expert systems, it is reasonable to build a unified integrated development environment, appearing to be one of the substantial components of CASE technologies [1]. This report proposes an approach to building an integrated development environment for specialized expert systems based on a combination of SWI-Prolog, Java, Python programming languages, PostgreSQL database management systems (DBMS), as well as telemetry tools, which make building tools universal in solving analysis problems design, testing, and commissioning of the finished product. Using the created integrated development environment, we have built an experimental expert system [2, 3] focusing on automating the analysis processes and forming requirements for the software applications and hardware being developed based on the built-in telemetry tools and taking into account the specifics of the corresponding subject area.

2 Methods

One of the main purposes of the ES is to solve tasks that are rather time-consuming for experts based on the accumulated knowledge base filled with data that reflect the professional experience and qualifications of experts in the relevant subject area [4].

An expert system that was built using an integrated development environment allows you to automatically generate and analyze the requirements for the applications being developed [2] in the relevant subject area based on the knowledge base and telemetry tools. In Fig. 1, the scheme of interaction between “developers” and “customer” is represented.

3 Telemetry

The telemetry tools are embedded in the integrated development environment to determine the necessary characteristics of software and hardware of workstations. The relevant information about these characteristics is used subsequently to fill the

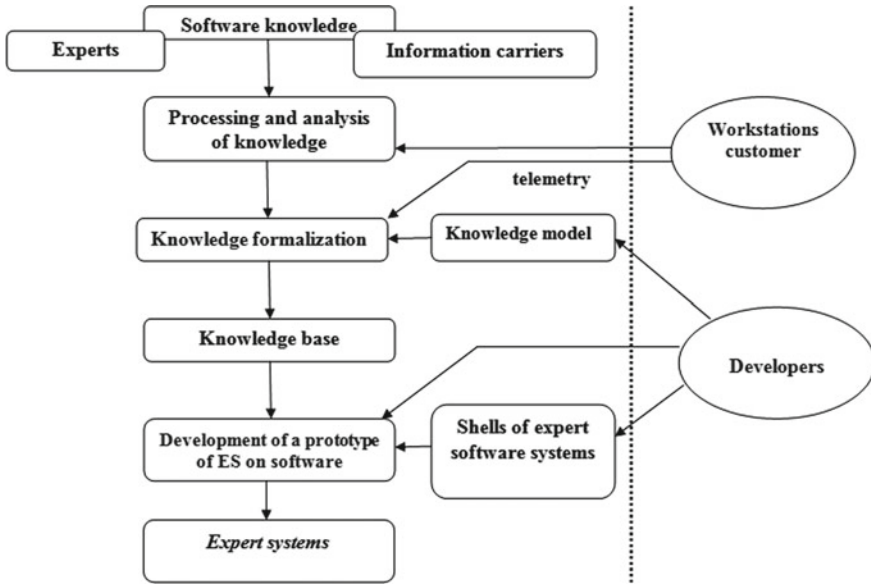


Fig. 1 The scheme of interaction between “developers” and “customer”

knowledge base of the developed expert system with new facts. Using the obtained characteristics on workstations in the knowledge base, the expert system produces reference information which will be later used for creating technical specifications to ensure successful implementation of the software product. Testing on the level of the specification is a very important process as it is cheaper than making a change in the final product. The structure of the system involves the use of client–server component, responsible for sending, collecting, and subsequent processing of data in the knowledge base.

Software developers are offered a wide range of selection of the necessary characteristics for workstations in the network including:

- (1) Operating system (version information, discharge, latest update)
- (2) ROM (memory information)
- (3) GPU (frequency, memory)
- (4) CPU (core frequency, cache)
- (5) IP addresses and computer names of the network, etc.

Intelligent data analysis is the process of detecting previously unknown, non-trivial, practically useful, and accessible interpretation of knowledge necessary for decision-making [5] in various spheres of human activity. Applying an intelligent data analysis to software design process, we form the necessary characteristics of the workstations. For an implementation of that, we used SWI-Prolog, Python, and the necessary modules included in the integrated development environment:

//pandas module and DataFrame object for analyzing characteristics

```

ra = pd.DataFrame ({
    'PC name': [socket.gethostname()],
    'IP': [socket.gethostbyname(hostname)],
    'OS': [platform.system()],
    'OS version' : [platform.release()],
    'CPU capacity': [sys.platform],
    'CPU' : [proc_info.Name],
    'RAM' : [system_ram],
    'Date:': [time.ctime()],
}, index = ['Work station'])

```

In data mining, there are the following stages:

- Identification of patterns (free search);
- Predictive modeling;
- Exception analysis.

At the stage of identifying patterns, a dataset is examined to find hidden patterns. Free search is represented by such actions: the identification of patterns of conditional logic, the identification of patterns of associative logic, the identification of trends and fluctuations.

In the course of solving the problem of searching for associative rules, regularities are found between related events in a dataset. The distinction of association from the tasks of classification and clustering is that the search for patterns is carried out not on the basis of the properties of the analyzed object, but between several events which occur simultaneously.

4 Results and Discussions

Using the data in the knowledge base, the expert system generates requirements and recommendations for system and hardware characteristics for the qualitative integration of a software product into workflows of various subject areas that will permit to satisfy basic needs and expectations of software products customers.

The system of measuring the characteristics of workstations is a set of measured characteristics, units of measurement, measurement scales, and connections established between the elements [6].

The «pandas» module in Python provides the possibilities for a convenient combination of several workstations characteristics using Series, DataFrame, and Panel objects with various types of dataset logic and relational algebra. Below, the program code in Python for structuring the characteristics of workstations for further data analysis is represented:

Example of the workstation:

```
>>ra = pd.DataFrame ({
    'PC name': [socket.gethostname()],
    'IP': [socket.gethostbyname(hostname)],
    'OS': [platform.system()],
    'OS version' : [platform.release()],
    'CPU capacity': [sys.platform],
    'CPU' : [proc_info.Name],
    'RAM' : [system_ram],
    'Date:': [time.ctime()],
    }, index = ['First work station'])
```

Combining workstations using the «concat»() function:

```
>> result = pd.concat(frames)
```

The result of measurements of characteristics (Fig. 3) in the selected scale was the identification of similarities and differences in the characteristics of hardware and software of workstations using the Apriori algorithm, searching for association rules in data mining in Python.

Education on association rules (hereinafter, associations rules learning—ARL) is quite often applicable in real-life method of finding relationships (associations) in a dataset. For the first time, Piatetsky-Shapiro G [20] wrote about this in detail in “Discovery, Analysis, and Presentation of Strong Rules” (1991).

There are a number of basic concepts in ARL:

- Support
- Confidence
- Lift (support)

Support

To check when a transaction is a single element set result, (1) is used:

$$\text{supp}(X) = \frac{|\{t \in T, X \in t\}|}{|T|} \quad (1)$$

where

X is a data element to determine the similarity in a set of sets in all transactions.

T is the number of transactions.

To check several items (Fig. 2) with transaction (2):

$$\text{supp}(x_1 \cup x_2) = \frac{\sigma(x_1 \cup x_2)}{|T|} \quad (2)$$

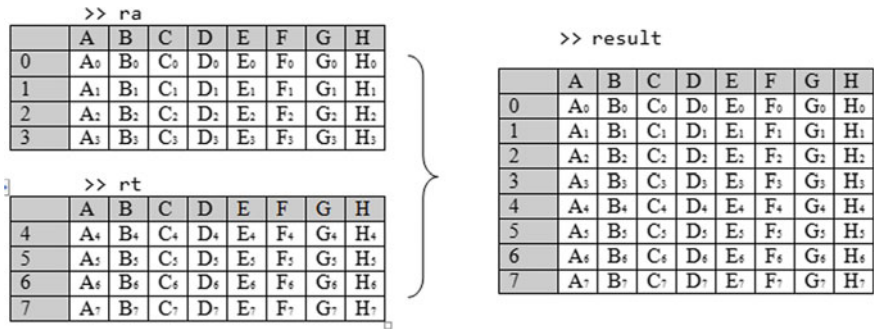


Fig. 2 Using the «concat» () function to combine workstations

where

σ is the number of transactions containing x_1 and x_2

T is the number of transactions

In the example of collecting and analyzing performance requirements:

$$\text{supp} = \frac{\text{Windows transactions and resolution 1280:768}}{\text{all transactions}} = P(A \cup B) \quad (3)$$

The result is presented in the form of a diagram created with the matplotlib module (Fig. 3) in Python:

$$\begin{aligned} \text{supp} &= 80\% \text{ (Windows OS)} \\ \text{supp} &= 100\% \text{ (Screen resolution)} \end{aligned} \quad (4)$$

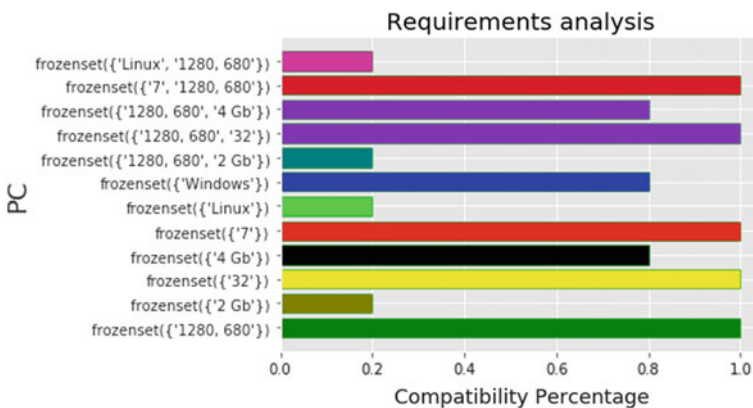


Fig. 3 The result of the analysis of the characteristics of workstations

Confidence

This indicator is designed to test the entire set of result elements, (5) is applied:

$$\text{conf}(x_1 \cup x_2) = \frac{\text{supp}(x_1 \cup x_2)}{\text{supp}(x_1)} \tag{5}$$

where

$\text{conf}(x_1 \cup x_2)$ —the rule is checked, namely in which transactions the rule “A” $\text{supp}(X)$ is executed, in the elements A to H “results”:

$$\text{supp}(X \cup Y) \tag{6}$$

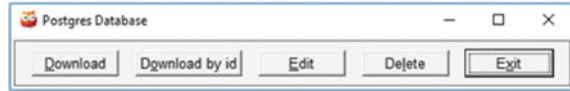
The programming code for implementing the Apriori algorithm in the Jupyter Notebook integrated development environment (Fig. 4):

```
import pandas as pd
from apyori import apriori
plt.style.use('ggplot')
plt.yticks(y_pos, labels)
plt.ylabel('PC', fontsize=18)
plt.xlabel('Compatibility', fontsize=18)
plt.title('Requirements', fontsize=20)
plt.show()
support=df.iloc[0:12]['support']
products=df.iloc[0:12]['items']
simple_bar_chart(support,products)
```

Fig. 4 System for measuring characteristics

	items	support
0	(1280, 680)	1.0
1	(2 Gb)	0.2
2	(32)	1.0
3	(4 Gb)	0.8
4	(7)	1.0
5	(Linux)	0.2
6	(Windows)	0.8
7	(2 Gb, 1280, 680)	0.2
8	(32, 1280, 680)	1.0
9	(4 Gb, 1280, 680)	0.8
10	(7, 1280, 680)	1.0
11	(Linux, 1280, 680)	0.2

Fig. 5 The main menu for working with database



Below, a listing of the `soft.pl` knowledge base implemented in the SWI-Prolog (logic programming language), in which there are facts and rules that ensure the identification of facts of compatibility of database management systems with a software product, is represented.

An example of the implementation of the query language in SWI-Prolog (compatibility with the standard SQL) in the production model of knowledge representation (Fig. 5):

```
s_compatible('PostgreSQL').
s_compatible('MySQL').
s_compatible('Oracle Database').
# software compatibility
soft_compatible('PostgreSQL').
soft_compatible('MySQL').
soft_compatible('Oracle Database').
# Rules in the knowledge base for determining the compatibility of software products
s_compatible(X) :- soft_compatible(X)

getTables:-
odbc_current_table(PostgreSQLProlog,Table),
write(Table).
getColumns(Table):-
odbc_table_column(PostgreSQLProlog,Table,Column),
write(Table-Column).
```

5 Conclusion

As a result of the study, we developed an experimental version of a specialized expert system based on logical rules for the creation of the workstation characteristics. These characteristics provide an efficient process of software and hardware product development. The expert system was built using the integrated development environment of expert systems. In the nearest future, using the implemented integrated development environment, we intend to create an expert system for medical applications.

Acknowledgements This work was supported by the research grant of Kazan Federal University.

References

1. H.A. Muller, R.J. Norman, J. Slonim, *Computer Aided Software Engineering* (Springer, New York, 1996)
2. R.A. Burnashev, A.V. Gubajdullin, A.I. Enikeev, Specialized case tools for the development of expert systems. *Adv. Intell. Syst. Comput.* **745**, 599–605 (2018). https://doi.org/10.1007/978-3-319-77703-0_59
3. R.A. Burnashev, N.S. Yalkaev, A.I. Enikeev, Data structuring and data processing for the information intellectual applications. *J. Fundamental Appl. Sci.* **9**, 1403–1416
4. A.M. Kamalov, R.A. Burnashev, Development of the expert system prototype «medexpert» for differential disease diagnostics. *Astra Salvensis* **2017**, 55–64 (2017)
5. F. Telnov Yu, *Intelligent Information Systems* (Moscow International Institute of Econometrics, Informatics, Finance and Law, Moscow, 2004), C. 11
6. K. Burov, Detection of knowledge in data warehouses. *Open Syst.* № 5–6, 67–77 (1999)

A Component-Based Method for Developing Cross-Platform User Interfaces for Mobile Applications



Marek Beranek  and Vladimir Kovar

Abstract A common problem with User Interface (UI) frameworks is their closed architecture forcing designers to implement user interfaces from a limited set of components. This lack of extensibility limits the usability features of the application and may result in an unappealing user interface. Additionally, poor reusability associated with UI development leads to low application development productivity and high project costs. In this paper, we describe the Unicorn Universe User Interface framework uu5. uu5 is a component-based framework designed to support rapid development of reliable and scalable cross-platform mobile applications. The uu5 framework simplifies UI development using specialized components that improve user experience and facilitate integration with React and other commonly used UI libraries. We illustrate the uu5 design method with an example of uuCourseKit education delivery application.

Keywords Component-based method · Cross-platform user interfaces · Mobile applications

1 Introduction

Development of user interfaces for mobile applications is a challenging, labor-intensive task that involves different types of skills, ranging from creative high-level design to highly technical developer skills. To achieve the best user experience results, user interfaces are often *handcrafted* to suit a particular business use case, reducing reusability across different projects, and resulting in low application development productivity. To improve application development productivity, User Interface (UI) must be built from reusable components, extending existing components and developing new components only when no existing components have the required

M. Beranek (✉) · V. Kovar
Unicorn College, V Kapslovně 2767/2, 130 00 Prague 3, Czech Republic
e-mail: marek.beranek@unicorncollege.cz

V. Kovar
e-mail: vladimir.kovar@unicorn.eu

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_17

functionality. Another important benefit of UI component reuse is a consistent user experience across the entire application portfolio. UI development frameworks play an important role in component-based development of user interfaces; however, most frameworks have a closed architecture forcing developers to implement user interfaces from a limited set of components. This lack of extensibility limits the usability features of the application and may result in an unappealing UI on some devices (e.g., mobile phones).

In this paper, we describe the Unicorn Universe User Interface framework *uu5*, designed to support rapid development of reliable and scalable cross-platform mobile applications. The *uu5* framework simplifies UI development by using specialized components that improve user experience and facilitate integration with React (<https://facebook.github.io/react/>) and other commonly used UI libraries. *uu5* implements a comprehensive and extensible graphical system and typography based on the Material Design standard developed by Google.

In the next section (Sect. 2), we review related literature on UI development methods and frameworks. In the following sections, we discuss component-based UI development using the *uu5* component library (Sect. 3) and the *uu5* component design process (Sect. 4). Section 5 illustrates the *uu5* method using the *uuCourseKit* education delivery application. Section 6 contains our conclusions.

2 Related Literature

Methods and tools for developing cross-platform user interfaces for mobile applications have been the subject of recent research interest both in academia and by industry practitioners [1–4]. Melamed et al. have investigated the use of pervasive computing to deliver applications and services on mobile phones, evaluating a number of platforms including J2ME and native smartphone development [5]. The paper introduces HTML5, describes its advantages and drawbacks, and concludes that HTML5 is a good solution for creating and distributing pervasive media applications. Another study compares native code and Web code development for mobile applications [1]. Authors conclude that hybrid solutions will play an important role in the future. In another recent publication, the authors report on a comparative analysis of cross-platform development approaches for mobile applications and conclude that HTML5 will play a key role in the future [2]. Experiences gained during an industrial project concerned with building user interfaces for database access are discussed in [6]. The authors describe a systematic approach to analysis of standards and conventions for the design of user interfaces for various mobile platforms, focusing on interoperability of different systems such as HTML5, Java, and .NET. Kruchten et al. describe the 4 + 1 Architectural View Model for the Rational Unified Process [7]. View scenarios consist of a set of use cases that describe sequences of interactions between objects and processes. UML 2.5 specification describes use cases that capture system behavior with users and IoT devices represented as *actors* that interact with the system through use cases [8]. Brambilla et al. [9] propose an extension

to the OMG Interaction Flow Modeling Language (IFML) for mobile applications and describe their implementation experience with the development of automatic code generators for cross-platform mobile applications based on HTML5, CSS, and JavaScript. The authors illustrate their approach implementing a popular mobile application and provide a productivity comparison with traditional approaches. In another publication, authors investigate the application of model-driven development of user interfaces for Internet of Things (IoT) systems [9]. Acerbis et al. [10] propose a comprehensive tool suite (the WebRatio Platform) for model-driven development of Web and mobile applications. The tool supports developers in the specification of domain and user interaction models using extended versions of the OMG IFML language. The extensions include primitives tailored for Web and mobile application development.

In summary, it is evident that there is a trend toward the use of HTML5-based technologies for the implementation of cross-platform mobile applications. However, at present, there is no widely accepted standard framework forcing developers to use a combination of various technologies, programming languages, and libraries for the development of mobile applications. Furthermore, most of the above approaches do not provide effective support for reuse of UI components.

3 Component-Based UI Development Using the uu5 Framework

In traditional (*fat client*) client/server applications, user interfaces are implemented using components with properties and methods that are activated by application events. This approach facilitates high levels of component reuse. HTML-based user interfaces are typically constructed from a set of HTML elements that are not associated with properties and methods, making reuse of UI components difficult. For example, if the requirement is for a data table that displays a set of data elements and enables sorting and filtering, identical HTML elements and JavaScript code have to be implemented repeatedly in different applications and across different projects. With the emergence of HTML5, customizable and reusable UI components have become feasible, significantly improving the potential for reuse. HTML5 applications typically run as a single-page application (SPA) [11], and when combined with suitable frontend technologies (e.g., React) have several important advantages [12]:

- The amount of transferred data from the server is minimized as the user interface is loaded only once.
- The page does not flicker as it does not need to be fully re-rendered.
- As the page does not need to be re-loaded, the JavaScript contained within the page is compiled only once when the page is first loaded, substantially improving the response time.
- Advanced features and capabilities of HTML5 and JavaScript result in a user experience that is comparable with a native UI.

- Resources (i.e., processor and memory) of the client device are used for rendering and for executing user interface logic, avoiding having to generate the HTML5 code on the server. The server executes Application Programming Interface (API) calls and needs lower resources, making the application less expensive to run.

3.1 *uu5 Library*

Unicorn uu5 is a platform and a library for building user interfaces based on HTML5 and JavaScript [13]. The uu5 library integrates React framework [14] and Material Design graphical system [15]. A number of typography and responsivity principles are adopted from the Bootstrap framework [16]. The uu5 library facilitates development of responsive, *mobile-first* applications with UI that adapts to a specific device, ensuring that the user experience is comparable with native applications (i.e., applications designed to run on a specific platform, e.g., iOS or Android). The uu5 library can be used for any device that supports a Web browser, including smartphones, tablets, desktop and laptop computers, and smart TVs. uu5 is ideal for developing single-page applications and produces applications that can be easily controlled via a keyboard, mouse, or touch. The uu5 library is an open-source library, and its license is derived from the standard BSD 2.0 license.

3.2 *uu5 Components*

A uu5 component is an application element (that can be composed of a combination of HTML5 elements) placed on a page or within an application. Composition of components allows developers to create complex components that can be encapsulated within a visual use case to display text or to perform input validation. Component behavior is controlled via settings and depends on the context that the component is deployed in. More complex components can display data tables, provide interactive charts, or read and process QR (Quick Response) codes. uu5 is based on React and manages components along the React lifecycle [17]. uu5 extends React rules with its own rules, designed to increase efficiency of the UI development process. Every component is stored in a separate JavaScript file and has its own style sheets file, so that, for example, the BigCalendar component is implemented using the big-calendar.js and big-calendar.less files. uu5 recommends using the backward-compatible language extension *less* (Leaner Style Sheets) for style sheets. Table 1 lists the main elements of uu5 components.

The uu5 library defines the following types of components [18]:

- User Interface Entry point (UVE) component represented by an HTML file with linked libraries.
- Route-specific content of a page.
- Other types of application components.

Table 1 uu5 component elements

Mixins	<i>Mixins</i> extend a component, for example with the functionality to react to change of its size. Mixins support multi-level inheritance
Statics	The <i>Statics</i> object contains constants that are initialized when the library is first loaded into a page
PropsTypes	Data types of component properties are defined and validated against <i>PropsTypes</i>
DefaultProps	<i>DefaultProps</i> are used to define the default properties for all <i>propsTypes</i>
Component lifecycle	<i>Component lifecycle</i> describes the implementation of the methods required by React component lifecycle, as described in Sect. 3.3
Interface	The component <i>interface</i> includes public methods that can modify component properties or behavior without the need to re-render the parent component
Overriding methods	<i>Overriding methods</i> are used to override component default behavior and are typically implemented through a <i>Mixin</i>
Private methods	<i>Private methods</i> can be called only within the component code
Render	Every React component must have the <i>render</i> method

3.3 uu5 Component Lifecycle

Similar to React, every uu5 component has its own lifecycle supported by a set of methods which are called during mounting, running, and unmounting of the component. Figure 1 illustrates the sequence of methods calls and the component lifecycle. When a component is created, the following methods are called: *componentWillMount*, *render*, and *componentDidMount*. If component properties change, the *componentWillReceiveProps* method is called and then the methods: *shouldComponentUpdate*, *componentWillUpdate*, *render*, and *componentDidUpdate* are called. If a parent component is unmounting, the *componentWillUnmount* method is called.

3.4 Single-Page Applications

A single-page application (SPA) is a technique for organizing HTML5 pages and communicating with the server. An application using this technique creates user interface by dynamically changing the current HTML5 page. This avoids loading a new page after each click event, which is commonly the case with traditional Web applications. All the necessary HTML5, CSS, and JavaScript elements are loaded either during the initialization of the application or on-demand in the background. This technique is also suitable for developing cross-platform mobile applications.

Because a SPA loads only a single page, there are no changes in the Web browser address bar when interacting with the user, and thus no need to save the browsing history. This affects the behavior of the *Previous* and *Next* browser buttons. Activating

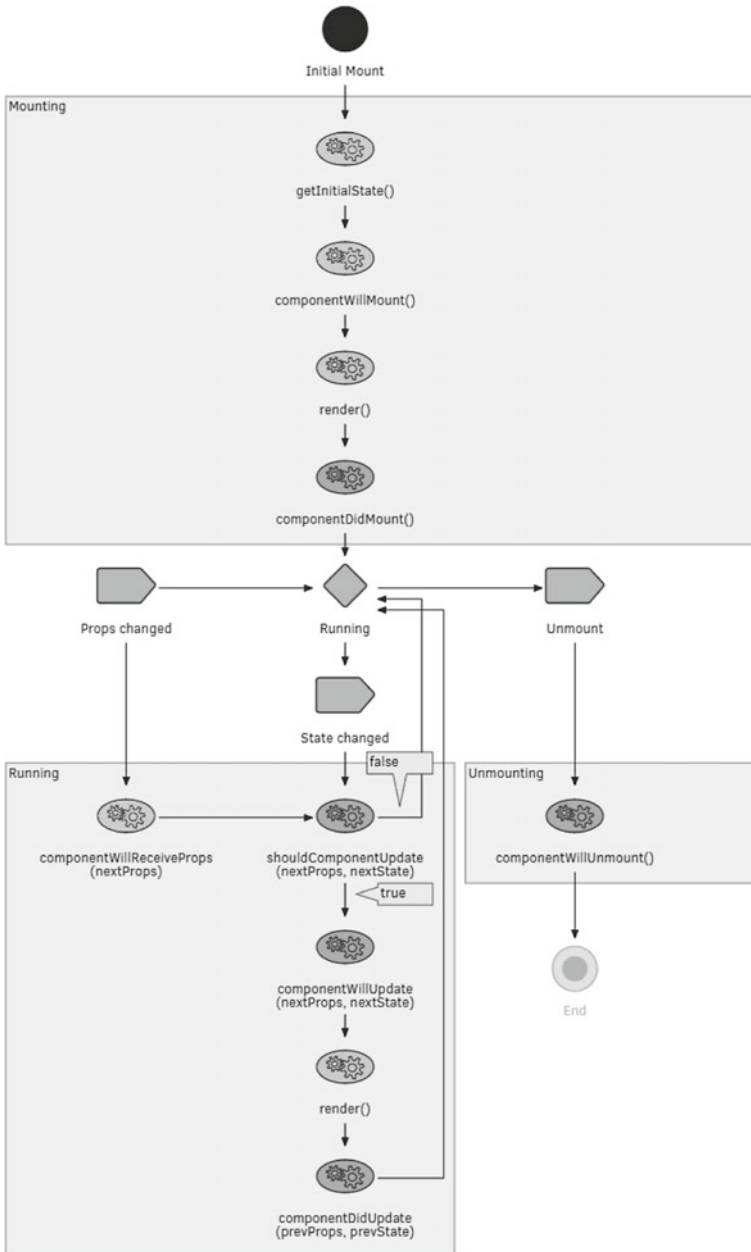


Fig. 1 React component lifecycle [14]

the Previous button causes the browser to display the page that preceded the SPA, not the previously displayed page. uu5 deals with these problems by using the router component (<UU5.Common.Router>).

A typical SPA layout consists of several parts; some are static (e.g., application header, main menu navigation, and application footer), and some are continuously changing according to the corresponding visual use case. uu5 deals with this layout requirement by using the page component (<UU5.Bricks.Page>).

3.5 Responsivity

The main advantage of the uu5 *mobile-first* approach is that Responsive Web Applications can be adapted to different devices including mobile phones, tablets, and desktop computers. As these devices have different screen size, the framework needs to define the basic range of sizes that the application can render. For example, Bootstrap defines five sizes: extra small devices (portrait phones), small devices (landscape phones), medium devices (tablets), large devices (desktops), and extra-large devices (large desktops) [19]. Bootstrap uses a grid technique that supports responsivity by dividing the screen into 12 notional columns and rearranging the columns according to screen size ensuring that the UI is best suited to the device it runs on. So that, for example, a component on a mobile phone extends over the entire screen, but on a desktop computer, it spans only three columns, allowing the display to contain four times as much information.

The uu5 column component (<UU5.Brick.Column>) forms the basis of responsive behavior. The most important property is *colWidth*, which defines how many of the 12 columns the component spans and what screen size the settings apply for. For example, the settings “xs-12 s-12 m-6 l-4 xl-3” result in rendering the component on a mobile device (xs = extra small device) across all 12 columns, on small tablets or landscape phones (s = small device) it will be rendered across the entire screen, but on large tablets (m = medium device), the component will be rendered only over half of the columns (i.e., 6 columns), allowing two components to be rendered next to each other. On desktop devices (l = large device), the component will span four columns, and three components will be rendered on the screen next to each other. On a high-resolution device (xl = extra-large device), the component spans three columns, and four components will be rendered on the screen.

In order to make responsive behavior work as intended, the uu5 row component (<UU5.Bricks.Row>) has to encapsulate the column components. The row component ensures that all of its columns on mobile devices are lined up vertically. Vertical centering is difficult to achieve with standard CSS styles; the Flexbox layout module supports vertical centering and helps UI developers to overcome this problem.

3.6 Component Communication

Effective communication between components is critical when implementing complex UI applications. A number of different component communication methods can be used in practice. In uu5 (and React), the basic mechanism for communication between components is via component properties and by using component interface methods. When developers need to work with an instance of a specific component, React uses the *ref* property that saves the instance of a component to an instance variable. This is useful in situations where *child* components need to communicate with a parent component using interface calls (e.g., a parent form component communicates directly with a modal component via its interface methods using the *ref* property). In situations where there is no *parent-child* relationship, the uu5 framework provides a central component register (CCR) service that enables the registration of component instances under a specific key. This makes it possible to reference the saved component instances from any other component and call its interface methods.

3.7 Localization

Most applications need to be available in different languages. Native applications have built-in support for multiple languages, but multilingual support has been an issue for cross-platform applications. The uu5 framework is natively multilingual; it provides a Language Sensitive Item (LSI) component that responds to a change of the environment language.

3.8 Dynamic Rendering of Components

More complex cross-platform applications often consist of a range of components from different libraries provided by different vendors or development teams. The use of multiple libraries may result in a large amount of code being downloaded at runtime, and this can impact on application performance, in particular, if the Internet connection is slow. In order to alleviate such issues, the uu5 framework supports dynamic loading of components during runtime. The uu5 Component Registry service keeps track of library versions and always loads the latest version of the requested library.

4 uu5 Component Design Method

The uu5 methodology defines component design steps for different components types. Table 2 shows the design information used for different component types.

Basic information contains a brief summary of the component design, including a diagram that shows graphically the structure of the component and its main methods. In the case of a User View Entry (UVE) component, it contains UVE and its routes. Routes are represented by a table that contains all UVE routes. SPA component is represented by a diagram which contains routes (including route parameters that are part of a route URL) and events. Properties are represented by a table that contains a list of component properties, including component name, data type, and a default value. If a component has an interface, it is described in the interface section. Component static data and list of Mixins are also described. Finally, if a component is made up of other components, then these components are included in the component list section.

Components can be composed of other components using a combination of sequence, selection, and iteration operations. This approach is inspired by the Jackson Structured Programming (JSP) method [20]. The uu5 framework applies the JSP method for designing the structure of components.

Component diagrams are used to show the internal structure of components. An example of a component diagram is illustrated in Fig. 2.

The uu5 framework uses the Unicorn Universe Business Modeling Language (uuBML)—a visual modeling language based on Unified Modeling Language (UML). All techniques for building complex components based on sequence, selection, and iteration are fully supported by the uuBML visual notation. Key elements of uuBML notation used for component design are listed in Table 3.

Table 2 Design information and component types

Information	UVE	Route	Comp.
Basic information	X	X	X
Diagram	X	X	X
Routes	X		
SPA component	X		
Mock up designs	X	X	X
Graphic designs	X	X	X
Parameters		X	
Properties	X	X	X
Interface	X	X	X
Static data	X	X	X
Mixins	X	X	X
Component list	X	X	X

Fig. 2 Component diagram [21]

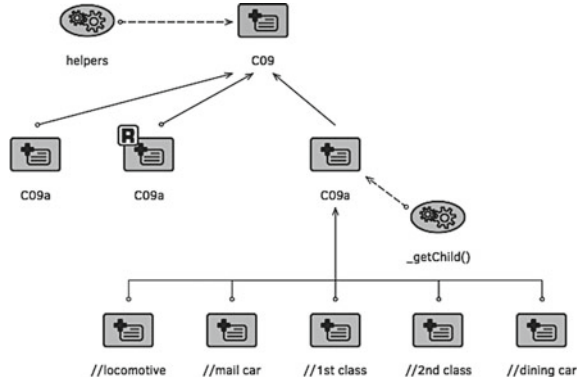


Table 3 Key elements of uuBML notation used in component design

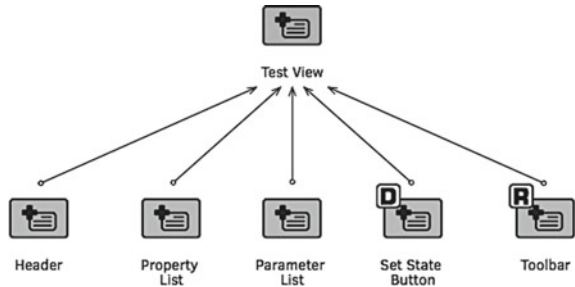
Symbol	Description
	Component
	Sequence
	Selection
	Iteration
	Method
	1:1 relationship represents existence of a component method

4.1 Sequence

The sequence is one of the main structures for building complex UI components. A sequence represents a set of components from which a complex component is constructed. Components can have labels that indicate that a component is only visible under certain conditions (D—disabled components), or that a component is rendered only under certain conditions (R—rendered component).

Figure 3 shows the use of a sequence of nested components within the Test View component that is made up of Header, Property List, Parameter List, Set State Button, and Toolbar nested components. The Set State Button component is disabled (D) and under specific conditions will be unavailable. The Toolbar component will be rendered (R) only under specific conditions.

Fig. 3 Sequence of nested components



4.2 Selection

A selection is used in situations where a complex component is composed of several nested components, but only one of the components is being displayed at the time, depending on a condition controlled by a method. A selection is often used in wizards that are composed of several steps, with only the active step displayed at a time. A wizard is implemented as a complex component, and the steps are implemented as nested components. Figure 4 shows the use of a selection of nested components of the Test Wizard component. The wizard has three steps represented by nested components: Basic Properties, Property Administration, and Parameter Administration. The `_getChild()` method returns a component which represents the active step within the Test Wizard component. Another example of the use of a selection is in situations where an application needs to display different nested components based on the device type (e.g., cards on smartphones and tablets and a table on desktop computers), or different type of view (map view vs. grid view vs. card view).

Fig. 4 Selection of nested components

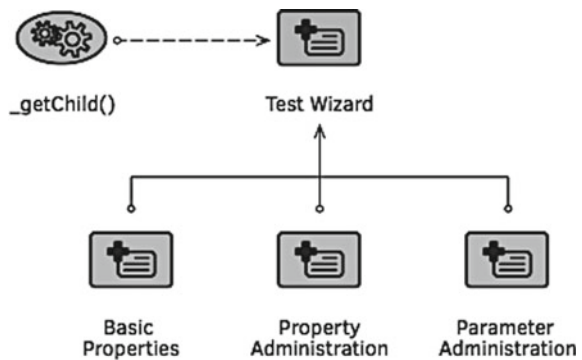
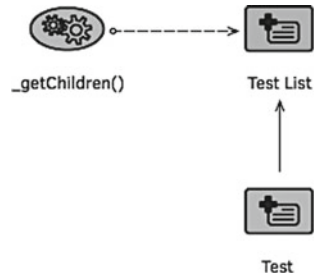


Fig. 5 Iteration of nested components



4.3 Iteration

An iteration is used in situations where a nested component has to be displayed several times. An example of an iteration is a table that displays a list of employees represented as a set of rows in the table (each row is a nested component).

Figure 5 shows the application of iteration in the Test List component which shows a list of performed tests visualized as a card. Each test is implemented as a nested component called Test. The `_getChildren()` method returns the number of times that the nested component is displayed.

5 Case Study

In this section, we illustrate the application of the uu5 framework for the design of cross-platform user interfaces of the uuCourseKit education delivery application. uuCourseKit is a cloud-based application that supports the development of variety of courses that can be accessed by students and external participants. The uuCourseKit application delivers self-learning online courses that use quizzes and knowledge cards to progressively evaluate course participants. Each course consists of blocks, and each block consists of topics. Each topic consists of lessons that are composed of interactive questions and knowledge cards [22].

The uuCourseKit application has three user view entry points (UVEs): Course Study, Course Content Administration, and Student Administration. The numbers of routes and components related to each UVE are shown in Table 4.

We have selected the CourseMenu route and the CourseMenuBlock components to illustrate the application of the uu5 component-based approach for UI design,

Table 4 Number of routes and components

UVE	No. of routes	No. of comp.
course	14	67
executivesContent	7	32
executivesStudent	1	10

showing component diagrams and the visual representation of the structure of the user interface.

The goal of the selected route is to provide the main portal and an entry point for the students of the course. Figure 6 shows graphical designs of the CourseMenu route. The screens illustrate different sizes of the topic cards that could be used on devices with different screen size.

Figure 7 shows a component diagram that depicts the structure of the CourseMenu route, including nested components and methods from which the route is constructed. The CourseMenu route consists of the body component, that consists of a sequence of six components: course welcome button row, prerequisite test and scan test row, block list, test menu button row, course rating button row, and the background image. The double forward-slash (“//”) in the component name indicates that the element is assembled from native uu5 components or their extensions and does not need to be separately specified. The prerequisite test and scan test rows are rendered under specific conditions as indicated by the label R. These components consist of sequences of two components: prerequisite test button and scan test button, also rendered under specific conditions. The conditions under which the components are rendered are described in the route design. The diagram shows the methods that are called when the prerequisite test button or scan test button is activated. The underscore “_” in front of the method name indicates a private (helper) method.

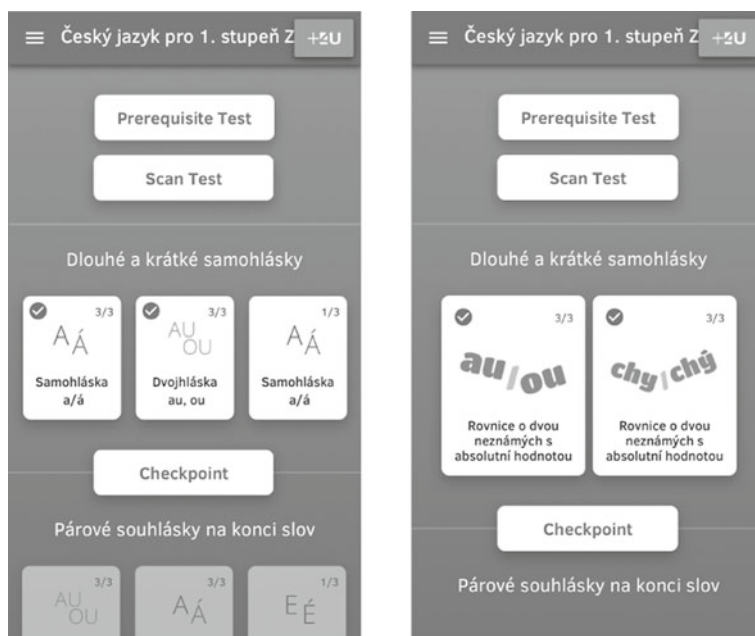


Fig. 6 Graphical designs of CourseMenu

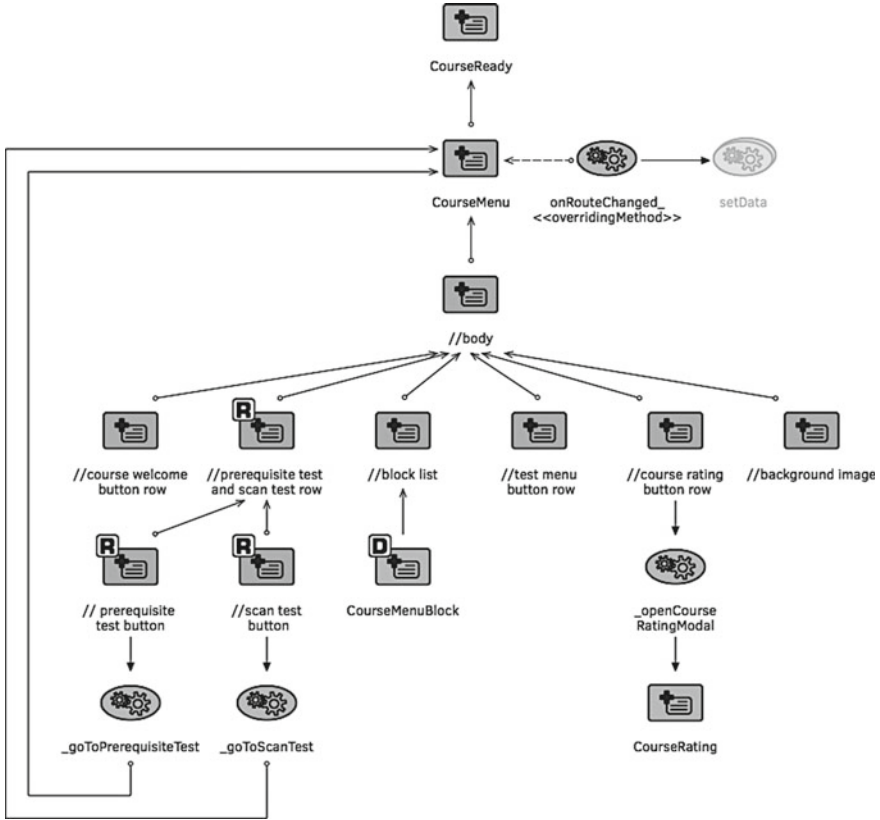


Fig. 7 Component diagram of CourseMenu

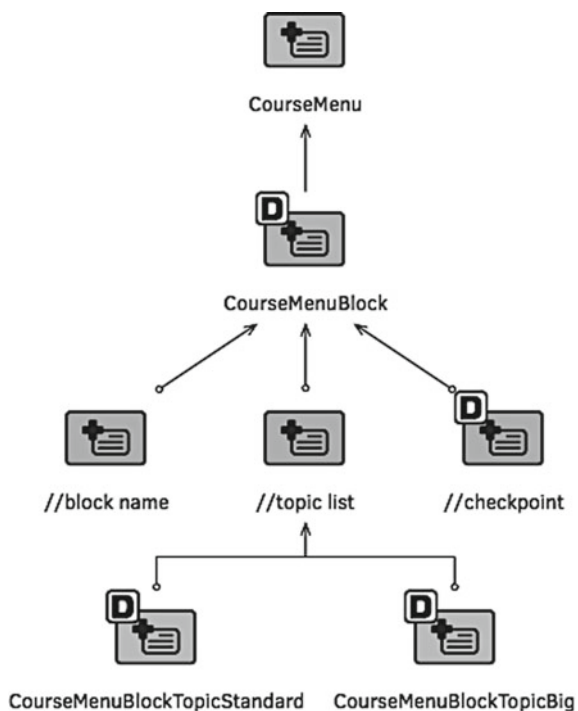
Block list consists of the iteration of the CourseMenuBlock component. The diagram also illustrates that the `_openCourseRatingModal` method is called from course rating button row. This helper opens the CourseRating component. The CourseMenuBlock and CourseRating components are described in greater detail in a separate design following the same rules and structure as described in Sect. 4. Figure 8 depicts the graphical design of the CourseMenuBlock component displaying different sizes of the topic cards (similar to the illustration in Fig. 6).

The CourseMenuBlock component structure is shown in Fig. 9. The component consists of a sequence of three components: *block name*, *topic list*, and *checkpoint*. The topic list component consists of a selection of two components displaying block topics in two sizes: *standard* and *big*. The topic list component, checkpoint, and both CourseMenuBlockTopic components are visible under conditions specified in the component design. Both CourseMenuBlockTopic components are described in greater detail in a separate design.



Fig. 8 Graphical design of CourseMenuBlock

Fig. 9 Component diagram of CourseMenuBlock



6 Conclusions

We have argued that most UI frameworks suffer from lack of extensibility forcing designers to implement user interfaces using a limited set of components. Furthermore, poor reusability associated with UI development leads to low application development productivity and high cost. We have described the uu5 component-based UI development method and illustrated this method using the uuCourseKit education delivery application. We have described the role of the uu5 library in supporting the implementation of cross-platform user interfaces based on HTML5 and JavaScript. The use of well-structured, loosely coupled UI components makes the development and maintenance of software easier and faster, improving code readability, and reducing software defects. Another major long-term benefit of component-based development is that applications running on different devices share the same code base, avoiding complex release management, and reducing the cost of application development.

References

1. A. Charland, B. Leroux, Mobile application development: web vs. native. *Commun. ACM* **54**(5), 49–53 (2011)
2. S. Xanthopoulos, S. Xinogalos, A comparative analysis of cross-platform development approaches for mobile applications, in *Proceedings of the 6th Balkan Conference in Informatics* (ACM, Thessaloniki, Greece, 2013), p. 213–220
3. J. Harjono et al., Building smarter web applications with HTML5, in *Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research* (IBM Corp., Toronto, Ontario, Canada, 2010) p. 402–403
4. P. Smutný, Mobile development tools and cross-platform solutions, in *Carpathian Control Conference (ICCC), 2012 13th International*, IEEE (2012)
5. T. Melamed, B. Clayton, *A Comparative Evaluation of HTML5 as a Pervasive Media Platform* (Springer, Heidelberg, 2010)
6. A. Holzinger, P. Treitler, W. Slany, *Making Apps Useable on Multiple Different Mobile Platforms: On Interoperability for Business Application Development on Smartphones* (Springer, Heidelberg, 2012)
7. P. Kruchten, *The Rational Unified Process: An Introduction*, 2nd edn. (Addison-Wesley Longman Publishing Co., Inc., 2000), p. 320
8. K. Fakhroudinov, *UML 2.5 Diagrams Overview* (Cited: 30 May 2018). Available from: <https://www.uml-diagrams.org/uml-25-diagrams.html> (2018)
9. M. Brambilla, A. Mauri, E. Umuhoza, Extending the Interaction Flow Modeling Language (IFML) for model driven development of mobile applications front end, in *Mobile Web Information Systems: 11th International Conference, MobiWIS 2014, Barcelona, Spain, August 27–29, 2014. Proceedings*, ed. by I. Awan et al (Springer International Publishing, Cham, 2014), p. 176–191
10. R. Acerbis et al., Model-driven development based on OMG’s IFML with WebRatio web and mobile platform, in *Engineering the Web in the Big Data Era: 15th International Conference, ICWE 2015, Rotterdam, The Netherlands, June 23–26, 2015, Proceedings*, ed. by P. Cimiano et al (Springer International Publishing, Cham, 2015) p. 605–608

11. ASP.NET - Single-Page Applications: Build Modern, Responsive Web Apps with ASP.NET. 2018 [Cited: 30 May 2018]; Available from: <https://msdn.microsoft.com/en-us/magazine/dn463786.aspx>
12. uu5g04 – Documentation, (Cited: 30 May 2018); Available from: <https://uuos9.plus4u.net/uu-bookkitg01-main/78462435-ed11ec379073476db0aa295ad6c00178>
13. D. Flanagan, *JavaScript—The Definitive Guide* (O’Reilly, Sebastopol, CA, 2006), p. 497
14. Facebook, *React—A JavaScript Library for Building User Interfaces* (Cited: 30 May 2018); Available from: <https://reactjs.org/index.html>
15. Google, *Material Design*. [Cited: 30 May 2018]; Available from: <https://material.io/>
16. Bootstrap, (Cited: 30 May 2018); Available from: <https://getbootstrap.com>
17. Facebook, *React. Component* (Cited: 30 May 2018); Available from: <https://reactjs.org/docs/react-component.html>
18. V. Kovar et al., *What Makes Up a Component*. uu5 Library—Documentation 2018 (Cited: 30 May 2018); Available from: <https://uuos9.plus4u.net/uu-dockitg01-main/78462435-ed11ec379073476db0aa295ad6c00178/book/page?code=howToDesignAComponent>
19. *Bootstrap Layout Overview* (Cited: 30 May 2018); Available from: <https://getbootstrap.com/docs/4.1/layout/overview/>
20. L. Ingevaldsson, *Jackson Structured Programming: A Practical Method of Program Design*, 2nd Revised edn. (Chartwell-Bratt, 1986)
21. V. Kovar et al., in *Example 09—Train Example*. uu5 Library—Documentation 2018 (Cited: 30 May 2018); Available from: https://uuos9.plus4u.net/uu-dockitg01-main/78462435-ed11ec379073476db0aa295ad6c00178/book/page?code=ee09_01
22. M. Beranek, V. Kovar, V. Vacek, Design of a learning management system for small and medium sized Universities and Colleges, in *International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government* (CSREA Press, USA, 2017) p. 22–25

Improvement of Vehicles Production by Means of Creating Intelligent Information System for the Verification of Manufacturability of Design Documentation



Irina Makarova , Ksenia Shubenkova , Timur Nikolaev and Krzysztof Żabiński 

Abstract The article presents one of the options for solving the optimization problem, the interaction between designer and technologist in the transition to the concept of Industry 4.0, by creating an intelligent information system for the verification of manufacturability design documentation of automotive company. The conceptual scheme is presented, and the interaction of modules and algorithms are described. The adequacy of the proposed solution is checked by conducting a multifactor computer experiment. Verification and validation of the system are based on real-world examples.

Keywords Manufacturability · Intelligent information system · Digital twin

1 Introduction

Reasonable and rational management and development of all spheres of human activities, including automotive industry, is associated with intellectualization. The high level of motorization and the globalization of markets force automakers to search for new solutions, to constantly improve both the vehicle's design and production technology. Hence, it is possible to withstand considerable competition in the markets only through the continuous development and application of innovative solutions. Latest technological achievements provide a transition from industrial automation into a new, fourth, stage of industrialization. Industry 4.0 is the German initiative's brand name that defines the prospects of future production. The combination of the

I. Makarova · K. Shubenkova · T. Nikolaev
Kazan Federal University, Naberezhnye Chelny, Russia
e-mail: kamivm@mail.ru

K. Shubenkova
e-mail: ksenia.shubenkova@gmail.com

K. Żabiński (✉)
Institute of Computer Science, University of Silesia, Katowice, Poland
e-mail: kzabinski@us.edu.pl

Internet's dissemination growth, mobile devices, the development of data analysis methods, the "Internet of Things" and machine learning is changing the expectations and requests of consumers. Digitalization helps to focus on the customer; so, mass production of a new item allows the industrial manufacturing of an individual product. Interconnected industry allows everything from design to manufacturing to be done through teamwork between products and machines, and between machines themselves. At the same time, the actual task is to organize the interaction between the designer and the technologist, i.e., checking the design documentation for manufacturability. Numerous studies have been devoted to study various possible improvements of these processes.

2 Problem's State-of-the-Art

Industry 4.0 has become a frequent topic at a number of conferences and even in the mass media. The document [1] deals with industry 4.0 in terms of benefits and risks that are not related to social aspects, but solely with serial production data in the automotive. Article [2] aims to determine the digital twins' role in industrial conditions for industry 4.0. The article analyzes how the concept originally born in the aerospace sector can be useful for the production sphere.

The article [3] discusses the possibilities of using the modeling potential within the framework of the smart factory concept. In the article [4], a manual on rethinking the product development processes is presented.

Simulation of products and production processes is widely used in the engineering phase. Faster optimization algorithms, increasing computer power, and the amount of available data can be used in the simulation area for real-time management and optimization of products and production systems—a concept often called digital twin (DT). The article [5] identifies the functionality and data models needed to provide real-time geometry, and how this concept allows you to move from mass production to the more individual one. The authors of the article [6] believe that since 3D CAD data includes not only model figures, but also part structures and specifications, you can reuse data in various types of post-production activities.

Automotive design depends on the purpose of usage. The document [7] discusses the development of automotive technology, taking into account the setting and verification of targets, as well as methods for optimization of this critical aspect of the engineering design process.

Authors of the article [8] proposed a multi-component structural units synthesizing method with maximum structural productivity and manufacturability. Authors use a multi-purpose genetic algorithm in combination with FEM analysis to obtain optimal Pareto solutions. Trade-offs between structural rigidity, total weight, components manufacturability (size and simplicity), and the compounds number have been analyzed. The authors of the article [9] argue that the previous projects' study is important to prevent repetition of problems and is usually implemented using parallel development and design for production (DFM). The purpose of this article is

to indicate the key factors for the effective reuse of production experience, which is believed to help reduce costs and improve product quality. Article [10] proposed a new method for supporting reliable disassembly planning in uncertainty conditions of the input product's quality.

In the authors' view [11], ICT innovations have begun to change traditional products to intelligent, smart ones. These changes, associated with the product, imply the need for new engineering processes. When developing the product, it is necessary to concentrate not only on the early phase of its lifecycle, but also on the phase of product usage. Personalization is a new production paradigm for meeting the customer's diverse needs. Document [12] proposes a framework for the effective creation of personalized products. The authors proposed an integrated structure to support personalization and is shown on the example of a personalized bicycle. The articles [13, 14] are devoted to data collection technologies for the creation of cyber-physical systems and DT.

3 Materials and Methods

The main quality criteria for technological machines include their performance, accuracy, and reliability. In addition, there are many other criteria. Criteria for the mechanisms quality measured quantitatively can be divided into five groups: kinematic and dynamic, energy, thermal, reliability, geometric and weight [15].

Qualitative evaluation characterizes the technological design of the construction in general, based on the engineer's experience—on the basis of ensuring interchangeability and permissible errors in units and aggregates mounting. Qualitative analysis is carried out by the method of expert assessments; quantitative evaluation is performed by statistical analysis methods. The quantitative estimation is based on engineering and calculation methods and is carried out according to design and technological features, and it can be performed according to planned indicators, when the technical specification sets the basic indicators of the technological quality of the product (TQP), and on non-planned indicators—when an alternative to TQP arises to choose the most rational design solution of a number of equivalents with respect to the properties under consideration. Constructive and technological consistency is one of the main principles of the most expedient preparation of production. The application of this principle allows us to consistently analyze the processes of science research, experimental design, and technological developments. It is known, for example, that when designing new products of mechanical engineering and instrument making, up to 80% of constructive solutions passes from product to product. Interaction between technologists and designers is carried out in the system named Teamcenter. This system allows you to process notifications from designers by technologists, but does not allow you to analyze design documentation (DD) for manufacturability.

The proposed information system should be integrated with two systems available at PC "KAMAZ": NX and Teamcenter. Integration with two systems will allow designers and technologists to coordinate notifications in a faster way by: automatic

product analysis; issuing a preliminary result; storage of the conclusion and results in a single place; feedback between technologists and designers; no need to download products and work in NX for technologists. The two-way interaction between a technologist and a designer at the stage of designing and starting manufacturing of a new product allows not only to speed up the processes, but also to optimize them. The main criteria for evaluating the product for manufacturability are the coefficient of novelty and complexity. Formalization and automation of these actions allow you to optimize the technologists' operation. We propose automation of the product's manufacturability evaluation process by implementing the algorithm calculating the corresponding coefficients (Fig. 1). The individual surface complexity C_i depends, first of all, on its curvature degree Cur (flat, single, double) and also on the additional attributive information's volume A associated to a given surface (roughness, geometric tolerances). In addition, each surface is included in some structural element F_j of the part. To take into account the elements complexity, one must take into account the surfaces number N adjacent to the one under consideration. The bigger this number, the more the element is geometrically complex.

Mathematical models were developed to formalize the manufacturability evaluation process. The i th surface complexity function:

$$C_i = f_1(Cur) + N^2 + R + T_{geom}^3 + T_{dim} + N_{dim}^2 \tag{1}$$

where f_1 is the surface curvature function.

$$f_1(Cur) = \begin{cases} 1, & Cur = 0 \\ 2, & Cur = 1 \\ 8, & Cur = 2 \end{cases} \tag{2}$$

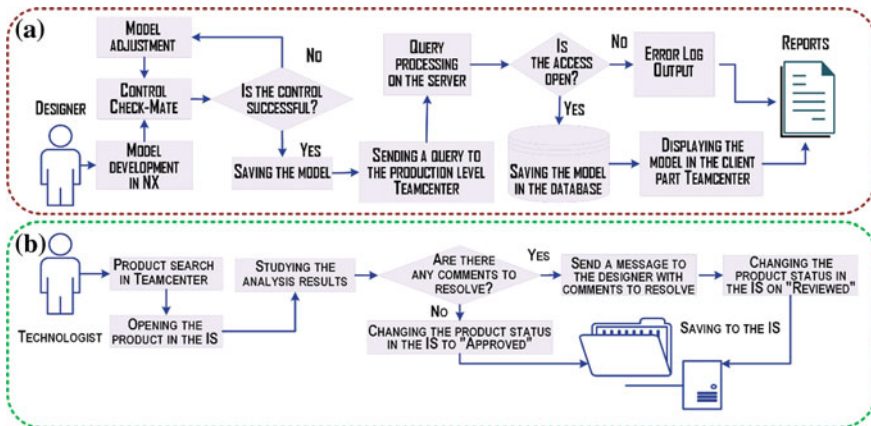


Fig. 1 Activities algorithm for a the designer; b the technologist

where N is the conjugate surfaces number; R —surface roughness coefficient ($R = 1, 2$); T_{geom} —the geometric tolerances number, affixed to the surface; T_{dim} —the number of actual dimensions on the surface; N_{dim} —actual number of dimensions on the surface (not taken into account when calculating the surfaces of the base sample). Constants for formulas (1) and (2) are obtained by the expert estimates method and can be refined in further research time.

The detail’s complexity consisting of m constructive elements is expressed as the relative sum of each element complexities:

$$C = \frac{\sum_{j=1}^m C_j}{m} \tag{3}$$

To determine the complexity of the projected units in comparison with its base sample, it is necessary to determine the complexity of each part, and then calculate the complexity increment:

$$C_{\text{com}} = C_1 - C_2 \tag{4}$$

where C_1 is the complexity of the compared product and C_2 is the complexity of the base model.

The complexity degree definition is as follows

$$C_{\text{cd}} = \frac{|C_{\text{com}}|}{C} \cdot 100\% \tag{5}$$

where $|C_{\text{com}}|$ —increase in complexity; C —a product complexity quantitative indicator (if $|C_{\text{com}}| \geq 0$, then $C = C_1$, otherwise $C = C_2$).

An algorithm reducing the analysis laboriousness of large assembly units inter-project (mutual) unification has been developed

$$\begin{aligned} \text{If } m < n_{\text{max}} \quad K_{\text{MU}} &= \left[\frac{\left(\sum_{i=1}^H n_i - Q \right)}{\left(\sum_{i=1}^H n_i - n_{\text{max}} \right)} \right] \cdot 100\% \\ \text{If } m > n_{\text{max}} \quad K_{\text{MU}} &= \left[\frac{\left(\sum_{i=1}^H n_i - Q \right)}{\left(\sum_{i=1}^H n_i - m \right)} \right] \cdot 100\% \end{aligned} \tag{6}$$

where H is the total number of the product groups under analysis; m —the list of names of component parts uniting the product group; n_i —the total components number in the i th product group; $Q = \sum_{i=1}^m q_i$ —the total components names number used in the products group under analysis; q_i —the name of the i th constituent part number; n_{max} —the maximum components number from the product groups being analyzed.

Repeatability factor allows you to estimate the duplication (repeatability, borrowing) percentage of the product components under comparison. The calculation is carried out according to the formula:

$$K_R = \frac{N_S - N}{N_S} \cdot 100\% \tag{7}$$

where N is the number of different constituents; N_S is the total number of constituents.

The modularity coefficient, which is the ratio of the parts number included in the blocks to the total number of vehicles parts, is calculated by the formula:

$$K_{mod} = \frac{N_{as.un.}}{N_{tot.}} \cdot 100\% \tag{8}$$

where $N_{as.un.}$ —the number of parts included in the assembly units; $N_{tot.}$ —total number of parts in the product (Fig. 2).

In addition, a methodology for calculating the applicability coefficient with recognition of original parts and units having the status of “EP” (Experimental Production) has been developed. The goal is to check the number of components adopted from other assemblies. The calculation is carried out according to the formula:

$$K_{app} = \frac{n - n_n}{n} \cdot 100\% \tag{9}$$

where n_n —is the number of novel components; n —is the total number of components.

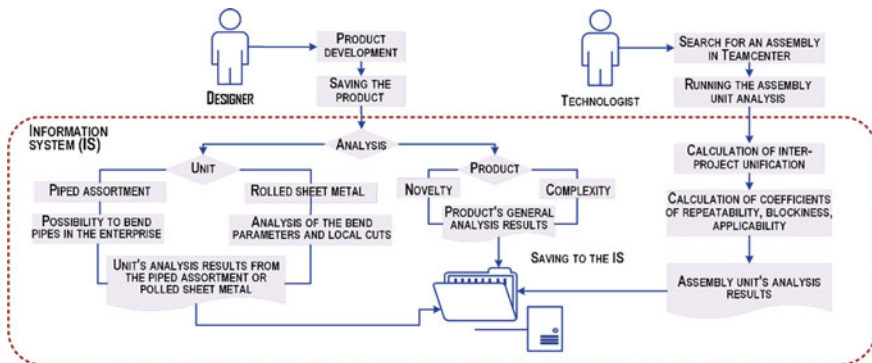


Fig. 2 Interaction of a designer and a technologist in IS

4 Results and Discussion

Manufacturability evaluation implies at the first stage analysis of each detail which is a part of the assembly unit. 3D model of a sheet metal product is analyzed for the following attributes: input data verification (material, thickness of roll-sheet material, bending radius); control accessories of “product’s material” to the category of rolled sheet products; calculation of previously created product’s minimum allowable bending radius R_{min} mathematical model; comparison of the minimum permissible bending radius R_{min} with the product’s bending radius R : if $R_{min} \leq R$, then the product is ready for production; if $R_{min} > R$ products, then correction of CD is necessary.

Conclusions about the product’s manufacturability are based on the results of analyses of the following scenarios: (1) The product manufacturability is in production. (2) The product manufacturability is in production with correction of DD: (a) it is required to increase the bending radius of the product, so that $R_{min} \leq R_{bend}$, and (b) the material of the product is missing. (3) There is no bending of the element on the 3D model. (4) It is required to choose a product, the material of which is roll-sheet steel.

Analogically, the analysis of the distance from the bending line to the protrusion angle of roll-sheet products obtained by cold sheet stamping/flexible is performed; rounding radii of products internal corners subjected to heat treatment; heights of the bent straight part of a shelf on product from rolled sheet metal; products from pipe assortment to possibility of cold bending; release of folds on products with ledges of rolled sheet. To verify the proposed algorithms adequacy, calculations were made using the proposed methods and forecasts of reduction in labor times using the developed IS. For example, let us consider a comparison of the parts’ 3D models shown in Fig. 3a, b.

The basic model contains three flat surfaces and three single curvature surfaces. The outer surface of the cylinder is marked with a roughness, and the flat ends of the bushing have parallel’s profile. The element’s complexity is calculated by the formulas 1–3. The model of the new element has three new flat surfaces and changes its complexity (Table 1). We obtained the calculated complexity equal to 19.3, the increment in complexity with respect to the base sample according to the formula 4 is 9.3, which is 48% with respect to the base sample (formula 5). To determine the

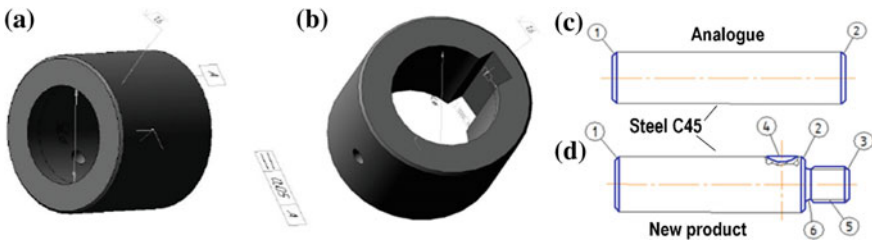


Fig. 3 a, b 3D models of basic and new products; c, d determining the product’s novelty

Table 1 Parameters comparison of the basic and new product

Surface	Curvature	Conjugate surfaces' number	Roughness	Geometrical tolerance	Actual dimensions	New dimensions	Complexity
Cylindrical (outer)	1	3	1.2	0	3	0	15.2
Cylindrical (internal)	1	3	0	0	3	0	14
Cylindrical (inner surface of the hole)	1	2	0	0	3	0	9
Flat (butt)	0	2	0	0	2	1	8
Flat (butt)	0	2	0	1	2	1	9
Flat (chamfer)	0	2	0	0	2	0	7
Flat (lateral surface of the groove)	0	4	0	0	4	4	37
Flat (lateral surface of the groove)	0	4	0	1	4	4	38
Flat (groove)	0	4	0	0	4	4	37

product's novelty, the following steps are needed: find analog; compare materials and components; define new elements, and then calculate the novelty degree as per formula:

$$N = \frac{N_{um1}}{N_{um2}} \cdot 100\% \quad (10)$$

where N_{um1} —number of new elements in relation to the product analog; N_{um2} —total number of elements in a new product.

An exemplary comparison of two products is shown in Fig. 3c, d. The products are made of the same material (Steel C45). The new product contains four new elements—a threaded surface (5), a chamfer (3), a groove (6), and a keyway (4). There are six elements in all. The calculated complexity degree, relative to the analog is as follows:

$$H = \frac{4}{6} * 100\% \cong 67\% \quad (11)$$

The degree of novelty “ N ” when choosing an analog should tend to 0. Thus, it is necessary to choose an analog having the lowest value of “ N .” If the degree of novelty is 70% or more, the product is completely new. With a novelty degree of 20–70%—it is a new type of a product, if less than 20%—it is a product of minor novelty.

5 Conclusions

Taking into account the results obtained, it can be observed that the developed system may reduce labor times by a factor of 8. It can be obtained by automation of design documentation analysis on manufacturability. This system will surely accelerate the interaction between designers and technologists (Fig. 2).

References

1. M. Švingerová, M. Melichar, Evaluation of process risks in Industry 4.0 environment, in *Proceedings of the 28th DAAAM International Symposium*, p. 1021–1029
2. R. Rosen, G. Wichert, G. Lo, K.D. Bettenhausen, About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine* **48–3**, 567–572 (2015)
3. G. Mahler et al., Cellular communication of traffic signal state to connected vehicles for arterial eco-driving, in *IEEE 20th International Conference on Intelligent Transportation Systems*, Yokohama (2017), p. 1–6
4. M. El-Sayed, Rethinking the automotive design and development processes for product realization. *SAE Technical Paper* (2008). <https://doi.org/10.4271/2008-01-0861>
5. R. Söderberg et al., Toward a Digital Twin for real-time geometry assurance in individualized production. *CIRP Ann. Manuf. Technol.* **66**, 137–140 (2017)

6. B.M. Hapuwatte, F. Badurdeen, I.S. Jawahir, Metrics-based Integrated Predictive Performance Models for Optimized Sustainable Product Design, p. 25–34
7. P. Lomangino, K. Sohn, Methods for target-setting and rule compliance in automotive engineering, in *International Mechanical Engineering Congress and Exposition. Design Engineering*, vol 1 and 2, Washington, DC, USA (2003), p. 33–38
8. N. Lyu, K. Saitou, Topology optimization of multi-component structures via decomposition-based assembly synthesis, in *29th Design Automation Conference*, Parts A and B. Chicago, Illinois, USA (2003), p. 269–281
9. P. Andersson, Manufacturing experience in a design context enabled by a service oriented PLM architecture, in *10th International Conference on Advanced Vehicle and Tire Technologies*, Brooklyn, New York, USA (2008). p. 257–265
10. M. Colledani, O. Battaïa. A decision support system to manage the quality of End-of-Life products in disassembly systems. *CIRP Ann. – Manuf. Technol.* **65**, 41–44 (2016)
11. M. Abramovici, J.C. Göbel, P. Savarino, Reconfiguration of smart products during their use phase based on virtual product twins. *CIRP Ann. – Manuf. Technol.* **66**, 165–168 (2017)
12. C. Tan et al., Product personalization enabled by assembly architecture and cyber physical systems. *CIRP Ann. – Manuf. Technol.* **66**, 33–36 (2017)
13. T.H.-J. Uhlemann, C. Lehmann, R. Steinhilper, The digital twin: realizing the cyber-physical production system for Industry 4.0. *Proc. CIRP* **61**, 335–340 (2017)
14. G.N. Schroeder et al., Digital twin data modeling with automationml and a communication methodology for data exchange. *FAC-Papers On Line* **49–30**, 012–017 (2016)
15. P. Zawadzki, Methodology of KBE system development for automated design of multivariant products, in *Advances in Manufacturing*, ed. by A. Hamrol, O. Ciszak, S. Legutko, M. Jurczyk. *Lecture Notes in Mechanical Engineering* (Springer, Heidelberg, 2018) p. 239–248

Group Delay Function Followed by Dynamic Programming Versus Multiscale-Product for Glottal Closure Instant Detection



Ghaya Smidi and Aicha Bouzid

Abstract This paper fits in the context of glottal closure instant (GCI) detection. Based on the algorithmic steps of DYPSA method and those set by the Multiscale-Product of LPC residual signal, MP(e(n)), method, it is proposed to compare the effect of Group Delay function followed by Dynamic Programming with that of the MP method, with a view to GCI detection. Results of this comparative approach are presented in terms of reliability, accuracy, and cost of execution in a clean environment. Also, the robustness of the performances of these two methods is compared in the presence of white and babble noise. According to all comparison made, we can conclude that the Group Delay function followed by Dynamic Programming is inefficient in the GCI detection compared to the MP method.

Keywords Glottal closure instant · Group Delay function · Dynamic Programming · DYPSA · Multiscale-Product · LPC residual signal

1 Introduction

The measurement of glottal closure instant (GCI) is a basic requirement in many speech processing applications, such as the speaker identification [1], the speech synthesis [2], the diagnosis of pathologies [3], the analysis stages in the context of handling different acoustic features of a variety of voice qualities [4]. Different methods are known by their reliability in terms of GCIs detection, among which we found the Group Delay (GD) method followed by Dynamic Programming (DP) applied to the LPC residual signal, namely DYPSA algorithm [5]. In fact, the GD is a function based on the characteristics of the global phase of the phase-minimum signals, like LPC residual signal. Also, Multiscale-Product (MP) method was applied

G. Smidi (✉)

Shaqra University, Riyadh, Kingdom of Saudi Arabia

e-mail: Smidi_ghaya@yahoo.fr

A. Bouzid

Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, SITI, Tunis, Tunisia

e-mail: bouzidacha@yahoo.fr

© Springer Nature Singapore Pte Ltd. 2020

X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_19

in the same signal for GCI detection [6]. The LPC residual signal, noted $e(n)$, is obtained by the inverse filtering of the pre-accentuated speech signal by the transfer of the estimated vocal tract. Given that these two last methods are based on the LPC residual signal, we propose to compare them, in terms of GCI detection performances; GD function followed by DP versus MP method.

As regards to the GD function [7], the operating principle is as follows: for the LPC residual signal, the average slope of the phase spectrum is zero [8]. The offset version of the same signal will have a phase spectrum like the original, but with an average slope proportional to the displacement. This means that the average slope of the phase spectrum is a function of the location of the excitation pulse. The characteristics of the systems manifest themselves as fluctuations in the phase spectrum, whereas the mean slope of the phase spectrum is determined by the moment of excitation with respect to the origin of the times. The GD function is smoothed using a three-point median filter to eliminate all discontinuities. The phase slope is calculated at each sampling instant to obtain the phase slope function. If the excitation time is in the middle of the frame, then the phase slope is zero. Therefore, the positive zero crossings of the phase slope function correspond to the significant excitation times. Finally, the candidates not considered by the phase slope function are recovered using the technical projection [5]. The DP is an algorithm for solving optimization problems, relying on a cost function.

As for the Multiscale-Product (MP), it is defined as the product of the coefficients of the wavelet transform for different scales. It allows by appropriate choice of scales, to raise the line of maxima and to reduce the presence of noise [9].

This paper is structured as follows: Sect. 2 presents an algorithmic comparison between DYPSA and $MP(e(n))$, in order to underline the effect of the GD function followed by DP versus that of the MP method when both applied on $e(n)$ signal. Section 3 presents the comparative results of GCI detection in terms of reliability, accuracy, and cost of execution, in a clean environment. Section 4 presents the comparative results of GCI detection in a noisy environment. Conclusion and perspectives are finally presented.

2 Proposed Comparison Approach

The proposed comparison approach examines the GCIs detection performance offered by the GD method followed by the DP versus the GCIs detection performance offered by the MP method when both applied on the LPC residual signal. Knowing that the application of GD method followed by the DP on the LPC residual signal ($e(n)$) is known under the title of DYPSA method, it is proposed to compare DYPSA with the MP method applied on the same signal noted $e(n)$. Figure 1 outlines the various steps of the respective GCIs detection algorithms: DYPSA, $MP(e(n))$.

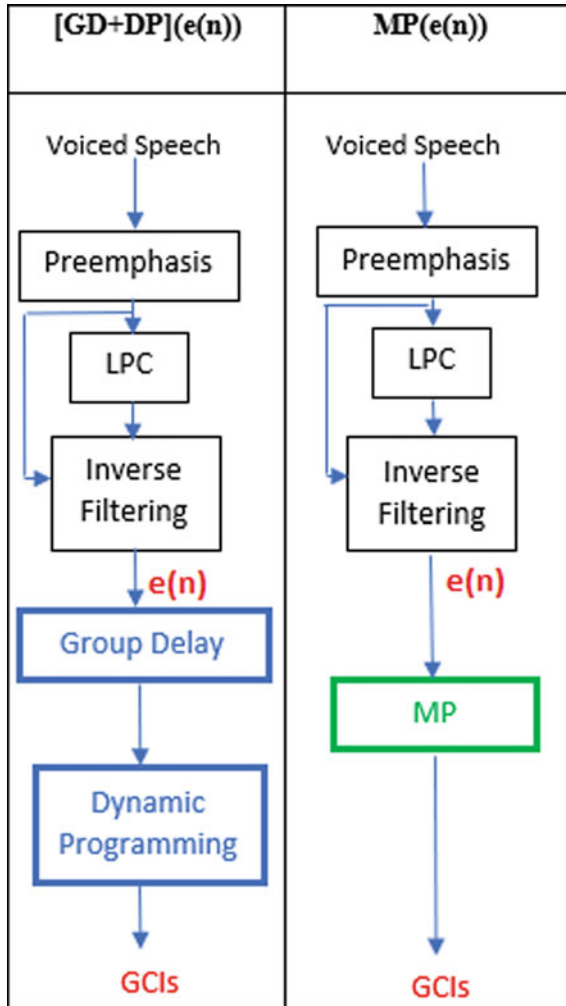


Fig. 1 Algorithmic comparison between DYPSA method and MP(e(n)) method, for GCI detection

3 Comparative Results in a Clean Environment

3.1 Comparison of Reliability and Accuracy

Table 1 shows the performance of the methods considered in a clean environment in terms of IDR, MR, FAR, IDA, and accuracy to ± 0.25 ms, using both Keele University database and the BDL-ARCTIC database.

Table 1 GCI detection reliability and accuracy given by [(GD + DP)](e(n)) versus MP(e(n)) on Keele and BDL databases

Databases	Method	IDR (%)	MR (%)	FAR (%)	IDA (ms)	Accuracy to 0.25 ms (%)
Keele	[(GD + DP)](e(n))	81.3	15.74	2.98	0.55	71.57
	MP(e(n))	97.8	1.83	0.38	0.30	92.07
BDL	[(GD + DP)](e(n))	95.54	2.12	2.34	0.42	83.74
	MP(e(n))	98.29	1.62	0.11	0.21	96.65

Evaluation metrics of the proposed comparison approach are defined as follows:

- **IDR(%)**: Identification Rate; the percentage of larynx cycles for which we detect only one GCI.
- **MR(%)**: Missing Rate; the percentage of larynx cycles for which no GCI detected.
- **FAR(%)**: False Alarm Rate; the percentage of larynx cycles for which more than one GCI is detected.
- **IDA(ms)**: Identification Accuracy; the standard deviation of the timing error distribution.
- **The accuracy to ± 0.25 ms(%)**: the percentage of detections for which the timing error is smaller than this bound.

Considering different comparative results given by Table 1, we note that:

- On the Keele University database, MP insures by far the best GCI identification rate (97.80 vs. 81.30%); this is due to the very high rate of missed GCI (15.74%) recorded during application of [(GD + DP)] on e(n). Regarding the precision, the MP applied to e(n) records a good accuracy, 92% of temporal errors are less than ± 0.25 ms, versus a mediocre accuracy provided by the application of GD followed by the DP on the same signal (71.57% of accuracy to 0.25 ms).
- On the BDL-ARCTIC database, the rate of missed values given by [(GD + DP)](e(n)) is visibly reduced, improving thus remarkably the identification rate given by this method (95.54%). Nevertheless, this latter rate remains much lower than this given by the MP (98.29%), because the false alarms rate still quite high. The accuracy to 0.25 ms given by applying the MP is exceptional (96.65%), clearly exceeding that given by applying the Group Delay function followed by the Dynamic Programming on the same signal.

This comparison allows us to conclude that for a reliable and accurate detection of GCIs, the MP method is much more efficient than the Group Delay function followed by Dynamic Programming. Histograms given by Figs. 2 and 3 presenting respectively the Identification Rate (%) and the Accuracy to 0.25 ms (%) given by MP(e(n)) compared to those given by [(GD + DP)](e(n)) confirm this last conclusion.

Fig. 2 Identification Rate (%) given by MP(e(n)) compared to that given by [GD + DP](e(n))

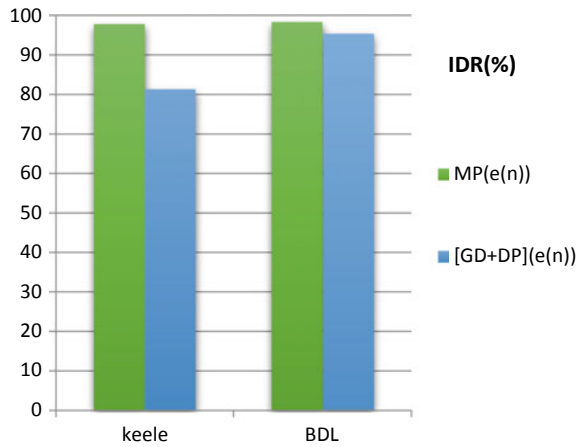
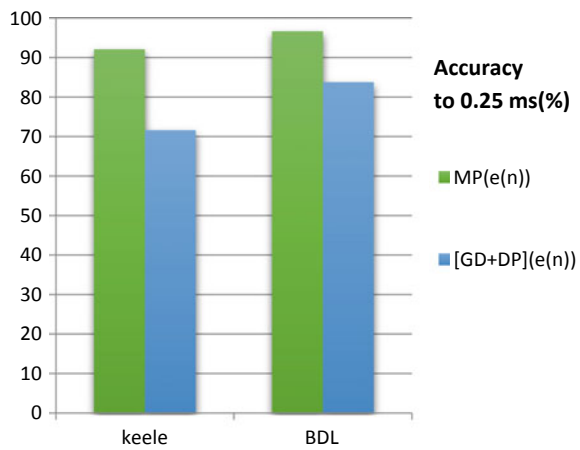


Fig. 3 Accuracy to 0.25 ms (%) given by MP(e(n)) compared to that given by [GD + DP](e(n))



3.2 Comparison of Execution Costs

In order to provide a complete comparison between MP and [GD + DP], when both applied to e(n) signal, a study of computational complexity is described in this section. In the case of DYPSA and MP(e(n)), we have to calculate the LPC residual signal, which makes the same computational load for the two methods; what remains concerns the load of the GD and the DP compared to that of MP. Specially, the load of calculation of Dynamic Programming is much heavier because of the large number of false GCIs candidates that must be eliminated, also the minimization of the cost vector, which concerns the measurement of waveform similarity during Dynamic Programming, presents a high computational load due to the large number of executions necessary to find the optimal path [10]. In order to compare

Table 2 Relative calculation time (RCT) for the evaluated methods, averaged across all BDL database speakers

Method	CPU(s)	RCT (%)	Duration of all BDL database(s)
MP(e(n))	526.176	16.24	3240
[(GD + DP)](e(n))	644.76	19.9	

the computation complexity, the relative computation time (RCT) of the considered methods is evaluated on the BDL database. The RCT in % is given by Eq. 1:

$$\text{RCT (\%)} = 100 \cdot \text{CPU time (s)} / \text{sound duration (s)} \quad (1)$$

Table 2 shows the averaged RCT obtained for considered methods using BDL-ARCTIC database.

4 Noise Robustness

We propose in this section to compare the performances of MP(e(n)) method with those of [GD + DP](e(n)) method, in a noisy environment, and thus to compare the Group Delay function followed by Dynamic Programming to the Multiscale-Product. The robustness of both methods is evaluated in the presence of white and babble noise based on the Keele University database.

4.1 Robustness to White Noise

Figure 4 illustrates the robustness of reliability of the two methods against white noise with an SNR ranging from -5 to 10 dB.

By examining the representative curves presented by Fig. 4, between 10 and 0 dB, we note that the slope of the curve representing the identification rate of the method [GD + DP](e(n)) is much greater than that of the MP(e(n)) method. In fact, [GD + DP](e(n)) has 9.82% of degradation in terms of identification rate between 10 and 0 versus 3.38% of degradation for MP(e(n)). Indeed, the graph representing the rate of false alarms of the GCIs detected by [GD + DP](e(n)) has an increasing slope with noise between 0 and 10 dB. For an SNR less than 0 dB, both methods are almost similarly affected. However, under severe noise conditions (-5 dB), the MP method has the best identification rate close to 90% .

Figure 5 illustrates robustness of accuracy of the two methods against white noise with an SNR ranging from -5 to 10 dB.

As shown in Fig. 5, the rate of errors less than ± 0.25 ms of the GCIs detected by the [GD + DP] method is much more affected by white noise than that given by the

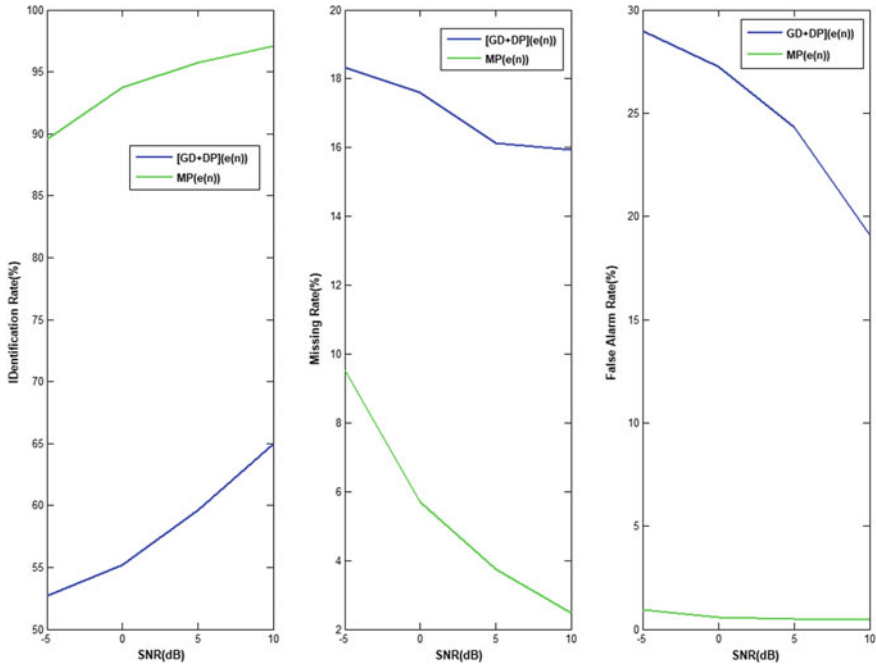


Fig. 4 Comparative illustration of the robustness of the reliability given by MP(e(n)) with that given by [GD + DP](e(n)), in the presence of white noise

MP method. The MP(e(n)) method keeps the best rate of errors less than ± 0.25 ms, under -5 dB of white noise (64.85%). The standard deviation of the two methods is similarly affected between -5 and 5 dB of white noise. But between 5 and 10 dB, the standard deviation of the MP method (e(n)) is a little less affected. The MP method exhibits GCI detection with the lowest standard deviation at -5 dB of white noise.

4.2 Robustness to Babble Noise

Figure 6 illustrates robustness of reliability of the two methods against babble noise with an SNR ranging from -5 to 10 dB.

By examining the curves given in Fig. 6, we can note that the reliability performances given by MP(e(n)) are remarkably affected in the presence of babble noise; in fact, the missing rate recorded by MP(e(n)) is strictly increasing with the severity of noise which masks the GCI’s peaks. Even though the identification rate given by [GD + DP] is constant between 0 and 5 dB, that given by MP remains much better regardless of the babble noise rate.

Figure 7 illustrates robustness of accuracy of the two methods against babble noise with an SNR ranging from -5 to 10 dB.

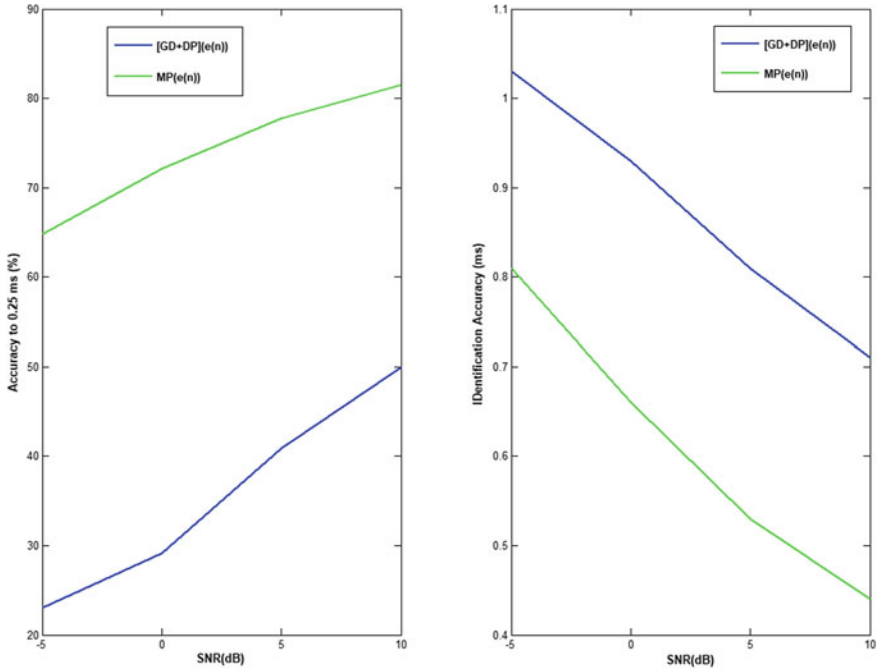


Fig. 5 Comparative illustration of the robustness of the accuracy given by MP(e(n)) with that given by [GD + DP](e(n)) in the presence of white noise

By examining the curves given in Fig. 7, it can also be deduced that accuracy given by MP(e(n)) in the presence of babble noise is remarkably affected; the MP keeps the best rate of errors less than 0.25 ms, but concerning the Identification Accuracy, [GD + DP] shows a little better standard deviation than that recorded by MP, respectively 1 ms versus 1.16 ms at -5 dB of babble noise.

5 Conclusion

This paper presents a comparative approach between two GCIs detection methods, both based on the LPC residual signal, to highlight the effect of the Group Delay function, yet followed by Dynamic Programming versus the simple application of the Multiscale-Product. By comparing the MP(e(n)) with [(GD + DP)](e(n)) in a clean environment, we find that the MP provides a higher identification rate with a much better accuracy and much lower execution time than that put by [(GD + DP)]. Under white noise conditions, reliability and accuracy of GCI detection given by MP are slightly affected compared to those given by [GD + DP]. Obviously, MP(e(n)) keeps the best performance under severe with noise conditions. Under babble noise,

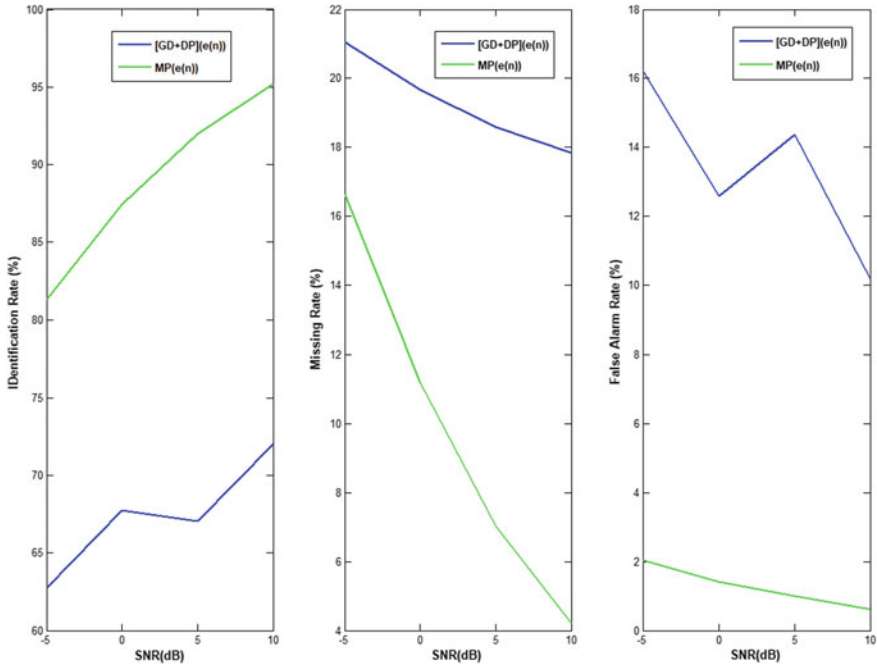


Fig. 6 Comparative illustration of the robustness of the reliability given by MP(e(n)) with that given by [GD + DP](e(n)), in the presence of babble noise

reliability and accuracy of GCI detection given by MP are visibly affected but this method keeps the best identification rate and the best accuracy to 0.25 ms at -5 dB of SNR. This leads to the conclusion that the Group Delay function followed by the Dynamic Programming is inefficient in the detection of GCIs, compared to the Multiscale-Product method. In future work, the efficiency of MP method compared to that of the Group Delay function followed by Dynamic Programming will be further highlighted, by comparing the YAGA algorithm with MP method applied on the derivative glottal waveform signal.

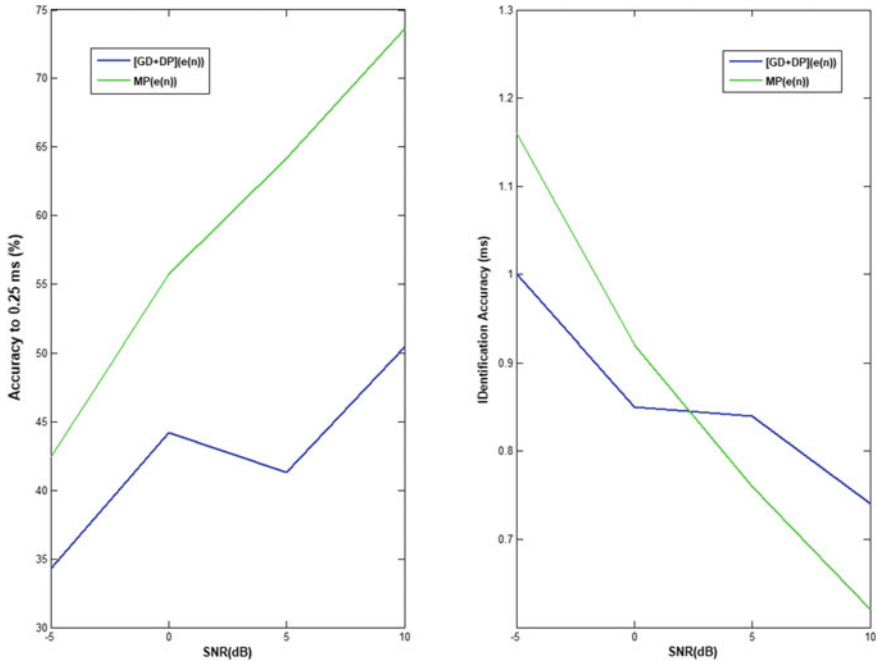


Fig. 7 Comparative illustration of the robustness of the accuracy given by MP(e(n)) with that given by [GD + DP]e(n), in the presence of babble noise

References

1. J. Gudnason, M. Brookes, Voice source cepstrum coefficients for speaker identification, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 31 Mar 2008. IEEE (2008), pp. 4821–4824
2. Y. Stylianou, Removing linear phase mismatches in concatenative speech synthesis. *IEEE Trans. Speech Audio Process.* **9**(3), 232–239 (2001)
3. M. Brookes, P.A. Naylor, J. Gudnason, A quantitative assessment of group delay methods for identifying glottal closures in voiced speech. *IEEE Trans. Audio, Speech Language Process.* **14**(2), 456–466 (2006)
4. J. Kane, C. Gobl, Evaluation of glottal closure instant detection in a range of voice qualities. *Speech Commun.* **55**(2), 295–314
5. P.A. Naylor, A. Kounoudes, J. Gudnason, M. Brookes, Estimation of glottal closure instants in voiced speech using the DYPSA algorithm
6. G. Smidi, A. Bouzid, N. Ellouze, Localisation Temporelle des Discontinuités du Résidu LPC par le Produit Multi-échelle Application à la Détection des GCI en Parole Voisée: e-STAN^oe-STA 2014–1
7. R. Smits, B. Yegnanarayana, Determination of instants of significant excitation in speech using group delay function. *IEEE*
8. R. Smits, B. Yegnanarayana, Determination of instants of significant excitation in speech using group delay function. *IEEE Trans. Speech Audio Process.* **3**(5), 325–333 (1995)
9. A. Bouzid, N. Ellouze, Produit multiéchelle pour la détection des instants d'ouverture et de fermeture de la glotte sur le signal de parole. *JEP* 525–528 (2006)

10. T. Drugman, M. Thomas, J. Gudnason, P. Naylor, T. Dutoit, Detection of glottal closure instants from speech signals: a quantitative review. *IEEE Trans. Audio Speech Language Process.* **20**(3), 994–1006 (2012)

Activity Logging in a Bring Your Own Application Environment for Digital Forensics



Bernard Shibwabo Kasamani  and Duncan Litunya

Abstract The use of cloud applications introduces new challenges to information system's security. The idea of applications accessible from multiple devices and hosted or provided by third-party organisations brings new complications to IT security. In situations where organisations are embracing Bring Your Own Applications (BYOA) and where they allow use of free to public cloud applications within their networks, it is important for IT security experts to consider how to secure their BYOA environments and also monitor how these applications are used and the flow of information. The aim of this research was to develop a digital forensics-based solution for securing BYOA cloud environment. This solution can be used to improve security in an organisation implementing BYOA. The research focuses on free to public cloud applications, whereby security challenges are identified and security measures proposed. The security measures are enforced through the development of a customized solution. The solution has been developed using rapid application development (RAD) system methodology. Using Geany editor and Python programming language, the prototype developed relies on digital forensics artefacts to gather information about the usage of BYOAs. The solution captures digital forensics artefacts and stores them into a database as logs of the activity on Google Drive application. The solution demonstrates how digital forensics artefacts can be used to enhance security in a BYOA environment.

Keywords Digital forensics · IT security · BYOA · SaaS forensics · Cloud forensics in SaaS · Security intelligence in SaaS · Google Docs forensics · Activity logging · Cloud logging · Google Docs logging

B. S. Kasamani (✉) · D. Litunya
Strathmore University, Nairobi, Kenya
e-mail: bshibwabo@strathmore.edu

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_20

241

1 Introduction

Companies are embracing cloud computing and other models spawned from cloud computing like bring your own application (BYOA). There are many factors behind this: employee productivity, staff mobility, costs—it is cheaper than enterprise solutions especially for smaller organisations and also considering licensing and other related software costs [1, 2].

However, the risks of adopting this model of computing are also significant. There is almost complete loss of control of applications and, to some extent, data. Organisations need to get back some of the controls [3, 4]. In some continents, especially for small and medium enterprises (SMEs), BYOA is very attractive but the same cannot be said in organisations where security is of high importance. Implementing BYOA involves balancing security requirement, budgets available and the risk appetite of the organisation.

Cloud computing is an approach that provides ubiquitous, convenient, on-demand network access to a shared pool of computing resources enabling organisations to increase computing capacity or capabilities without heavily investing in capital expenditure. Cloud computing has transformed how IT services are managed, accessed and delivered. There are various types of cloud computing delivery models; IaaS—infrastructure as a service, PaaS—platform as a service and SaaS—software as a service [5].

Bring your own application sometimes also referred to a build your own application is a growing trend that allows employees to use their preferred applications for work purposes [3]. If implemented as a strategy, it has a very low cost barrier and together with bring your own device they form the cornerstone of bring your own everything strategy [6]. BYOA in respect to this research specifically looks at those delivered as SaaS model and as a free service to the public. The companies “selling” these applications as a service make money by getting more people to use the Internet and also through selling advertisements.

There are a wide variety of free to public cloud applications, popularly referred to as consumer versions/applications [6]. Such applications continue to be adopted as BYOA, sometimes even without the knowledge of the organisation itself, in such situations they are said to operate within “shadow IT” of the organisation. Employees use these consumer applications to access enterprise systems and also store organisational data [6]. Some of these cloud applications do not need installation into the device. They can just be accessed via Web browsers.

However, BYOA introduces additional security challenges. Consumer adoption of smartphones has encouraged the culture of “apps” which is a popular moniker used to refer to applications. Mobile phones have morphed into mini-computers; now, applications are providing the same experience in computers and in mobile devices [7]. This is driven by user’s demand to have similar experience on computers and mobile devices. This in turn makes users more comfortable. Enterprise systems have responded by breaking down “big systems” into modules and allowing users to choose what they want or feel comfortable with.

BYOA implementation poses security threat especially through information leakage [8]. Information can be moved from one point to another, or other crimes or violations can occur in a BYOA environment. When incidents do happen, an organisation will only have two sources of information where they have complete control: their device and their network.

Organisations are embracing BYOA for various reasons. With Internet penetration increasing, and the average Internet speeds also increasing, more people are using mobile devices and home computers to access enterprise applications. BYOA model also means that not everything is under the absolute control of the IT department of the organisation [1]. A third-party cloud provider is added into the picture. For IT security experts and digital forensics practitioners, examiners and researchers, this is a new challenge.

2 Methodology

The system development methodology used in this research is rapid application development (RAD). RAD is an incremental model where a prototype is produced and improved in an iterative approach based on input from users and developers. RAD provides the abilities to quickly develop an application and to make modifications when needed. RAD reduced the traditional waterfall model into four steps. These four steps are a compressed version of the waterfall model and form a cycle of iterations. These four steps put more emphasis in analysis and design.

The key phases in RAD were:

System Analysis: The problem was defined as this stage; it was the initial process to gather the specific requirements of the system. This was the planning phase that determined the system scope. A qualitative-exploratory research approach was undertaken. This was intended to provide answers to underlying issues, gain insights into the problem and discover new ideas of tackling the problem. It was useful in getting the real security issues for BYOA: Google Drive, looking at other measures that can improve the security and discover new ideas for implementing security.

The exploratory research was best considering the little knowledge available on this subject. Data collection process was not through fixed-response questions, it allowed capturing of opinions and personal choices and even deviation from the subject line but not the research objectives. RAD and the qualitative-exploratory research technique were complimentary; focus group discussions were conducted during the initial system analysis stage and then on each subsequent iteration. The focus group consisted of five IT managers who work for small companies ranging from 10 to 20 users. These companies have Google Docs and Google Drive operating within their shadow IT. The team was used to refine the solutions' functionalities by providing personal, technical and expert opinion.

System Design: The requirements were then analysed and transformed into logical and then physical systems specifications and descriptions. The data collected during system analysis was derived into system processes and functionalities, and further visually analysed using unified modelling language (UML). Processes within the system were defined including all critical system components.

Development/Construction Stage: The designs were translated into code. Geany, a text editor, was used to write the Python code. The workspace: computer and development platform were prepped; installation of the required software was done. System code was generated as well as database descriptions. The “dummy” accounts for Google Docs were opened. The coding was implemented to achieve all identified functionalities to produce the prototype. A prototype was built and then tested, and feedback was provided that was used to refine the prototype. The feedback and modification cycle continued until a final, acceptable version of the system emerged. The initial prototype had limited functionalities and was improved with feedback from the focus groups until a final acceptable product was developed.

Testing: The system was tested, and its functionalities were evaluated. The intention was not only to identify errors but to identify weaknesses and areas of improvement. Weaknesses and improvements were then integrated in the next iteration to improve the prototype. System testing was done using the iterative approach of RAD. The development was evolutionary and the system was improved as prototypes were produced and reviewed by the focus group. Testing was integrated all through the development cycle; prototypes were tested during every iterative cycle.

RAD drastically reduced the time required to develop the application; it also gave greater control over project to the developer. End-user satisfaction level was high because of the continuous involvement through the feedback processes within the methodology. Reusability of prototypes saved on time.

3 Proposed Solution

3.1 Forensics Artefact Acquisition

When Google Docs is installed, the following artefacts are created on the default installation location which is `c:\users\\AppData\Local\Google\Drive\user_default`; three SQLite databases—snapshot, sync_config and uploader.

Snapshot.DB: This database contains seven tables. Within these tables is all the information of actual history of synchronisation between the local computer and the cloud.

Sync_config.DB: This database provides information about the users account email address, local root path and Google Drive version.

Google Drive incorporates Google Docs and offers Web-based office suites applications such as Word documents that allow users to create and edit documents online while collaborating in real time with other users. To access or work on Google Docs offline, an extension must be enabled in Google Chrome browser. Google Drive client installation maintains different profiles for each user “C:\Users\\AppData\Local\Google\Drive\user_default”.

On installation of Google Docs, different keys and values are recorded into the registry, and these keys and values can be used to identify the Google Drive client version and user folder for synchronisation;

- i. SOFTWARE\Microsoft\Windows\CurrentVersion\Installer\Folders\
- ii. SOFTWARE\Google\Drive
- iii. NTUSER\Software\Microsoft\Windows\CurrentVersion\Run\GoogleDriveSync
- iv. NTUSER\Software\Classes.

During installation, configuration files are saved within the installation folder in the user profile, the executable and libraries are stored in the bin subfolder, and prefetch files are created in Windows prefetch folder. Some information can also be derived from RAM analysis, using string searches. A process analysis of googledrivesync.exe produces the results in Table 1.

Table 1 Process analysis

Code	Label
00000000\004bcd6e	“VS_VERSION_INFO”
00000000\004bcdca	“StringFileInfo”
00000000\004bcdee	“040904B0”
00000000\004bce06	“CompanyName”
00000000\004bce20	“Google”
00000000\004bce36	“FileDescription”
00000000\004bce58	“Google Drive”
00000000\004bce7a	“FileVersion”
00000000\004bce94	“2.34.5075.1619”
00000000\004bceba	“LegalCopyright”
00000000\004bcd8	“Google”
00000000\004bceee	“ProductName”
00000000\004bcf08	“Google Drive”
00000000\004bcf2a	“ProductVersion”
00000000\004bcf48	“2.34.5075.1619”
00000000\004bcf6e	“VarFileInfo”
00000000\004bcf8e	“Translation”

3.2 Data Storage Requirements

The application needs to capture data from SQLite and store it into MySQL. The data needs to be dated appropriately. The solution is required to run independently on each client. Central storage is required to capture and store data from various clients. Data is stored as system logs for easier retrieval. The logs are stored for retrieval and review if required.

3.3 Data Classification Requirements

The entire SQLite data can be captured and stored, but for logging purposes not all data is required. There needs to be some elimination of unwanted data and some form of “normalisation” to remove unnecessary data from log files. The data available through digital forensics is in the following table representations: Table 2 contains information about files uploaded into the cloud.

Column definitions:

Doc_ID: The first half of the characters for this field remains constant for every file uploaded by a user in their own Google Drive. The second half of characters keeps changing. The first 13 characters are similar for all files uploaded under the same user account even using different computers. An assumption can be made that the first half is attributed to the user account used to upload the document into Google drive. It can be used to uniquely identify a user account with a file. The Doc_ID also serves as the link via HTTP to the file. The Doc_ID utilizes a numeral system, which

Table 2 Table Cloud_entry

Column	Data type
Doc_id	Text (primary key)
Filename	Text
Modified	Integer
Created	Integer
Acl_role	Integer
Doc_type	Integer
Removed	Integer
Size	Integer
Checksum	Text
Shared	Integer
Resource type	Text
Original size	Integer
Original checksum	Text

seems to be autogenerated. Files created online seem to have longer Doc_IDs and do not seem to follow any pattern. Uploaded files have 28 character names (there were some few exceptions), while files created online have longer Doc_IDs more than 28 characters.

Filename: Actual filename of the file in the cloud sync folder.

Modified: Last date modified. This is in UNIX timestamp; the number of seconds since 1 January 1970 at the time of modification.

Created: The date the file was created in the cloud; this field will remain empty if a file is created locally and uploaded into the cloud.

Acl_role: This column defines the creator of the document files that have been created and shared by other users and then downloaded into Google drive should have a value of 1. Files uploaded or created by a user in their own Google Drive display a value of 0.

Doc_type: This column should assign documents values based on the key below. However, it does not seem to work for documents created locally and uploaded. Documents that have been created and uploaded are assigned a value “1” while documents generated using Google Docs are appropriately assigned other values as follows:

Document Type List:

- i. 0 = place holder for folders
- ii. 1 = Appears to be a place holder type for various file extensions. All files uploaded to the drive have this number
- iii. 2 = Google Presentation/slides
- iv. 3 = Google Form
- v. 4 = Google Spreadsheet
- vi. 5 = Google Drawing
- vii. 6 = Google Document
- viii. 12 = Google Map
- ix. 13 = Google Site.

Removed: All tests on this field did not result to anything significant to note. The value remained 0, thus, no assumptions or conclusion can be made.

Size: Size of the file. Folders do not appear to have values even if there are files inside them.

Checksum: MD5 hash of the files. When files are created in the cloud, they do not appear to get an MD5 hash. They get MD5 hash if they are locally placed in the Google Drive or uploaded via the Web through the upload feature Google has.

Shared: Shared files and folders are assigned 1; those not shared are 0—a representation of Boolean true or false.

Resource_type: Files uploaded are only defined as files. Folders are appropriately named folder whether created offline or online. Files created online are defined as document.

Original_size: all test done did not result to anything significant to note, the field remained Null therefore, no assumptions or conclusions can be made.

Table 3 Table Local_entry

Column	Data type
Inode	Integer (primary key)
Volume	Text
Filename	Text
Modified	Integer
Checksum	Text
Size	Integer
Is_folder	Integer

original_checksum: all test done did not result to anything significant to note, the field remained Null therefore, no assumptions or conclusions can be made.

Table 3 contains information about files stored locally.

Column definitions:

Inode_number: Unique inode number assigned to each file. Under the local_relations table, it refers to the child_inode_number and connects to the parent_inode_number. The assumption is that it is a pointer reference to the file.

Volume: This column represents the volume serial in decimal. By running the command Vol C: in Windows, the same value in hexadecimal is retrieved. The value is the same as all files are stored locally on the same volume.

Filename: Actual filename of the file in the local default sync folder.

Modified: Last date modified. This is a UNIX timestamp, i.e. the number of seconds since 1 January 1970 when file was modified.

Checksum: MD5 checksum of the file, as per calculated in the local default sync folder of the computer.

Size: File size measured in bytes.

Is_folder: This column defines whether a resource is a file or a folder. File is 0 and folder is 1.

Table 4 contains references between files and folders in the cloud.

Column definitions:

Child_doc_id: references the doc_id of the file

Parent_doc_id: references the doc_id of the folder

Table 5 contains reference information between files, folders and disc drive/volume.

Column definitions:

Table 4 Table Cloud_relations

Column	Data type
Child_doc_id	Text
Parent_doc_id	Text

Table 5 Table Local_relations

Column	Data type
Child_inode	Integer
Child_volume	Text
Parent_inode	Integer
Parent_volume	Text

Table 6 Table mappings

Column	Data type
Inode	Integer
Volume	Text
Doc_id	Text

Child_inode: references the file inode
 Child_volume: serial of local volume
 Parent_inode: references the folder inode
 Parent_volume: serial of local volume.

Table 6 contains reference found in other tables.
 These columns are already represented in other tables.

Two tables; Table volume_info and Table_overlay_status remain empty and did not populate any data.

3.4 Data Output

From the tables above, the following information, displayed in Table 7 is derived to form logging information for Google Drive.

This requires that an operation to join the tables is conducted and appropriate primary keys and foreign keys are identified in the tables.

Join operation involves cloud_entry + local_entry.

Table 7 Logging output

Column	From table?
Doc_id	Cloud_entry
Filename	Local_entry
Checksum	Local_entry
Share	Cloud_entry
Acl_role	Cloud_entry
Modified	Cloud_entry
Volume	Local_entry

3.5 *Digital Artefacts Explained*

Doc_id: This field can be used to map files to users. Using the first 13 characters, it is possible to identify the person who originally uploaded a document to the cloud. This can only be done when looking for users within the organisation.

Filename: This can be used to identify the file and document type using the extension. Getting this data from the table `local_entry` can also help identify both Windows-created and Google Docs-created documents. Files created in Google Drive have the extension `.gdoc`, `.gmap`, `.gform` and so on, while local files only have usual Windows extensions.

Checksum: This MD5 hash can be used to uniquely identify a resource. This data is collected from `local_entry` because in the `cloud_entry` table files created online do not have this field populated.

Share: This field will inform whether a resource is shared or not.

Acl_role: This field will indicate whether the file was created by the users or downloaded from another Google Drive shared resource.

Modified: This is the only populated timestamp. It provides information when the file was last accessed. This information is picked from cloud entry so that online access can also be recorded.

Volume: This column shows under which volume the file currently resides on local device.

For the purposes of such an application, some artefacts outside Google Drive are also important, the computer name and the timestamp of when any logging information is collected.

3.6 *Application*

The application runs silently on system shutdown or logoff and system restart, grabbing required information and appropriately storing in log files. The application does not require input from user but is set to run by an administrator and always run in the background silently. The application should collect information and store them appropriately into a MySQL database.

3.7 *System Design*

Figure 1 shows all information available for capture. However, the information is more than required as per the system requirements. The image is an extract from DBVizualizer analyzing the relationships between the various tables.

The tables `local_entry` and `cloud_entry` are joined, but the unrequired columns are left out to produce a log. The `docID` is chosen as the primary key, and it is thus

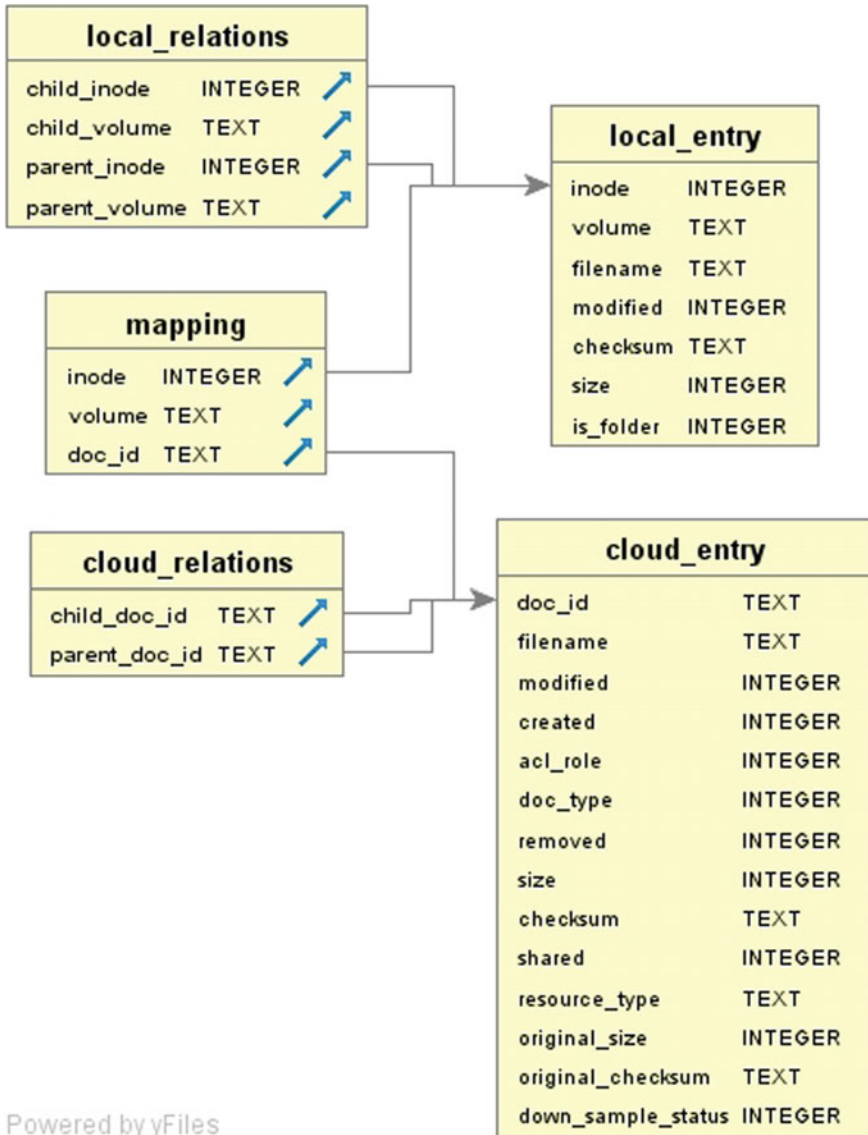


Fig. 1 SQLite tables

unique. The best way to capture the logs is through a trigger event. The trigger event should activate a background application, which silently creates the necessary log files. This event must be constant to ensure that the logs are created. The best trigger event is system shutdown or logoff or restart.

The system should pick data from SQLite database, create appropriate logs in a CSV file, and then transfer and store this data into MySQL. CSV is used as an intermediary to perform join actions from the required tables so that the data can be stored appropriately into MySQL.

3.8 Python Implementation Description

Step 1: Data Acquisition

The first operation is the capture of required information; this is done by joining of two tables `cloud_entry` and `local_entry` and only selecting the required information fields. This is achieved by capturing the data through creation of a CSV file.

Step 2: Connection to Database Server

The data is to be captured from various devices/computers. However, the captured information is stored centrally. Using MySQLdb Python module, a TCP/IP connection can be established.

Step 3: Create Database in MySQL

It is important to store data in a manner that it can be accessed and searched easily. If Google Drive is running in several computers, this information needs to be captured from all the computers. This application should have the capacity to create storage and appropriately store the logs for easy reference. A MySQL server is used to manage a database for each device/computer. The database stores daily activity logs. Python Code was used to create MySQL database and name it using computer name.

Step 4: Create Table and Insert Data

For purposes of daily logging, a table is created with an appropriate name (to make easy reference, the current date is used) to store the data. The data is then transferred from the CSV file into the MySQL table. Python Code was used to create tables in MySQL and insert data from CSV.

3.9 Use Interaction

Once the data is captured in the MySQL server, there is need to query and interact with the data. A simple user interface developed using Laravel Framework is developed.

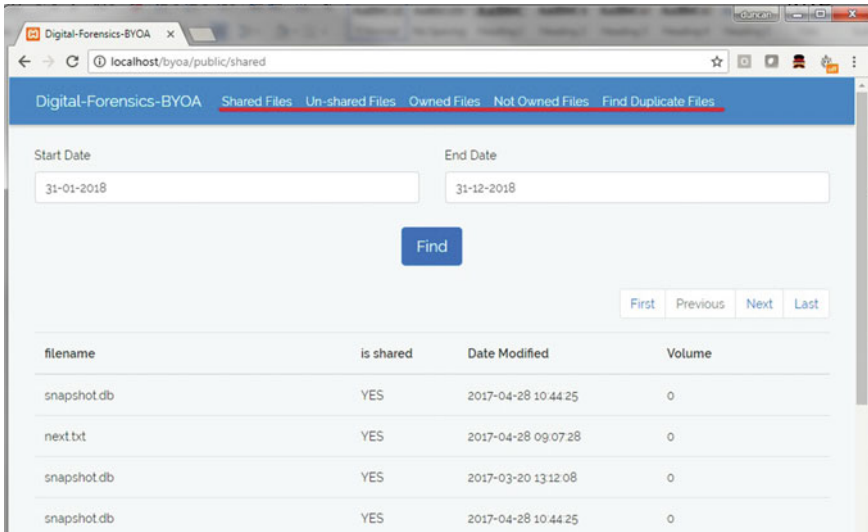


Fig. 2 A sample user interface

This simple user interface allows interaction with data stored in the database server. Though the user interface has predefined queries, it is only demonstrative and does not exhaust the type queries or interaction that can be performed on the data.

There is a functionality to filter using various criteria, like files shared/not shared, files stored in various user drives and are owned or not owned by the various users, duplicate files residing in the various user drives. Figure 2 shows a sample user interface.

3.10 System Testing

The testing process was undertaken in two forms: one is the testing of the application and whether it can efficiently run and capture the required data and the other is feedback from potential users on whether the tool adds value and improves the security setup in a BYOA environment.

Test 1: Data Acquisition → Creation of CSV File

This test is done to ensure the code for acquiring data and creating the consolidated CSV file works as required. Also, there is need to ensure that file activity is captured on the CSV file. The test steps are as follows:

- i. New files are created/copied.
- ii. Allow time for synchronisation.
- iii. Run the code to generate the CSV file.

- iv. Check whether new files are captured in CSV file.
- v. The expected result is that both files should appear. The file copied into the local repository should generate an appropriate “file Checksum”.

One file is copied into the local Google Drive repository and another is created online. The file copied is a JPEG: “tables-2.jpg”, and the online file created is “first spreadsheet”. Once the files have synchronised/replicated, the Python script for creating a CSV is executed. Figure 3 shows the results of the test.

The CSV file is created, and the details of the files are found in the CSV file contents. The red lines highlight the file names of the files created. The results are accurate, including the population of the checksum column for the local file and null for the online generated file.

Test 2: Creation of Device Database, Tables and Recording of data

This test involves testing the code that creates the database and tables and captures the data in the MySQL server. The database used for testing purposes is MySQL. The solution should create a separate database for each device, and under each database, tables are created and named according to dates for easier reference and referral. The steps for this test are as follows:

- i. The code is run.
- ii. A database should be created and named according to the computer the code ran from.
- iii. A table should be created under the current date.
- iv. The identified data should be captured into the table.
- v. The test is repeated daily for several days so as to confirm results.

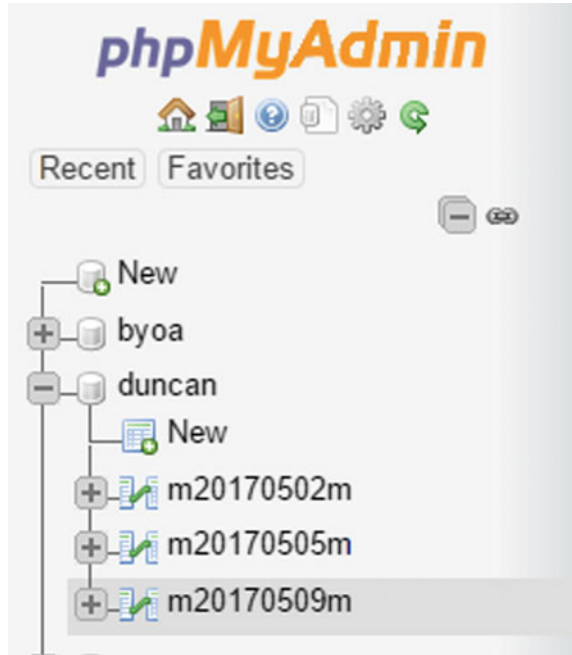
Figure 4 shows the test results which were successful.

A look into the MySQL server shows that a database similar to the computer name (Duncan) has been created and three tables within the database named using the format mYYYYMMDDm. The system managed to capture data for the computer “duncan” on three different dates: 02/05/2017, 05/05/2017 and 09/05/2017. A look into the tables also shows that data is appropriately captured.

Document ID	File Name	File Size	File Checksum	Resource Type	File/Folder Share	Date Modified	
2123	0800524H0C0cTVYVNGJwX0kVvYv8	634280	039f6c02e83948632de7b72d7739c3a	0	file	1437462461	
2124	0800524H0C0cTVJ5devVSTXNVLNU	182721	03304b33de94ea7af3c4e502f93c4c9e	0	file	1437663245	
2125	0800524H0C0cRhdELETwiTV0SFU	254902	d5192f68ca18186e8bed6876b2a653b7	0	file	1437462136	
2126	0800524H0C0cT2ZfalHwOHMIV0E	2497361	79759733770d08c1443187bd053ceb3c	0	file	1437760864	
2127	0800524H0C0cTvqNk13r8HbV08	464957	0fa0053f4a3af2975ca5c244610a0634	0	file	1435951079	
2128	0800524H0C0cTVVnWkZwGumY28	1266688	6c33c74cc150f8ade9088908f09240	0	file	1436284854	
2129	0800524H0C0cT112HFZCkAeHf	1389426	f8d362fc99118739f0e6a70f2a6ef	0	file	1436281296	
2130	19FZ244z9_MNGS1kxw87-rSR1ic			0	document	1492934161	
2131	1v0z1azr1A8ZL_cKqEYzyM-QcX0llymL			0	document	1492934211	
2132	1Vn0cUwFV0iWwHuDzUpkAa1RgZel_1st		5583	8b58ab7bd1dc71b50771efaa0974c3	0	document	1492934232
2133	0800524H0C0cT1V18hduXqJg6DU				0	file	1492940611
2134	1c0F9wZ6MkC193w1jH-IQ0ABev-Ty				0	document	1492953229
2135							
2136							

Fig. 3 Test results

Fig. 4 Test results



Test 3: Running Application in Different Device/Computer

This test was conducted by running the application in a different computer. The steps for the test are as follows:

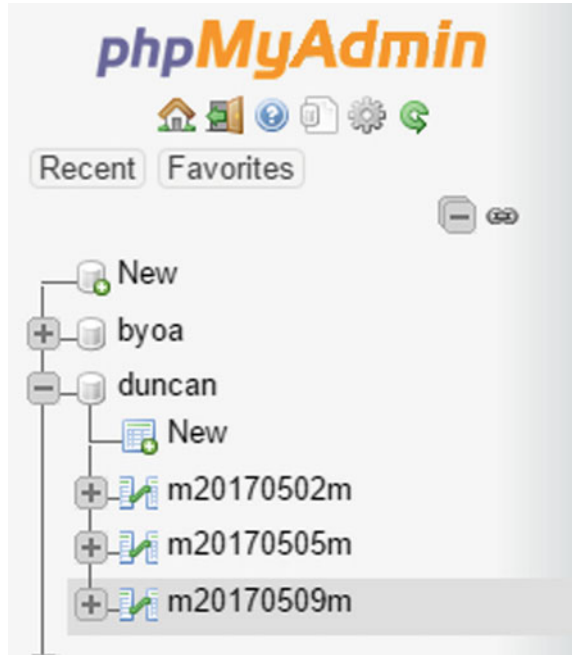
- i. Run application on different computer (named “litunya”).
- ii. New database is created with computer name.
- iii. New table is created with current date.

The results were positive, and a new database is created in the MySQL server and new table according to the date of execution which is 2nd November 2017. Figure 5 shows the old computer database with a red arrow and new computer database with a green arrow and the new table appropriately named.

4 Discussions

This research clearly brings out the role of activity logging and retrospective analysis. It shows the importance of activity logging and a comprehensive IT forensics support framework. An appropriate system development methodology and research technique was used to gather system requirements for a solution that can address activity logging requirements when implementing BYOA (Google Drive). Digital

Fig. 5 Test results (new database)



forensics artefacts were identified, how they can be obtained specifically for Google Drive and what they represent.

A study of the digital forensics artefacts is done to identify how the artefacts can be useful while logging BYOA activity. These artefacts are the core information for a logging application which can be used to enhance security. The application including a simple user interface is also developed to facilitate interaction with the collected data and sample queries tested. The system undergoes some simple tests to ensure that it is functional.

The proposed solution after undergoing repetitive enhancements and tests is finally evaluated through feedback from potential users. The feedback is positive, and the system can be used to enhance security through monitoring and logging of BYOA activities on devices.

The application provides actionable information. It collects information, which has been investigated and found to be relevant, from different computers. The information is stored in a structured manner that is easy to retrieve and query. In IT, security intelligence is a comprehensive approach that integrates multiple processes and practices designed to provide 360° protection. Log management or log data is an integral part of such systems, and this application demonstrates how BYOA can be monitored using digital forensics artefacts.

5 Conclusions

This research reviews the security challenges in BYOA. It investigated possible options that can be implemented to improve the security. BYOA is important to enterprises more importantly those with limited budgets. The proposed solution built for Google Drive demonstrates a concept that can work with most free to public applications including browsers. It provides a simple solution to logging of activities of Google Drive, allowing multiple devices to be monitored. The information is easily retrievable from the database.

As a limitation of this study, there is still need to accurately investigate the purposes of the various columns and tables that store data in SQLite on Google Drive usage. There seems to be a relationship between the Doc_ID and the user account. If such a connection can be confirmed and the how the field is generated, it may be possible to identify the user account using the Doc_ID.

Future research needs to take into consideration that cloud sync applications are now incorporating encryption. Dropbox already uses encryption—how could the activity of such an application be logged? The same may apply on Google Drive soon and other BYOA cloud applications.

Acknowledgements We thank all the authors, reviewers, editors-in-chief and participants for the great work. We thank the Strathmore University Institutional Ethics Review Committee (SU-IERC) for the ethical clearance to undertake this work. As required, all participants in this research duly consented before they could participate.

References

1. M. Rouse, in *Bring Your Own Apps (BYOA)*. Retrieved April 02, 2017, from TechTarget (2008). <http://searchsecurity.techtarget.com/definition/bring-your-own-apps-BYOA>
2. Comcast Business View, in *BYOD/BYOA: A Growing, Applicable Trend*. Retrieved April 05, 2017, from INC (2016). <http://www.inc.com/comcast/byod-byoa-a-growing-applicable-trend.html>
3. R. Patel, *Enterprise Mobility Strategy & Solutions*. Partridge (2014)
4. P.M. Grance, *The NIST Definition of Cloud Computing* (National Institute of Standards and Technology Laboratory, 2009)
5. J. Green, *Cyber Security: An Introduction for Non-Technical Managers* (Routledge, London, 2015)
6. W. Akpose, Cloud in the horizon: Opportunities and challenges for enterprises in the cloud economy. Gigma Associates (2014)
7. M. Mordhorst, How to help enterprises going mobile. Anchor Academic (2014)
8. R. Walters. Bringing IT out of the shadows. *Netw. Secur.* 1–20 (2013)

Game Theory for Wireless Sensor Network Security



Hanane Saidi, Driss Gretete and Adnane Addaim

Abstract This paper presents a comprehensive study on applications of game theory to face the security problems in wireless sensor networks. The dynamic nature of WSNs makes them vulnerable to different security challenges. In the literature, three techniques have been proposed to protect WSNs from malicious nodes, cryptography, trust-based methods, and game theory. In this paper, we review the major game theory approaches to overcome denial-of-service attacks and selfish behaviors in wireless sensor networks; we propose at the end the fusion of cryptography techniques and game theory for more privacy and security. Finally, we discuss some limitations of game theory solutions in WSNs.

Keywords Wireless sensor networks · Game theory · Security

1 Introduction

Wireless sensor networks are gaining more interest in the last years; their application is covering multiple fields such as healthcare monitoring and environmental sensing forest fire detection. These applications are growing thanks to the contributions of the researchers that try to improve their utilizations. Several works have been done to reduce the energy consumptions of the nodes and strengthen the links while reducing the time of data transmission. However, due to the hostile environments where WSNs are deployed, security represents a major challenge in these kinds of networks. The most well-known attacks against WSNs that degrade the performances of the nodes are denial-of-services and distributed DoS. To encounter these attacks, several security techniques have been proposed such as cryptographic mechanisms,

H. Saidi (✉) · D. Gretete · A. Addaim
Ibn Tofail University, Kenitra, Morocco
e-mail: hanan.saidi03@gmail.com

D. Gretete
e-mail: drissgretete@hotmail.com

A. Addaim
e-mail: addaim@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_21

reputation-based methods, and game theoretic approaches. In this paper, our focus is on the last protection technique. Game theory is an applied mathematics branch. It is precisely a strategy to cope with decision-making situations. John Von Newman and Oscar Morgenstern have invented game theory in 1944. Its applications have covered war strategies and then biology and economics to be widely used in science fields. Researchers thought of game theory techniques as a security solution for wireless sensor networks that involve safe and malicious nodes. This paper is divided into four sections: the first is about the main principle of the wireless sensor networks, the second will cope with the issues and attacks threatening the WSNs; the third will expand the game theory approaches; finally fourth is an approach to fusion cryptography techniques to game theory, for more security and reliability of the network.

2 Wireless Sensor Networks

2.1 Architecture

The popularity of WSN is tremendously growing. Their applications [1] in different fields are making the human life easier and safer since the sensing characteristics help with detecting WSN is a collection of sensor nodes distributed in an environment able to collect, to store, and to process data, and send it to neighboring nodes. The main feature of WSNs is that they transform the data collected from an environment into an electrical signal that can be processed. Figure shows the main components of a sensor node; the sensing unit composed of the analog-to-digital converter. The processing unit is the main component that exchanges the data saving, energy generating, and processing; other components include the power unit and the transceiver. Some sensors may contain more than these units such as location unit (Fig. 1).

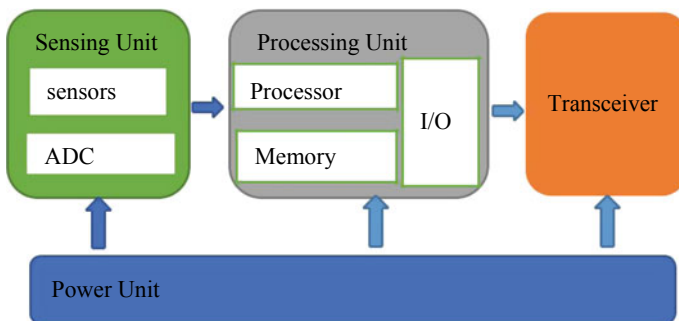


Fig. 1 Wireless sensor architecture

2.2 Routing Protocols

2.2.1 Traditional Routing

Routing protocols aim to find the best paths with low costs in terms of energy consumption constraints and link losses plus a small number of hops. Traditional routing techniques seem to be less performance compared to the new utilities given by the opportunistic routing approaches [2].

2.2.2 Opportunistic Routing

The main feature of the opportunistic routing is that it exploits the nature of the broadcast medium for data transmission [3]; each node participates at the sending depending on its distance from the destination. The next hop in the OR is selected after the transmission, whereas in the traditional routing the transmission node is selected before.

Opportunistic routing protocols [4] operate in three main steps: First data is broadcasted to a group of nodes (set) that have fully received the packet; this set runs a coordination protocol to choose the best node to forward the packet to the destination.

The coordination methods [5] are the most important parts of this chain because it makes the sending easy with a good quality. In the literature, the coordination methods are divided into: timer-based, token-based, and encoding-based.

Timer-Based

This type of coordination [6] is the easiest to implement, but one of its weakness is it allows duplication. To select the best relay from the network, the source sends the packet to all the nodes and then they are ranked by order of responding; the first to respond is the relay; this mechanism could also occur in a set of nodes to choose one of the CRS. Opportunistic wakeup MAC (OPWUM) [7] is a timer-based contention protocol for wireless sensor networks that allows selecting a relay with low cost of energy and preventing transmission duplication by allowing the relay to choose nodes from the neighboring.

Token-Based

This method has been proposed by Hosseinabadi and Vaidya [8]; a token sweeps all the nodes of the network starting by the destination (higher priority); if a relay is selected, an acknowledgment is injected in the token to avoid other nodes from transferring and thus avoid packet duplication, and this is the main advantage if this solution coped with timer-based coordination, but the cost in terms of control

Table 1 Comparison between TR and OR

	Broadcast nature	Number of relays	Relay	Selection	Time of candidate selection	Type of transmission
TR	Ignore	One	Fix		Before	Unicast
OR	Uses it	Multiple	Dynamic		After	Broadcast

packets and energy is high. Token DCF is a distributed MAC protocol that uses an overhearing technique to rank network stations for transmission on the wireless medium. The design goal of Token DCF is to decrease idle and collision time, which significantly improves the performance in terms of system throughput and access delay. O-ACK is an efficient MAC protocol, which improves the channel utilization by employing packet overhearing and eliminating explicit ACK frames. This protocol adjusts itself based on the surrounding environment. This protocol outperforms the DCF and Token DCF protocol.

Encoding-Based

Network coding is a technique [9] in which data is encoded at the source and decoded at the destination, to minimize the number of candidates and maximize the throughput there's basically no coordination overheard when an opportunistic routing is network coding-based which makes the wireless network duplication-free (Table 1).

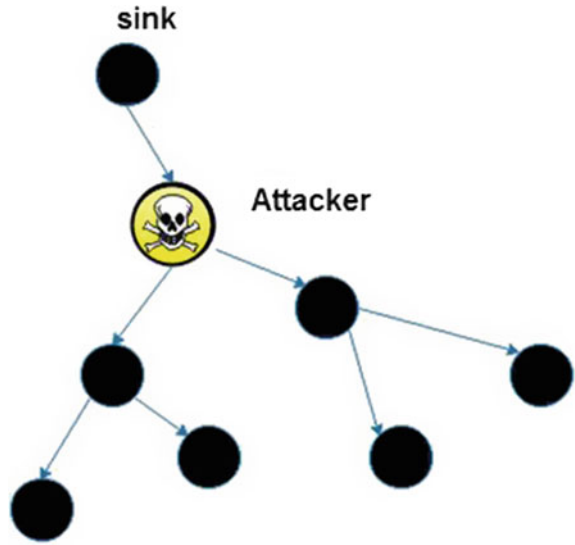
Attacks threatening WSNs

Attacks against wireless sensor networks could be on the hardware or on the routing protocols. We can notice less damage in network coding-based protocols due to the fact that they involve coding and decoding mechanisms but still both opportunistic routing and traditional routing are vulnerable to the same threats equally. Here, the major attacks in wireless sensor network schemes.

2.3 Blackhole Attack

Figure 2 shows that in a sinkhole or blackhole [10] attack, a malicious node collects all the traffic in the sensor network, and instead of sending them to the destination, it drops them; in another scenario, the attacker may spread false information between the nodes to make the illusion that the packets are correctly forwarded.

Fig. 2 Blackhole attack illustration



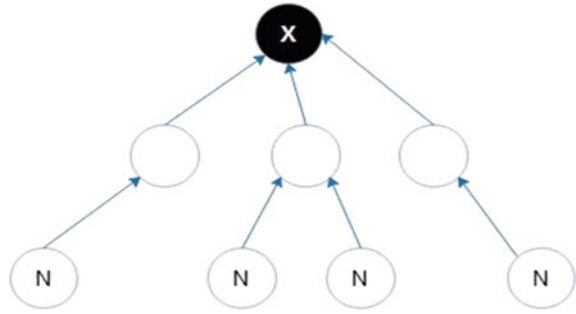
2.4 Sybil Attack

In the Sybil attack [11], a malicious node illegitimately takes multiple identities. The aim of this attack is to degrade the routing, data integrity, security, and energy. The peer to peer are more vulnerable to Sybil, whereas in wireless sensor networks it could be avoided by using correct protocols. Known targets of Sybil are the distributed storage, routing protocols, voting, data aggregation, resource allocation, and misbehavior detection.

2.5 HELLO Flood Attack

Discovery protocols in wireless schemes use HELLO messages to discover neighboring nodes. In a HELLO flood attack, the attacker uses these packets to saturate the network and consume its energy. The malicious node X in the figure has a powerful connection that allows it to send HELLO messages to a large number of nodes in a continuous manner. The neighboring nodes N will then try to answer it, even if they are located at far distances from the malicious node. By dint of trying to answer these messages, they will gradually consume all of their energy (Fig. 3).

Fig. 3 Sybil attack illustration



2.6 Denial-of-Services

Denial-of-service, also called denial-of-service attack, in a computer network is an attack carried out in order to harm the normal operation of this network.

There are many ways to proceed, and there is therefore a multitude of existing denial-of-service attacks. The state of the art in this field has the particularity that it includes two points of view: that of the attacker and that of the “defender.” It is essential to be able to define the model of an attack to be able to propose adequate countermeasures. In addition, more or less reciprocally, the protective mechanisms put in place over time push attackers (or researchers) to develop new attacks to circumvent them. Sensor networks are unfortunately very exposed to attacks denial-of-service, due to:

- Their extremely limited resources, and mainly in terms of energy;
- Their weak capabilities, which can introduce delays (latency in communications or processing time);
- Their exposure to physical attacks;
- The low reliability of the transmission medium, about confidentiality or collisions;
- Their remote management;
- The lack of centralized management (and the impossibility of knowing precisely the status of other nodes).

3 Game Theory

Game theory is a branch of applied mathematics that copes with decision-making situations involving multiple players. A game is generally based on a number of players that may behave in opposition or cooperation, to receive payoffs according their actions.

The following definitions are the fundamental rules of the game theory:

Players: An entity involved in a game beside a set of players; the players in a WSN are the nodes.

The strategy: It is a plan that gives the main rules that a game and the players will follow.

Payoffs: It's also called the utility is a positive or negative reward given to a player for a work or an action. Nash equilibrium: It consists of list of strategies, one for each player, which has a property that no player can unilaterally change his strategy.

3.1 Game Theory for Wireless Sensor Networks

The game theory methods used in wireless sensor networks are classified; non-cooperative game focuses on the behaviors between the nodes and their strategies; the players act in a selfish mode to cooperative games that represent a coalition between a number of nodes for the purpose to enhance the security of a network against internal and external attacks. The utility function of the cooperative game is constituted by three main parameters known as the security level between the set of nodes, the cooperation, and the reputation; the main categories of the cooperative game are:

Bargaining game: It is an agreement between the cooperating nodes to reach a Nash equilibrium solution. The nodes maximize the WSN performances to reach these goals. The parameters the node should be aware of are: fairness of resource allocation, no selfish nodes, the quality of services, the nature of the broadcast medium.

Repeated game: In strategic or static games, the players make their decisions simultaneously at the beginning of the game. On the contrary, the model of an extensive game defines the possible orders of the events. The players can make decisions during the game, and they can react to other players' decisions. Extensive games can be finite or infinite. A class of extensive games is repeated games, in which a game is played numerous times and the players can observe the outcome of the previous game before attending the next repetition.

Coalition game: A strategy to predict selfish behaviors, attacks, false alarms, and energy consumption to protect the network from malicious external and internal threats.

4 Application of Game Theory in WSN Security

Game theory is a technique used first in the economic field to predict the effect of a phenomenon when different parts are competing for the same resources. This scheme is used to secure opportunistic routing. Nodes are players in a routing game. Benign and malicious nodes are both parts of the game. The mathematical design of the game theory fixes the conflict between players aiming for the same goal. In opportunistic routing scheme, for example, benign nodes tend to find efficient and best path to transfer the packets, whereas malicious nodes tend to drop

messages; the evolution of game theory has settled a new mathematical model that nodes will follow when malicious appear in the network. In this section, we are going to list the existing game theory-based opportunistic routing protocols.

4.1 Aim

Is a game theory based protocol, AIM prevent selfish nodes behaviors in opportunistic routing protocol SOAR (time bases coordination), and make the energy consumption even equal between the nodes [12]. The source node uses a payment technique to make the relays transmit the packet to the destination. The forwarding nodes send the bids.

The authors designed a forwarding auction game. First the source chooses an amount of forwarding nodes, these relays determine the price that a node deserve to get to forward the packet. The aim of the auction game is to settle a pricing plan in order to optimize the transmission. AIM achieves The Bayesian Nash equilibrium solution to maximize the benefit of the nodes. This process requires an important amount of energy the authors did not forget this detail by including it in the auction game process.

4.2 COMO

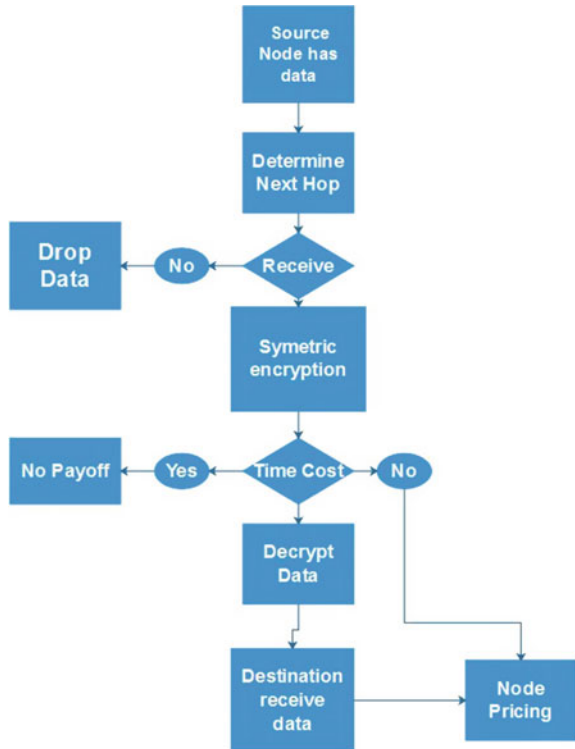
Is a game theory based protocol that stand for cooperation optimal protocol for multi-rate opportunistic routing and forwarding this protocol mainly aim to prevent selfish nodes to ensure fidelity of the players and reach maximized end to end throughput by using Nash equilibrium. When the author nodes obey the protocol, the network is optimized and the nodes get paid; the protocol measures link loss probabilities to include probe message [9]. COMO uses cryptographic elements to defend the probed messages from casting. The payment scheme ensures that nodes do not take advantage from link loss probabilities. COMO main metric is EATT (time metric). In real network, a source node pays the relays for transmitting packets by using the strongly Pareto, the relays cannot expand their utility if the other nodes utility does not decrease. The Authors have experimented COMO on ORBIT wireless scheme and prevent nodes misbehaviors, because nodes report the activity of their neighbors using traditional routing to compute the link loss probabilities. Moreover, when a node is reported, it gets paid and thus prevents selfish relays. Authors provided an extended amount of simulation to show the efficiency to prevent selfish behaviors.

5 Contribution and Discussion

Three main security techniques are separately used to secure WSNs which are: cryptography, trust based methods and game theory. In this section, we propose an approach to integrate cryptography to game theory methods to ensure a high confidentiality and integrity of data in WSNs (Fig. 4).

The following flowchart is a mixing between cryptography features and game theory pricing methods to motivate the nodes to act cooperatively. When the source wants to send a packet to a destination, it uses a protocol to select the next hop; this protocol can be traditional routing protocol or opportunistic; in our work, we aim for opportunistic routing as long as it gives the possibility to adapt the sending to the dynamic nature of the broadcast medium. After the selection, the next hop the data is encrypted using whether symmetric or asymmetric cryptography technique we assume the first one would avoid an important amount of processing because of the small processors capacities of the nodes. When the data is received by the relay a timer evaluates the normal duration of the packet to reach the destination, if it costs more than usual the node is punished and doesn't get any payoff, if not the data is decrypted and reaches the destination. The node is then rewarded.

Fig. 4 Game theory and cryptography combination



This method will surely increase the WSN security requirements:

- Confidentiality: The nodes are capable to communicate and understand each other.
- Integrity: Data should not be intercepted by a malicious sensor node.
- Authentication: Each sensor node uses the authentication information to verify that data comes from the benign node.
- Authorization: Nodes are allowed to perform some tasks like sending and receiving.
- Availability: The sensors must be ready when needed.
- Secrecy: The transit of data in the WSN should be encoded and decoded.

6 Conclusion

This paper has presented a comprehensive survey of the game theory methods and their application to secure wireless sensor networks. We have presented first an overview of the wireless sensor architecture and the security problems and attacks that face them. Then, we gave a clear explanation of the game theory techniques and their definitions. We finally presented our work on integrating game theory to cryptography techniques to enhance the security in WSNs. Which is to the best of our knowledge neglected, since the main focus is on energy saving and the robustness of these types of networks. Many works should be done to ensure a security for WSNs as long as they may operate in hostile environment and under threats. For that, we decided to extend this work by applying a cryptography and game theory combination to make it less energy and processing consuming that would be then our perspective for future projects.

References

1. M. Elappila, S. Chinara, D.R. Parhi, Survivable path routing in WSN for IoT applications. *Pervasive Mob. Comput.* (2017)
2. P. Spachos, L. Song, D. Hatzinakos, Comparison of traditional and opportunistic multihop routing in wireless networking scalability, vol. 1, pp. 182–185 (2012)
3. E. Efficiency, K. Zeng, Opportunistic routing in multihop wireless networks (2008)
4. A.A. Abins, N. Duraipandian, A survey on opportunistic routing protocols in wireless networks. *Proc. Comput. Sci.* **10**(3), 148–153 (2015)
5. Y. Xu, P. Scerri, M. Lewis, K. Sycara, Token-based approach for scalable team coordination
6. E. Rozner, J. Seshadri, Y. Mehta, L. Qiu, SOAR: Simple opportunistic adaptive routing protocol for wireless mesh networks. *IEEE Trans. Mob. Comput.* **8**(12), 1622–1635 (2009)
7. F.A. Aoudia, M. Gautier, O. Berder, OPWUM : opportunistic MAC protocol leveraging wake-up receivers in WSNs, vol. 2016 (2016)
8. G. Hosseinabadi, N. Vaidya, Token-DCF: an opportunistic MAC protocol for wireless networks (2012)

9. F. Wu, K. Gong, T. Zhang, G. Chen, COMO: A game-theoretic approach for joint multirate opportunistic routing and forwarding in non-cooperative wireless networks, vol. 1276, pp. 1–12 (2014)
10. A. Chhabra, V. Vashishth, D.K. Sharma, A game theory based secure model against black hole attacks in opportunistic networks (2017)
11. J. Newsome, D. Song, A. Perrig, The sybil attack in sensor networks: analysis & defenses (2004)
12. K. Zhang, R. Wang, D. Qian, AIM: An auction incentive mechanism in wireless networks with opportunistic routing (2010)

DSP Implementation of OQPSK Baseband Demodulator for Geosynchronous Multichannel TDMA Satellite Receiver



A. Chinna Veeresh, S. Venkata Siva Prasad, S. V. Hari Prasad, M. Soundarakumar and Vipin Tyagi

Abstract Geosynchronous satellite and development of its earth-based receiver has a lot of importance in satellite networks. The geosynchronous time division multiple access (TDMA) narrowband satellite receiver must have the ability to do proper baseband processing including frequency and time synchronization, phase offset correction, and baseband demodulation. This paper discusses the efficient implementation of baseband processing implemented in Texas Instruments (TI) multicore TMS320C6678 DSP for a field-deployable geosynchronous multichannel TDMA satellite receiver. It also discusses a complete discrete domain baseband processing technique for demodulating OQPSK-modulated bursts for a geosynchronous satellite receiver. In the proposed implementation, high multichannel capacity is easily attainable because of the reduced computational time and reduced complexity. Bit error rate (BER) performance exhibited in this implementation matches the requirements for a satellite receiver. The technique proposed in this paper is rather enticing due to the narrowed estimation error limit which is attainable with even a short preamble, especially for OQPSK-modulated bursts with its self-induced inter-symbol interference (ISI).

Keywords Additive white Gaussian noise (AWGN) · Matched filtering · Multicore DSP · OQPSK · Short preamble · Synchronization · Time division multiple access (TDMA) · Viterbi decoding

A. Chinna Veeresh (✉) · S. Venkata Siva Prasad · S. V. Hari Prasad · M. Soundarakumar · V. Tyagi
Centre for Development of Telematics, Electronics City Phase-1,
Hosur Road, Bangalore 560100, India
e-mail: veeresh@cdot.in

S. Venkata Siva Prasad
e-mail: vsp@cdot.in

S. V. Hari Prasad
e-mail: svhari@cdot.in

1 Introduction

Line-of-sight path between the geosynchronous satellite and earth stations ensure an AWGN-only impaired communication channel. Drifts in carrier frequency and phase are largely caused due to the oscillator mismatches between the transmitter and receiver. To attain the desired BER performance at the operating SNR, it is required to remove or compensate for these offsets in synchronization parameters. Estimating these error values is rather more challenging for narrowband case than that of wideband because of the stringent deviation tolerance requirement. As the data rate decreases, the probability of the constellation points crossing over its decision region increases even for small frequency offsets, which in turn implies an increased bit error rate. Hence, attaining a near perfect synchronization is desirable for satisfactory performance at low SNR.

Estimation of the synchronization parameters typically employs known sequences in TDMA systems. These data-aided techniques are widely discussed in [1, 2]. The performance accuracy of these methods has a direct relation to the length of the reference sequences employed with respect to the burst length. Nevertheless, these known patterns are necessarily employed in burst mode transmissions to assist in finding the burst position within the time slot period. Burst position acquisition, which is almost always the prior step in the burst synchronization, requires techniques like double correlation and differential correlation [3] which work well, even with uncompensated carrier frequency offset. Complete baseband synchronization using data-aided method is described in [4]. However, carrier recovery relying solely on short preambles or midambles often gives a very poor estimate of the frequency and phase errors. This motivated the use of non-data-aided carrier synchronization techniques [5, 6] for better results at the cost of increased complexity and computational time.

This paper focuses on attaining high accuracy burst synchronization in the baseband demodulator, with the complexity measure clearly between data-aided and non-data-aided techniques. While the burst alignment within the time slot is pinned down with the aid of reference sequence, carrier frequency, phase offset estimation, and tracking employ a combination of data-aided, non-data-aided, and decision-directed methods to obtain the desired precision without compromising on the computational time and complexity requirements.

The paper is organized as follows. Section 2 gives the system model. Section 3 explains the proposed algorithm comprehensively. Software architecture of the receiver, hardware test setup, and performance results are shown in Sects. 4, 5, and 6, respectively. The conclusion is drawn, and future works are given in Sect. 7.

2 System Model

The transmitted data is assumed to be organized into bursts. The burst consists of a known short preamble of length L_p OQPSK symbols, followed by convolution-encoded data of length L_d OQPSK symbols and guard period of length L_g symbols.

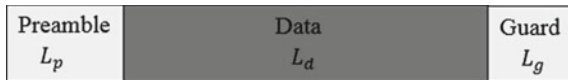


Fig. 1 Burst structure

The burst is pulse shaped by root-raised-cosine (RRC) filter with roll-off factor $\alpha = 0.4$ and transmitted over the AWGN impaired channel after up converting the baseband data. The received signal is represented as (Fig. 1)

$$r(t) = \sqrt{4/T} Re \{ \tilde{s}(t) e^{[j2\pi(f_c + \Delta f)t + \theta_0]} \} + w_1(t) \tag{1}$$

where $1/T$ is the baud rate and $Re\{\cdot\}$ is the real part. Here, f_c is the nominal carrier frequency, Δf is frequency offset which can be positive or negative, θ_0 is the phase offset, $w_1(t)$ is the additive white Gaussian noise with two-sided power spectral density of $N_0/2$ (watts/Hz), and $\tilde{s}(t)$ is the complex envelope of the OQPSK transmitted signal and is given as

$$\tilde{s}(t) = \sum_{k=0}^{L-1} S_{k,I} h(t - kT) + \sum_{k=0}^{L-1} S_{k,Q} h(t - kT - T/2) \tag{2}$$

where $S_{k,I} \in \pm 1$, $S_{k,Q} \in \pm 1$ are the in-phase and quadrature phase components of OQPSK symbol respectively, $h(t)$ is the impulse response of the transmit filter which is assumed to have the root-raised-cosine filter with roll-off factor of 0.4.

The received analog signal is sampled by an analog-to-digital converter (ADC) with sampling rate F_{s1} and down converted to baseband by mixer. It is then passed through low-pass filter and down-sampled at a rate F_s , where $F_s \geq 2(B + \Delta F_{max})$ and $F_s = F_{s1}/D$. ΔF_{max} is the maximum allowable frequency offset in baseband received signal, D is decimation factor, and $2B$ is the passband bandwidth. The received baseband signal is

$$r(nT_s) = \tilde{s}(nT_s) e^{(j2\pi \Delta f n T_s + \theta_0)} + w_1(nT_s) \tag{3}$$

3 Proposed Algorithm Flow

At the receiver, the proposed algorithm flow for baseband processing is depicted in Fig. 2. It performs the tasks of estimating coarse frequency offset using differential correlator branches and fine frequency offset estimation using maximum likelihood (ML) method. Since the burst is OQPSK-modulated, the start of burst of received signal should be accurate before applying to ML method, which can be achieved through

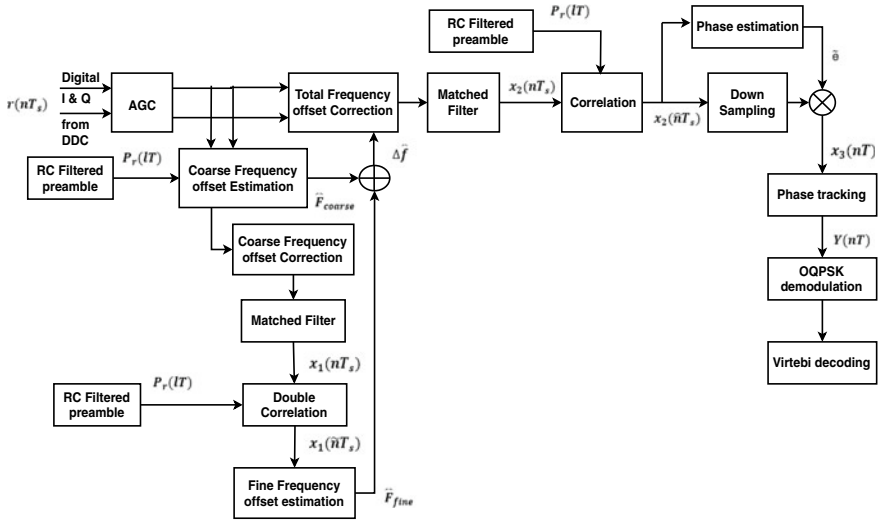


Fig. 2 Proposed algorithm flow

double correlation. The algorithm also performs the estimation and correction of the phase offset in the received samples before soft demodulation and decoding the data. Algorithm flow is described as below.

3.1 Frequency and Time Synchronization

Coarse Frequency Estimation: Coarse frequency offset estimation depicted in Fig. 3 consists of several differential correlator branches, each covering a certain frequency range. In each branch, the input signal is corrected by a different frequency offset. The spectrally shifted signals are differential correlated with the raised-cosine (RC) filtered OQPSK-modulated preamble sequence $[P_r(1T)]$. The RC filtered preamble is precomputed and stored for processing. The maximum search length over which the preamble needs to search depends on the guard period in the time slot within which the burst can shift. Coarse offset estimation is given by the path that gives the maximum value of the so obtained peaks among all the branches. The estimated offset correction is done which brings the frequency deviation within $\pm F_L$ Hz.

In Fig. 3, $F_{step} = 2 * F_L$, $N = (2 * \Delta F_{max}) / F_{step}$, and $\pm \Delta F_{max}$ are the maximum frequency deviation. Maximum value that can be chosen for F_{step} is such that the maximum frequency deviation after passing through the desired branch is less than the baseband filter bandwidth. There are $(N + 1)$ branches in total as shown. Here, matched filter is the receiver RRC filter with $\alpha = 0.4$.

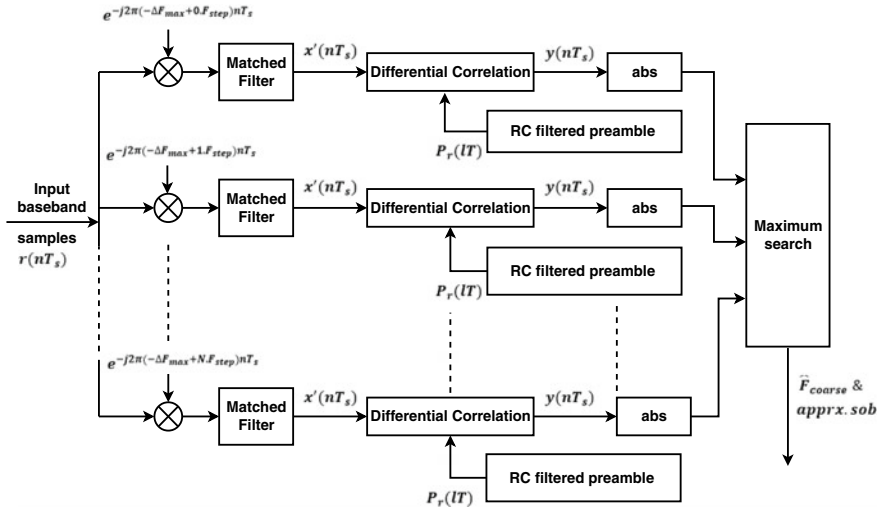


Fig. 3 Coarse frequency offset estimation

Sample corresponding to maximum correlation peak in each branch is not affected by the residual frequency offset with differential correlation as can be seen from the formula

$$y(nT_s) = \sum_{l=0}^{L_p-1} \mu_{n+lM', l}^* \mu_{n+(l+1)M', l+1} \tag{4}$$

$$\mu_{n+lM', l} = x'((n + lM')T_s)P_r^*(lT)$$

$$= x'(nT_s + lT)P_r^*(lT)$$

where $x'(nT_s)$, $n \in I$, is the frequency shifted and matched filtered received baseband samples in each branch. $P_r^*(lT)$ is the conjugate of the l th RC filtered preamble symbol and $M' = T/T_s$ is the baseband sampling factor in each branch.

The branch N' ($N' = 0, 1, \dots, N$), with the maximum peak $|y(nT_s)|^2$, gives the coarse frequency offset as

$$F_{coarse} = -(-\Delta F_{max} + N' F_{step}). \tag{5}$$

The coarse estimation can be brought further close to the actual frequency offset by searching in the range $[F_{coarse} - F_{step} \quad F_{coarse} + F_{step}]$ with a smaller $F_{step} = F_{step2}$ using the same coarse estimation procedure. This can be done iteratively for one or more times to bring it closer to the actual offset. A number of iterations done depend on the computational time required. Residual frequency offset in the burst is estimated and corrected with an ML-based fine frequency estimator, which can

estimate the maximum offsets in the range $[\frac{-1}{2T_s}, \frac{1}{2T_s}]$. The winning branch hence gives the coarse frequency-corrected and filtered samples, $x_1(nT_s)$ which is passed to the next block in the sequence.

Double Correlation for Start of Burst Estimation ($\tilde{n}T_s$): Differential correlation in the coarse frequency estimator block gives an approximate start of burst, *sob1* for the winning branch even in the presence of noise. There could, however, be chances of deviation from the exact start of burst by few samples due to filter introduced ISI for OQPSK-modulated signal. The accuracy in the start of burst estimation plays a major role in finding the fine frequency offset estimation using ML method. Hence, double correlation technique is employed in the algorithm which exhibits an improved performance as can be observed in Figs. 4 and 5.

$$y_1(nT_s) = \sum_{l=1}^{L_p-1} \left\{ \left| \sum_{k=l}^{L_p-1} X_{n+k-l}^* P_{k-l} X_{n+k} P_k^* \right| - \sum_{k=l}^{L_p-1} |X_{n+k-l}| |X_{n+k}| \right\} \quad (6)$$

$$X_{n+k-l} = x_1[nT_s + (k-l)T]$$

$$P_{k-l} = P_r^*[(k-l)T]$$

The first term inside the bracket in (6) is the magnitude of the correlation between $X_{n+k-l}P_{k-l}^*$ and $X_{n+k}P_k^*$. This correlation $\sum_{k=l}^{L_p-1} X_{n+k}P_k^*(X_{n+k-l}P_{k-l}^*)^*$ will be referred to as the double correlation with lag l . The second term inside the bracket in (6) is data correction term. The value of n at which Eq. (6) gives maximum value is \tilde{n} , the exact start of burst of the signal.

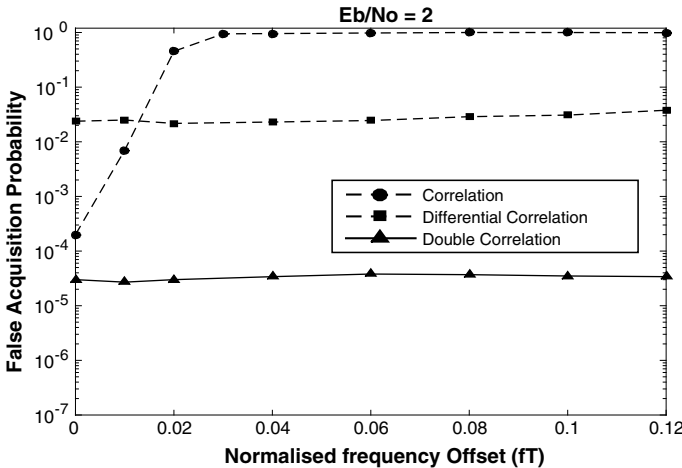


Fig. 4 False acquisition probability versus normalized frequency offset

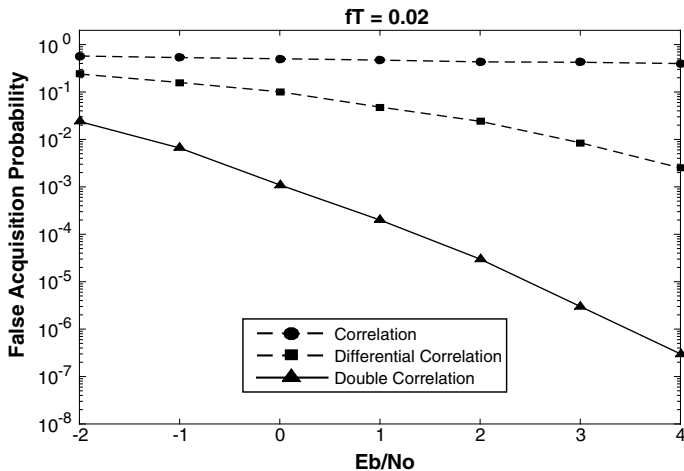


Fig. 5 False acquisition probability versus E_b/N_o

The robustness of the start of burst estimation techniques was examined by estimating the false acquisition probabilities for various normalized frequency offsets (fT) in the range $[0 \ 0.12]$, at $E_b/N_o = 2$ dB. The results are shown in Fig. 4. The performance of correlation-based estimator degraded rapidly as fT increased. The double correlation method outperforms the differential correlation. In the simulation shown in Fig. 5, the behavior of the sob estimators with respect to E_b/N_o was investigated at $fT = 0.02$.

Fine Frequency Offset Estimation: Fine frequency offset estimation is done over the filtered output of the coarse frequency offset-corrected data with start of burst \tilde{n} . Residual fine frequency F_{fine} is searched over $[-F_L \ +F_L]$ using ML method. The computational time can be reduced without compromising the accuracy by performing the fine frequency estimation in two steps:

(i) Approximate fine frequency, F_{fine1} is estimated using data-aided ML approach with the known filtered preamble. This does not yield a very accurate estimate as the preamble length is short; however, error in the estimation will be of the order of a few hertz for a narrowband signal.

Data-aided ML equation for the fine frequency estimation is written as,

set $w_{fine1} = w_i$, if w_i maximizes

$$\left| \sum_{l=0}^{L_p} x_1(\tilde{n}T_s + lT)P_r^*(lT)e^{-jw_i l} \right| \tag{7}$$

with $w_i = -w_L + iw_{fine_step}$

where $w_L = 2\pi F_L/F$ and $i \rightarrow 0$ to $2w_L/w_{fine_step}$, with F being the symbol.

Computational time can be further reduced by choosing a relatively larger $w_{fine_step} = w_{fine_step1}$ to search over $[-w_L + w_L]$ to obtain the DFT peak at some w_{fine1_temp} and then searching over $[(w_{fine1_temp} - w_{fine_step1}) (w_{fine1_temp} + w_{fine_step1})]$ with a smaller $w_{fine_step} = w_{fine_step2}$ to obtain a more precise w_{fine1} .

(ii) Search is performed around w_{fine1} for few tens of hertz using non-data-aided ML estimation technique to obtain a precise fine frequency estimate w_{fine2} . Number of samples, L' used should be adequately large to give a very precise frequency offset estimate. Modulation can be removed from the samples by using a nonlinear method given by

$$z(mT_s) = F[\rho(m)]e^{j4\phi(m)} \quad (8)$$

where $\rho(m) = |x_1(mT_s)|$, $\phi(m) = \arg[x_1(mT_s)]$

$$F[\rho(m)] = \rho^k(m), \quad \text{with } k = 2$$

Non-data-aided ML estimation equations are given as

set $w_{M_{fine2}} = w_i$, if w_i maximizes

$$\left| \sum_{n=0}^{L'-1} z(nT_s)e^{(-jw_in)} \right| \quad (9)$$

with $w_i = 4 * w_{fine1} - w'_L + i * w_{fine2_step}$

where $w'_L = (2\pi F'_L)/F_s$, with $F'_L = 4 * F'$, for F' being a frequency in few tens of hertz and F_s being the sample rate. Here $i \rightarrow 0$ to $(2w'_L)/w_{fine2_step}$, the total number of samples used for ML estimation is denoted by L' and $w_{fine2_step} = (2\pi F_{fine2_step})/F_s$ is the step increment in the angular frequency which defines the frequency resolution of the ML method. The fine frequency estimate is given by

$$w_{fine2} = w_{M_{fine2}}/4 \quad (10)$$

$$F_{fine} = w_{fine2}/(2\pi * T_s)$$

Total frequency offset estimated from the coarse and fine frequency offset estimates is compensated in the received baseband signal. The frequency-corrected burst is then match filtered with the RRC filter to obtain $x_2(nT_s)$. $x_2(nT_s)$ is correlated with the RC filtered preamble $[P_r(IT)]$ to find the peak correlation point which gives the exact start of the burst.

$$C_r(nT_s) = \sum_{l=0}^{L_p-1} x_2(nT_s + lT)P_r(lT) \quad (11)$$

3.2 Phase Offset Estimation and Correction

The value of $n = \hat{n}$, where $|C_r(nT_s)|^2$ represents the maximum peak, is the exact start of burst. At $\hat{n}T_s$, phase offset estimate can be obtained as

$$\tilde{\theta} = \arg[C_r(\hat{n}T_s)] \quad (12)$$

Synchronized data $x_2(\hat{n}T_s)$ is down-sampled to twice the symbol rate, and the estimated phase offset is corrected from it to obtain $x_3(nT)$.

3.3 Phase Tracking

Phase tracking is used to correct the very small uncompensated frequency offset remaining in $x_3(nT)$. Phase tracking technique employed is decision-directed, making use of the OQPSK hard decision-demodulated bits obtained from frequency, phase and time synchronized data $x_3(nT)$ are OQPSK-modulated and RC filtered to generate the estimates $\tilde{\beta}$. Residual phase offset values are obtained at symbol rate. $x_3(nT)$ and $\tilde{\beta}$ are down-sampled and passed to the decision-directed phase tracking loop which operates at symbol rate. L_{ISI} is the length of filter-induced inter-symbol interference and ρ is $0 < \rho < 1$. Optimum performance is observed at $\rho = 0.999$.

for $0 \leq k < L_p - L_{ISI}$

$$\begin{aligned} Z_0(kT) &= x_3(kT)\tilde{\beta}(kT) \\ Z_{avg}(kT) &= \rho Z_{avg}((k-1)T) + (1-\rho)Z_0(kT) \end{aligned} \quad (13)$$

for $-L_{ISI} \leq l \leq L_d - 1$

$$\begin{aligned} Z_0((L_p+l)T) &= x_3((L_p+l)T)\tilde{\beta}((L_p+l)T) \\ Z_{avg}((L_p+l)T) &= \rho Z_{avg}((L_p+l-1)T) + (1-\rho)Z_0((L_p+l)T) \\ \theta((l+L_{ISI})T) &= \arg[Z_{avg}((L_p+l)T)] \end{aligned} \quad (14)$$

The estimated residual phase offsets are corrected from the data portion of x_3 at twice the symbol rate and are given by:

$$\begin{aligned} Y(2kT) &= x_3(2(L_p+k)T)e^{-j2\pi\theta(kT)} \\ Y((2k+1)T) &= x_3(2(L_p+k+1)T)e^{-j2\pi\theta(kT)} \quad \text{for } 0 \leq k < L_d \end{aligned} \quad (15)$$

3.4 Soft and Hard OQPSK Demodulation

Since Q component has an offset by $T/2$ with respect to I component in OQPSK signal, alternate samples of Y are used to extract the ISI free I and Q components which are then mapped to even and odd bits, respectively. For each sample, based on the proximity of the corresponding I or Q component to the constellation point, 3-bit soft decision-demodulated output is obtained. Likewise, hard decision demodulation outputs are also obtained by considering I and Q components from alternate samples.

3.5 Soft Decision Viterbi Decoding

3-bit soft decision demodulator output bits corresponding to the complete data portion of the burst are passed to a soft input 1/2 rate Viterbi decoder to get the transmitted information bits.

4 Software Architecture of Receiver

The input IF signal is sampled by an analog-to-digital converter (ADC) at sampling frequency F_{s1} which is greater than the bandpass sampling frequency. The output from the ADC goes to the multicarrier digital down converters (DDC) which are implemented in the Xilinx's Kintex-7 FPGA using system generator blocks. The FPGA is programmed to have three DDC, each of which in turn handles four different channels. Each DDC consists of a direct digital synthesizer (DDS), a mixer, two cascaded integrator comb (CIC filters), an FIR filter, and scaling blocks. The DDC outputs are buffered alternatively in two RAMs. The block diagram is shown as below in Fig. 6.

For the IF at 70 MHz with 16 MHz bandwidth, the ADC samples at a rate of 40.6 Msp. DDC brings it to baseband with a numerically controlled oscillator (NCO)

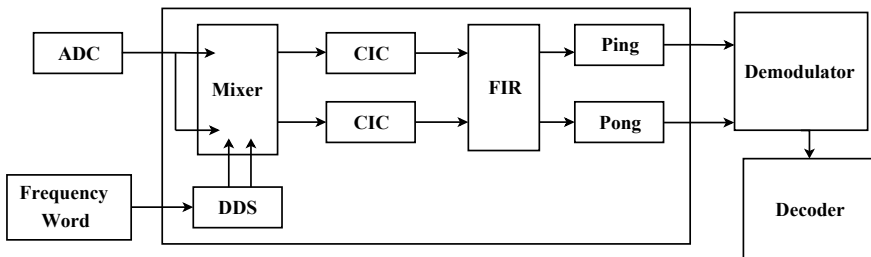


Fig. 6 Software architecture of receiver

Table 1 Parameters of ADC and DDC

ADC o/p rate or DDC i/p rate (Msps)	40.6
CIC decimation	725
FIR decimation	2
DDC clock (MHz)	162.4
FIR filter baseband cutoff (KHz)	12.4
DDC o/p sample rate (Ksps)	28
DDC o/p symbol rate (Ksps)	3.5
DDC o/p samples/symbol	8
Modulation	OQPSK
No. of channels/DDC	4
No. of DDC/FPGA	3

and further reduces the sampling rate to 28 Ksps using CIC and FIR filters for each channel. Hence for a baud rate of 3.5 Ksps, eight I/Q samples are generated per symbol per channel. The I/Q samples of the 12 channels per FPGA are contiguously stored in two alternating ping and pong buffers, such that, while data is fed to DSP from one RAM, the sampled data for the corresponding burst period is stored in the other RAM, with the storing and feeding toggling between the alternating RAMs with each time slot. The input/output parameters are shown in Table 1.

Baseband processing is done in TMS320C6678 TI eight core DSP operating at 1.25 GHz frequency. The 4 MB shared on-chip SRAM provided by multicore shared memory controller (MSMC) can be accessible by all the eight cores. As shown in Fig. 7, DDC outputs are fed to the DSP using external memory interface (EMIF16). The DSP is programmed to handle I/Q samples of 12 channels. These samples are stored in contiguous buffers in multicore shared memory (MSM SRAM), which is shared across all the cores. Core 0 functionality is dedicated to the reception and shared memory buffering of the DDC output samples. Baseband synchronization and demodulation algorithms run on core 1, core 2, core 3, and core 4 which are programmed to handle three channels per time slot. The demodulated soft decision outputs from these cores are fed into the Viterbi decoders programmed in Virtex FPGA. The Virtex FPGA has Viterbi decoder to handle the demodulated data of the 12 channels. Interfacing between DSP and Virtex FPGA again uses EMIF16. Core 5 receives the decoded data which is passed to the Core 6 for IP packetization and interfacing to the external world. The synchronization algorithm is implemented using fixed-point C and optimized using TI intrinsic operators and other compiler techniques in code composer studio (CCS), an integrated development environment (IDE) from TI. DSP can very well be programmed to increase the number of channels per DSP. The use of C intrinsics reduces the overall computation time of the demodulation algorithm. Complex multiplications and additions are done effectively using intrinsics that operate on 64-bit numbers, thus reducing the time taken. DSP uses

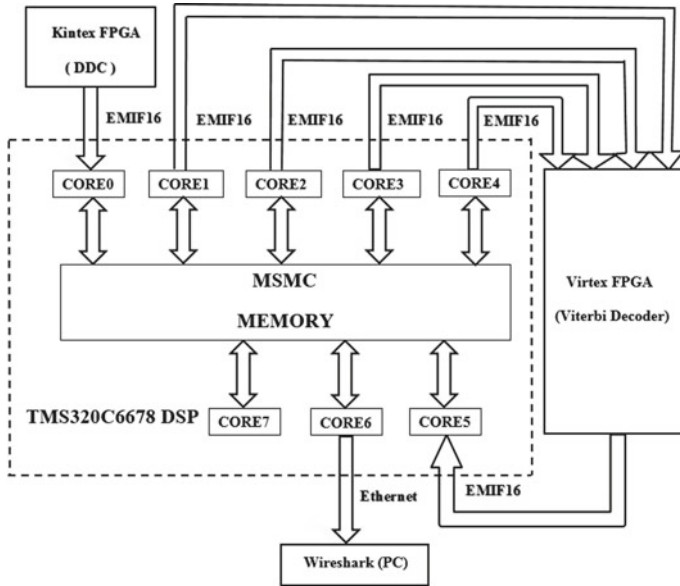


Fig. 7 Baseband processing of receiver in multicore DSP

seven cores for burst reception, synchronization, demodulation, and finally packetization. The total processing time for every burst is 35.125M clock cycles (1 clock cycle = 1/1.25 ns). So DSP can support 12 channels.

5 Hardware Test Setup

Hardware test setup is depicted in Fig.8. The IF signal carrying the OQPSK-modulated burst is generated from E4438C ESG vector signal generator. This IF is given to the programmable noise generator which adds the required AWGN. The IF output of the noise generator is passed on to the device under test (DUT). DUT has the IF card where ADC is implemented, the Kintex-7 FPGA which performs DDC functionality, demodulator in DSP, and decoder in Virtex FPGA.



Fig. 8 Hardware test setup

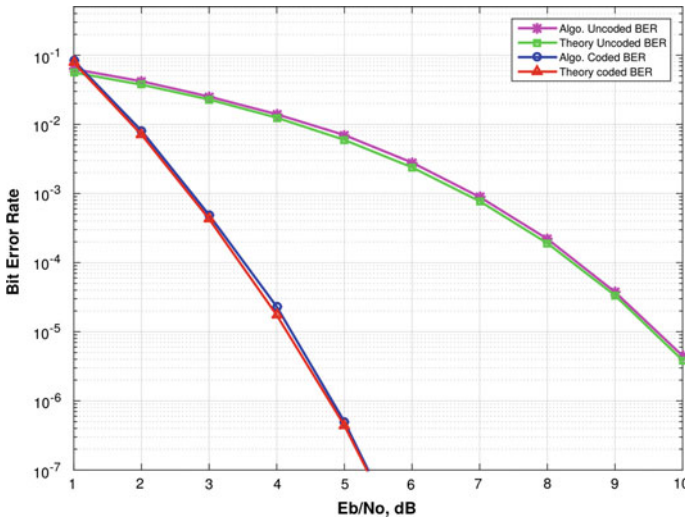


Fig. 9 BER performance versus E_b/N_0

6 Performance Results

E4438C is programmed to generate IF signal with a carrier frequency of 70 MHz and offset within ± 7.5 kHz, carrying OQPSK-modulated encoded bursts with 7 kbps bit rate. Each time slot is 120 ms duration, where transmission time is 110 ms and guard time is 10 ms. A short preamble of length $L_p = 32$ symbols is used in the burst. Pulse shaping filter set at the baseband is RRC filter with $\alpha = 0.4$. BER is obtained by varying noise in the noise generator for E_b/N_o in the range from 0 to 10 dB for different nominal frequency offsets.

Coded and uncoded BER obtained is plotted in Fig. 9. Slight degradation from the ideal as can be seen from the figure is caused by ADC quantization error. Demodulator algorithm in the DSP cores takes 28.1 ms for its execution which allows for up to 3 channels to be processed per core per DSP with all the interfacing delays. Hence, it provides a high multichannel capacity even with a single DSP.

7 Conclusion and Future Work

This paper discusses a complete discrete domain baseband processing technique for demodulating and decoding OQPSK-modulated bursts for a geosynchronous satellite receiver. It also touches up on the IF processing section involved. The algorithm is able to attain a very close synchronization of the received burst with the transmitter. Baseband synchronization and demodulation are completely implemented in C

language to run on TMS320C6678 TI multicore DSP. BER performance exhibited is appreciable for a satellite receiver. High multichannel capacity is easily attainable because of the reduced computational time and complexity involved. Future work is in the direction of receiver design for variable data rates.

References

1. U. Mengali, M. Morelli, Data-aided frequency estimation for burst digital transmission. *IEEE Trans. Commun.* **45**(1), 23–25 (1997)
2. S.M. Kay, A fast and accurate single frequency estimator. *IEEE Trans. Acoust. Speech Signal Process.* **37**, 1987–1990 (1989)
3. Z.Y. Choi, Y.H. Lee, Frame synchronization in the presence of frequency offset. *IEEE Trans. Commun.* **50**(7) (2002)
4. K. Vasudevan, Synchronization of bursty offset QPSK signals in the presence of frequency offset and noise, in *Proceedings, IEEE TENCON*, Hyderabad, India, Nov 2008
5. A.A. D’Amico, A.N. D’Andrea, R. Reggiannini, Efficient non-data-aided carrier and clock recovery for satellite DVB at very low signal-to-noise ratios. *IEEE J. Select. Areas Commun.* **19**(12), 2320–2330 (2001)
6. M.K. Nezami, R. Sudhakar, H. Helmken, DFT-based frequency acquisition algorithm for large carrier offsets in mobile satellite receivers. *IEE Electron. Lett.* (2001)

Textile Sensor-Based Exoskeleton Suits for the Disabled



P. R. Sriram, M. Muthu Manikandan, Nithin Ayyappaa and Rahul Murali

Abstract The exoskeleton suit is designed to make a person who is completely or partially paralyzed below the waistline capable of walking, running with minimal effort. This suit is mainly focused on people suffering from paraplegia. The special ability of the following suit is that it does not make the person bulky rather it retains the natural formation and structure of a human. The following suit is based on textile-based sensors that determine the movement and motion of the forearms. The advanced sensors placed in the forearm calculate the pattern and swinging motion of the forearm and formulates the action. The calculated motion of the forearm provides a gesture-based input for the control system. Thus, it provides a quicker response time for the manipulators to act according to the scenario. The suit consists of lithium-ion battery inside a backpack which powers the manipulators. A control system is further initiated to the actuators placed in thigh, knee, and ankle of the leg. The cutting-edge technology will be able to improve the muscle movement in the legs and help to permanently resolve the issue instead of relying on the use of wheelchairs for the rest of their lives.

Keywords Computational geometry · Graph theory · Hamilton cycles

P. R. Sriram (✉) · M. Muthu Manikandan · N. Ayyappaa · R. Murali
Department of Electronics and Communication Engineering, Sri Venkateswara
College of Engineering, Sriperumbudur, Tamil Nadu, India
e-mail: sriram.pr97@gmail.com

M. Muthu Manikandan
e-mail: muthu.manikandan.m@gmail.com

N. Ayyappaa
e-mail: ayyappannithin006@gmail.com

R. Murali
e-mail: luharrahu909@gmail.com

1 Introduction

Paraplegia is an impairment in motor or sensory functions of the lower extremities. It is usually caused by spinal cord injuries or a congenital condition that affects the neural elements of the spinal canal. The American spinal injury association (ASIA) classifies the above condition to be extremely severe. This suit is designed to help the disabled people capable of performing their everyday activities in ease. The suit is designed using ultra-light aluminum alloys which provide maximum ease to the user. Self-balancing technologies have been implemented to achieve stability. There are textile-based sensors which are present in the sleeve of the shirt which the user wears on. These sensors are the new technology that we have implemented to the already existing exoskeleton technology. These sensors calculate the angle of the arm bending and process the required criteria to provide the required speed for the suit to perform with. It boasts of high flexibility and rigidity. The suit consists of a lithium-ion battery producing power for the control system. The backpack also comprises of the gearbox which regulates the speed in which the suit operates. This is further connected to the manipulators, actuators, and the sensors in the arm. The objective of the suit is to provide maximum efficiency and comfort to the user.

2 Construction

2.1 *Limb Support*

This is categorized as the functioning region. Aluminum alloy-based limb support is designed using CAD software and manufactured. This limb support consists of three parts; they are present in the thigh region, limb region, and the ankle region and are named L1, L2, L3, respectively. These are further connected with two pneumatic drive system actuators situated in the hip-thigh joint, limb joint, and ankle joint named A, B, C, and D, respectively. This followed by a gearbox connected with dc motors. The following is similar in both the legs.

2.2 *Backpack*

This is categorized as the powerhouse and the processing unit of the suit. It consists of a lithium-ion battery powering the TM4C1294NCPDT controller inside the backpack. It also comprises of a three geared gearbox. This ensures three various speed for movement—walking, fast walking, and running. The backpack is designed in a sleek and slim way to ensure that the suit does not become bulky. The battery provides sufficient power for the user to cover a maximum distance of 13,000–14,000 steps while walking and 4000–8000 steps while running, depending on the speed.

The average person walks nearly 7000–12,000 steps daily. Hence, the following suit is sufficient enough for the user to carry out his day-to-day actions.

2.3 Backpack

A textile pressure sensor for muscle activity and motion detection are equipped in the forearm of the disabled person to track the dynamic movement of the person. Example: running, walking, and climbing. A force pressure sensor is also equipped in the bottom of the heel in the suit to detect if the movement is continuous. A dedicated switch is found in the suit which helps the user to switch between walking and staying still. This can prevent unwanted movement while performing normal activities.

3 Working

3.1 Textile Sensors

The sensors used in this research finding are based on a tri-layer structure with a capacitive force among them along with a pressure detecting neutral dielectric layer. The purpose of the sensors is to detect the muscular movements of the higher limbs. On shrinkage of the muscles, it broadens and this in turn promotes the pressure increment on the muscle. This principle of the sensor allows the suit to determine the required motion of the person. Whether to walk, run, or climb stairs. Not only the motion of the arm can be detected, but also the activity of the biceps and triceps can be measured accurately.

These accurate readings help us to figure out the exact angle in which the forearm is swung during a movement. In brief description, we understand that when the human tends to walk the forearm is swung at an angle between 180° and 90° . Similarly, during running, the forearm is swung at an angle between 90° and 270° . From the above context, we discover the dynamic motion of the forearm differentiating walking and running. The sensors located in the forearm identify such change in the angle and sends the information to the microcontroller system. Another interesting notion on human body motion is that, the human hand is first initiated as a gesture and the opposite leg is swung in parallel, completing a movement (Figs. 1 and 2).

So basically the movement of the suit basically depends on three major aspects.

- The angle of the forearm swinging movement
- The hand which swings
- Activity of biceps and triceps.

This information is fed into the control system.

Fig. 1 Walking and running

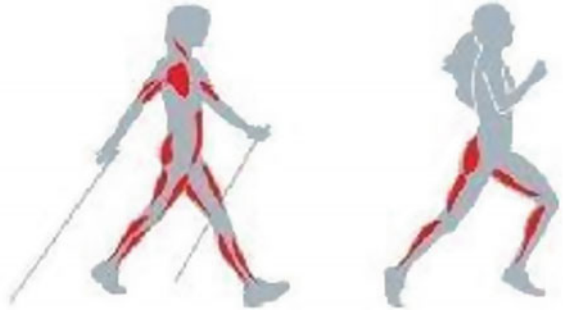
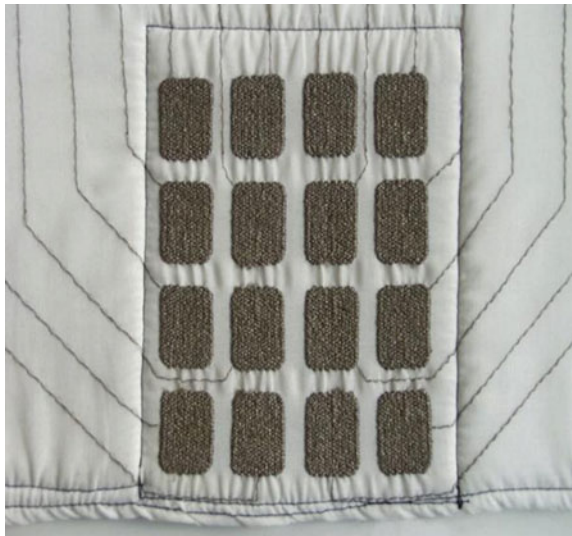


Fig. 2 Positioning of the pressure sensing elements consolidated under fabric

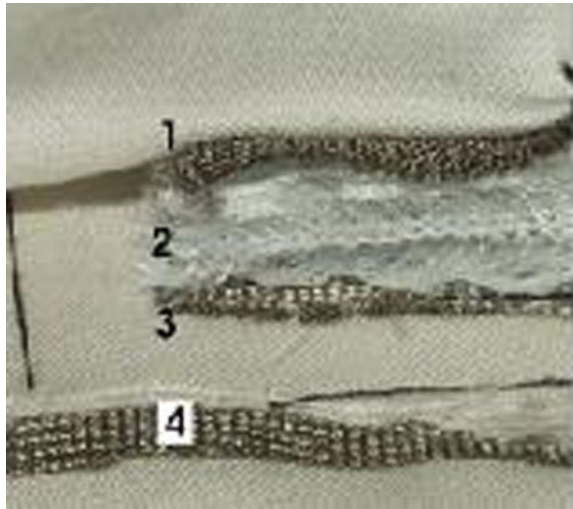


4 Control System

Control system is the place where all the analytical operations take place and all the operations to be carried out are calculated here. This control system is embedded into the backpack of the suit and along with the gearbox and the battery. TM4C1294NCPDT is the primary microcontroller used in the following suit. This is the main control unit of the exoskeleton. This is the main functioning unit of the suit. The control system is present in the backpack along with the battery and the gearbox unit. The control system facilitates total control of the suit by regulating the speed of the suit by feeding the output to the gearbox. This is the brain of the suit. All functions are carried out in this block. Self balancing of the suit is achieved using stabilizing mechanism using the control system (Fig. 3).

The left region of the frontal lobe commands the right side of the human hands and legs, whereas the right region of the frontal lobe commands the left side of the

Fig. 3 Structural alignment of the integral sensing element. Conductive yarn on 1 and 3, high fiber spacing compressible layer



human hands and legs. On the course of walking a systematic sync of the limbs are observed. When right side of the lower limb is put forward the left upper limb moves in forward and vice versa. These postulates are strictly followed on the design and construction of the exoskeletal wearable. The image portrays the command path.

In the figure, is the body of person who is paralyzed from the waistline because he had an injury. So, to deal with this problem a textile sensor is fixed rigidly on both the arms. When left-hand swings in front, the angle will decrease and reach a threshold point. This threshold point is that angle at which the arm swing angle is minimum. When a person walks his arms move in a specific manner. The textile sensors are used to calculate the angle of the arms and send the information to the computing system. The sensor sends these readings to the microcontroller. Immediately the microcontroller sends signals to the solenoid-operated DCV (directional control valve). While it is decreasing its angle, there is a DC motor which is fixed to the flow-control valve is revolving, and thereby the stroke length of the piston cylinder is increasing. The speed of the exoskeleton depends on the angle detected by the textile sensors and as a result the control system and generates the output required which triggers the gearbox and increases or decreases the speed (Fig. 4).

There are two pneumatic actuators one at the back-thigh muscle and one at the calf muscle. The back-thigh muscle actuator actuates gradually as soon as it gets information from the sensors. These sensors ensure free owing movement of the user and do not cause any disruptions for the user. These sensors also ensure comfort for the user.

The encoder is deployed to summarize the rotations undergone by the direct current (DC) powered motor and relay it to the microcontroller on board to direct the motor to rotate in the contrary directions with the identical number of rotations and the two actuators are pulled back every time the feet is in contact with the surface

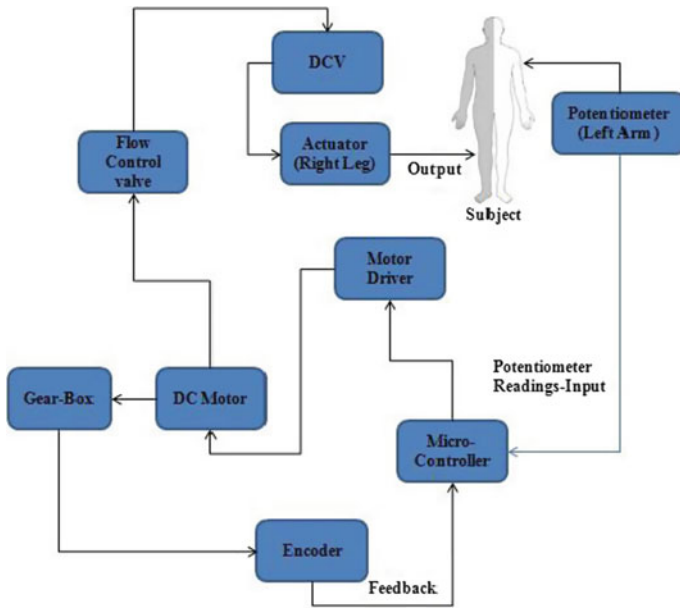


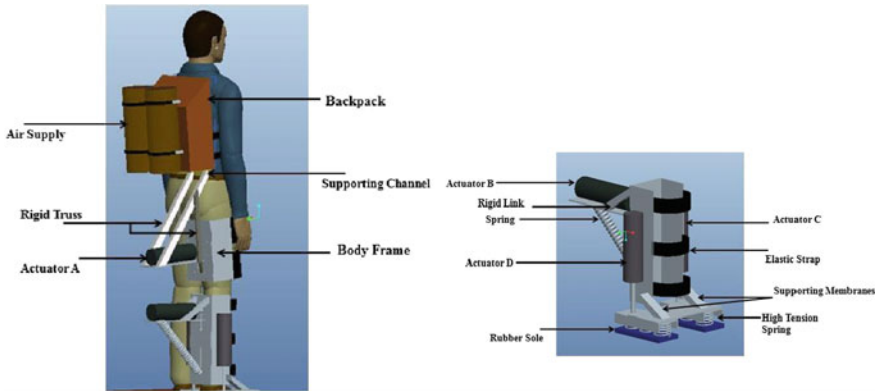
Fig. 4 Active control system

using a force detecting sensor, which is placed on the exoskeleton suits bottom, where the force is sensed.

5 Leg and Thigh Assembly

5.1 Leg Assembly

The lower functioning infrastructure comprises of three pneumatic actuators B, C, and D, respectively. When actuator B extends, the upper body mainframe refuses to coordinately move with the nether body since it is attached to the robust overlying body frame which is bridged to the substructure of the foundation of the cylinder and predominantly due to the adverse tension spring. Eventually, an oscillating movement which is habitual in an individual gait cycle. Actuators C and D are always in an extended state to assist the ruptured leg derived from the response forces. This facilitates the linear motion of the suit in almost any unfavorable terrain. Furthermore, an adaptable strap which assists in providing a composed rigid bridge connecting the main upper frame and the supporting lower structure which is variable.



5.2 Thigh Assembly

The actuator A is coordinated through a pseudo-motion that the augmented stroke of cylindrical piston achieves paramount angles, which is susceptible for movement. Immense vibrations are produced in accordance with the pneumatic pressure. Principle reinforcement is provided for the safety of the structure. The actuating movements are enforced to the rigid structure of the metallic extension. Thus, the expansion/nullification will induce actuating movement of the thigh tissue and this in turn leads to providing mobility. The metallic frame structure consists of high-grade aluminum alloys for high rigidity and low mass. Therefore, there is null forward transference of potential to the thigh tissue. The underface of metallic structure is laminated with occult material providing gushing inner fabrication. Elastic belts are aligned with the motive to adapt to the metallic structure with utmost ease along with adjustable capabilities with respect to the subject. The welding of robust links with metallic frame structure provides utmost rigidity.

6 Conclusion

Thus, the suggested methodology can assist the humans suffering from paraplegia. Paralysis is an ideology widely described in the paper. The idea of a textile sensor in a suit is an unexplored and opens new arenas for research development. This ideology being in its sapling stage currently, a huge evolution can be expected and that would aid the biomedical engineering to revolutionize med-tech to aid non-motile paralyzed humans to move around by making them walk.

References

1. L. Gui, Z. Yang, X. Yang, W. Gu, Y. Zhang, Design and control technique research of exoskeleton suit, in *Proceedings of the IEEE International Conference on Automation and Logistics*, Jinan, China, 18–21 August 2007
2. <http://emedicine.medscape.com/article/320160-overview>
3. <http://www.chiro.org/ACAPress/BodyAlignment.html>
4. Varicap series datasheet from MACOM
5. J. Myer, P. Lukowicz, G. Troster, Textile Pressure Sensor for Muscle Activity and Motion Detection, Austria
6. P. Van Torre, P. Vanverdeghem, H. Rogier, Correlated shadowing and fading characterization of MIMO off-body channels by means of multiple autonomous on-body nodes, in *Proceedings of the 2014 8th European Conference on Antennas and Propagation (EUCAP)*, Hague, The Netherlands, 6–11 April 2017
7. C.J. Walsh, D. Paluska, K. Pasch et al., Development of a lightweight, underactuated exoskeleton for load-carrying augmentation, in *IEEE International Conference on Robotics and Automation* (IEEE, 2006), pp. 3485–3491
8. M.S. Cherry, D.J. Choi, K.J. Deng et al., Design and fabrication of an elastic knee orthosis—preliminary results, in *ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (American Society of Mechanical Engineers, 2006), pp. 565–573
9. E. Ambrosini et al., The combined action of a passive exoskeleton and an EMG-controlled neuroprosthesis for upper limb stroke rehabilitation: first results of the RETRAINER project, in *IEEE International Conference on Rehabilitation Robotics*, 2017, pp. 56–61
10. J. Kamer, W. Reichenfelser, M. Gfoehler, Kinematic and kinetic analysis of human motion as design input for an upper extremity bracing system, in *Proceedings of the IASTED International Conference Biomedical Engineering*, 2012, pp. 376–383
11. T. Zahari, A.P.P. Abdul Majeed, M.Y. Wong, P. Tze, A. Ghaffar, A. Rahman, Preliminary investigation on the development of a lower extremity exoskeleton for gait rehabilitation: a clinical consideration. *J. Med. Bioeng.* **4**(1), 1–6 (2015)
12. A. Agrawal, A. Dube, D. Kansara, S. Shah, S. Sheth, Exoskeleton: the friend of mankind in context of rehabilitation and enhancement. *Indian J. Sci. Technol.* **9**(S1) (2016). ISSN 0974-5645
13. A.K. Seo, J. Lee, Y. Lee, T. Ha, Y. Shim, Fully autonomous hip exoskeleton saves metabolic cost of walking, in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4628–4635
14. E. Sejdic, K. Lowry, J. Bellanca, M. Redfern, J. Brach, A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**, 603–612 (2014)
15. D. Jarchi, C. Wong, R. Kwasnicki, B. Heller, G. Tew, G.Z. Yang, Gait parameter estimation from a miniaturized ear-worn sensor using singular spectrum analysis and longest common subsequence. *IEEE Trans. Biomed. Eng.* **61**, 1261–1273 (2014)

A Comparison of Indoor Positioning Systems for Access Control Using Virtual Perimeters



Brian Greaves , Marijke Coetzee  and Wai Sze Leung 

Abstract Integrated smart technologies are fast becoming the norm in modern work and home environments for providing interactivity and ease of use. Greater interconnectivity, however, enables greater risk of misuse. Logical assets in such environments are protected by logical access control. However, if a logical asset is given a physical form, it no longer has the same protection due to logical and physical access control not being well integrated into physical spaces. Great strides have been made to protect assets in physical spaces by geographically placing a security perimeter around them. Geo-fencing enables the demarcation of a virtual perimeter around locations to protect them from unwarranted access. A limitation of geo-fencing is that location cannot be determined accurately indoors as positioning technologies such as GPS are ineffective, and tag or active positioning systems are easily subverted. This research explores indoor positioning systems to define virtual perimeters in indoor spaces for access control to be performed even when topological changes may occur.

Keywords Indoor positioning systems · Virtual perimeters · Access control

1 Introduction

Physical spaces such as office buildings are becoming more integrated with smart technologies to enhance the level of interaction between users and building services [1]. Security services for such environments need to consider the protection of the physical and cyber assets they contain. It is not sufficient to prevent access to a room containing assets, but to limit what can be done while within the room itself [2].

B. Greaves (✉) · M. Coetzee · W. S. Leung
University of Johannesburg, Auckland Park 2006, South Africa
e-mail: bgreaves@uj.ac.za

M. Coetzee
e-mail: Marijkec@uj.ac.za

W. S. Leung
e-mail: wsleung@uj.ac.za

In a laboratory, one may be able to use some equipment but not others if they do not have sufficient clearance. If the restricted equipment is moved, it should remain restricted. Thus, topological changes must be considered when designing security [2].

Location-based access control is an active field of research that uses location information in access decisions with sufficient granularity to suit the needs of the space [4]. Protection can be provided by creating a virtual perimeter around devices to perform access control when entities are in sufficient proximity [3]. Virtual perimeters exist in the form of outdoor GPS or 3G geo-fences [4] which provide a boundary to trigger actions when crossed. However, due to the inaccuracy of GPS and 3G indoors, such systems would not be effective in an office or laboratory environment [5].

When designing an indoor equivalent of a geo-fence it is important that a system is accurate, unobtrusive, and cost-effective. Currently, beacons, tags, and Wi-Fi positioning systems are popular choices in systems with low-security requirements [6].

This research makes a contribution by comparing various forms of indoor positioning systems to create the equivalent of an indoor geo-fence system, a virtual perimeter.

The following section presents a scenario as a basis for this research and followed in Sect. 3 by a discussion on access control and indoor virtual perimeters. Section 4 presents a set of access control requirements for the scenario and Sect. 5 discusses indoor positioning systems that could potentially satisfy them. Finally, Sect. 6 compares the indoor positioning systems and Sect. 7 concludes the research.

2 Motivating Example

An office building is a physical space with offices, meeting, and printer rooms. The topology of the building determines which spaces are reachable with many offices being open plan. Some offices may be restricted and if an employee is granted access to an office via an access token, he has satisfied its reachability requirement.

Within the offices, employees have laptops which store sensitive documents. These laptops are physical entities, but should someone interact with them, there could be repercussions in the physical or cyber realms. For example, Alice is a finance officer with a laptop storing sensitive documents. Logical access control prevents others from viewing documents on her laptop, but nothing is stopping a colleague from snooping over her shoulder as she works. Thus, logical access control fails to protect the cyber document in the physical realm. Should a clerk of lower clearance be granted access to Alice's office to consult with her, the documents should not be viewable to him.

Furthermore, if Alice goes to another room, any sensitive information on her screen should still be protected. Similarly, if she prints a document in the printer room, it should not be readable by those without sufficient clearance. Since other

employees may already be in the office, access control cannot be done at the door. Therefore, both physical and logical access control must be used in unison to protect cyber assets. A perimeter can thus be created around Alice's laptop to prevent viewing if a clerk is too close. Therefore, determining the accurate location of all entities is important.

To address this consideration, the next section discusses indoor access control.

3 Access Control for Indoor Physical Spaces

Access control is the security service that limits access to resources and services for legitimate users of a system [7]. Access control falls into either the *physical* or *logical* domains. This section discusses physical and logical access control and the use of location therein. Access control and virtual perimeters are then defined.

3.1 *Physical and Logical Access Control*

Physical access control is defined as the ability to permit or deny access to a physical space based on an entity's identity and permissions [7], often using doors and locks.

Logical access control limits access to cyber assets such as files or services based on rules or permissions that allow subjects to perform actions on objects [7]. Additional factors such as roles or context can provide greater granularity for access decisions.

The scenario illustrated that both physical and logical access control needs to be performed and cannot, therefore, be treated as individual domains [8]. Thus, the focus shifts inward from perimeter defense to protect objects inside the perimeter. To address this, the system must know where entities are to prevent acts such as snooping.

3.2 *Location and Context for Access Control*

Due to being in a shared office space, it is not sufficient to know which room two entities are in, as in common building access systems [9]. The system must know where they are relative to one another [10]. Should the screen not be viewable by the clerk, work can continue normally. If the clerk moves to view the screen, snooping should be prevented even if the laptop is moved by Alice. Some objects such as printers may not move so anyone in a given area should not be able to view printed documents [11]. Thus, access control should use absolute and relative location interchangeably.

In the scenario, the locations of the employees and the access levels that they have provided the context for the space. Access control enforcement is affected by

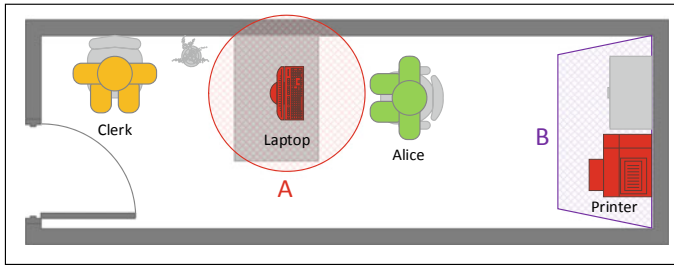


Fig. 1 Conceptual examples of virtual perimeters

the context of the space [7]. Thus, it is imperative to know the location and clearance of entities.

3.3 *Virtual Perimeters*

In a prison, only if the prisoner escapes by crossing the physical perimeter is there a need for action. A physical perimeter can be translated into a virtual counterpart by instead considering the area where a screen is viewable. If the clerk is out of viewing range, then no action is required until he crosses the virtual perimeter to view the screen. Therefore, a perimeter can be established around the screen and then actions can be taken when the distance between the clerk and screen is low enough. Virtual perimeters, therefore, can be expressed using absolute coordinates [4] or relative distances.

A and B in Fig. 1 show a relative and absolute virtual perimeter around Alice's laptop and the printer, respectively. As the laptop can move but the printer does not, the access control system should be able to use relative distance from the laptop and absolute coordinates dependent on the type of object and the nature of the space in question. Therefore, the following section presents a set of requirements to address the security needs of the scenario using virtual perimeters for access control.

4 Access Control Requirements for Indoor Virtual Perimeters

Based on the previous work using the same scenario [12], the researchers have identified a set of requirements for indoor virtual perimeters to be used in access control. Here, follow the three most pertinent requirements presented as a basis for later comparison:

1. *Virtual perimeters need to be enforced:* For access control decisions, the system must prevent the unwarranted viewing of sensitive information using subject and object location. If within viewing range, the information should be hidden [10].
2. *Sensors should accurately identify and locate physical and cyber entities:* Sensors must be able to locate subjects and objects accurately enough to determine if a subject has crossed a virtual perimeter so that access control can be enforced while still being usable and cost-effective in an office work environment [13].
3. *Logical access control should be enforced in a physical space:* Access control must protect physical representations of logical objects so that they have the same level of protection as their cyber counterparts [14].

These requirements bring into question the capabilities of indoor positioning systems. Physical entities must be tightly coupled with their logical counterparts so that subjects do not view screens while being reported as outside of viewing range by sensors [8]. To address this, the next section discusses current indoor positioning systems.

5 Indoor Positioning Systems

Indoor positioning systems (IPSs) determine the location of entities in buildings. In Fig. 1, there is a need to ensure that the clerk is not able to view the information on Alice's screen. Inaccurate location information could result in false positives or negatives, resulting in incorrect action being taken. Thus, location accuracy is important.

There are four primary categories of IPSs [10]; *direct sensing*, *dead-reckoning*, *triangulation*, and *pattern recognition*. Each of which are now briefly discussed:

Direct sensing uses sensors installed at the position of location measurement [10, 15]. For example, *barcodes* are scanned at fixed locations to provide information from a location database [10]. *Infrared* systems [10] broadcast an infrared beam that imparts location information to a reader when passed through. *Ultrasound* systems [10] use short-range transmitters to localize subjects that pass under them. Passive *RFID* tags are unpowered and inexpensive, while active tags are expensive and maintenance-heavy but yield accuracies of mere centimeters (passive) to 3 m (active) [10]. Except for infrared and active RFID, these approaches benefit from being inexpensive, simple, and reliable but are inflexible as they have fixed installation locations.

Dead-reckoning is a positioning system that estimates a subject's location based on a previous estimate or known location. Devices use embedded sensors to infer position based on the previous known locations. *Sensor-based dead-reckoning* systems use accelerometers or magnetometers with an accuracy of roughly 1.62 m dependent on the sensor [16]. *Bluetooth low-energy beacons* [16] are accurate to 1.7 m but require the installation of many beacons. *Camera-augmented dead-reckoning* systems [17] use camera images to compare against a pre-mapped building with a

sensor-dependent accuracy of 5–30 cm. Dead-reckoning systems have a great deal of variance in accuracy but benefit from low setup costs. However, if the measurement device is removed, it becomes impossible to position the subject as the system is tracking the device [10].

Triangulation compares location readings from three known points to localize subjects. *RFID* triangulation systems are low cost and accurate to 72 cm (active) and 15 cm (passive) [18] with a high density of sensors. *Infrared* triangulation boasts an accuracy of 20–30 mm [19] but is a costly investment [15]. *Ultrasound* triangulation has an accuracy of 5 cm [15] but is expensive and inflexible. *Wi-Fi* triangulation produces errors of up to 2 m but can be implemented using existing hardware [6]. However, all the IPSs in this category cannot locate subjects if the sensor device is removed [13].

Pattern recognition systems are either computer vision-based or fingerprinting-based. *Monocular computer vision* systems use a single camera to determine the presence of subjects in line-of-sight [14]. *Stereoscopic systems* use calibrated dual cameras to determine depth with an accuracy of 10 cm [15]. *Computer vision* approaches, however, tend to suffer from problems with occlusion and poor lighting [15] but approaches have been implemented to address these problems [20]. Both approaches are inexpensive using off-the-shelf cameras but require enormous computing power to process images. *Signal distribution* or *fingerprinting* [10] relies on a pre-recorded map of data to compare sensory readings to. *Magnetic field measurement* (MFM) [13] enables devices to locate themselves by finding landmarks in a building's magnetic field map with an accuracy of 2.9 m but requires an inexpensive MFM device to be carried at all times.

It is important to note that the methods and systems presented above are not exhaustive, but rather serve to provide a general basis for comparison in the next section.

6 Comparison of Indoor Positioning Systems

Table 1 provides a condensed version of the information presented above. Each IPS is listed in its *category* with columns for *location type*, stating if it uses absolute or relative location and its expected *accuracy*. *Sensor location* indicates if it requires fixed or carried sensors. *#Sensors* lists the number of sensors required and the expected *cost*.

Section 4 stated the three access control requirements derived from the scenario, each of which are now discussed with respect to the IPSs above.

The first requirement of *virtual perimeters needs to be enforced* and requires the specification of virtual perimeters in absolute coordinates or relative distance using techniques such as floatable geo-fences [4]. Accurate sensors are then required to determine if the virtual perimeter is crossed. However, token-based systems are not ideal as the token can be removed and direct sensing does not lend to changes in topology [13].

Table 1 Comparison of indoor positioning systems

Category	Type	Loc. type	Accuracy	Sens. loc.	# Sensors	Cost
Direct sensing	RFID	A	<3 m	Fixed + subject	1 Receiver/location 1 Tag/subject	Low
	Barcode	A	Exact	Fixed	1 Barcode/location 1 Reader/subject	Low
	Infrared	A	Line-of-sight	Fixed	1 Transmitter/location 1 Receiver/subject	High
	Ultrasound	A	Line-of-sight	Fixed + subject	1+ Transmitter/location 1+ Receiver/subject	Low
Dead-reckoning	Phone sensing	A + R	~1.7 m	Fixed	1 Sensor/initial location 1 Smart device/subject	Low to med
	Camera sensing	A	5–30 cm	Fixed	1 Smart device/subject	Low
Triangulation	RFID	R	15–72 cm	Fixed + subject	1+ Receiver/location 1 Tag/subject	Low
	Infrared	R	20–30 mm	Fixed + subject	1+ Receiver/location 1 Tag/subject	V. High
	Ultrasound	R	~5 cm	Fixed + subject	1+ Receiver/location 1 Tag/subject	High
	Wi-Fi	R	<2 m	Fixed + subject	1+ AP/location 1 Smart device/subject	Low
Pattern recognition	Mono camera	R	Line-of-sight	Fixed	1 Camera/location	Low
	Stereo camera	R	~10 cm	Fixed	1+ Camera/room	Low
	MFM	A	~2.9 m	Device + subject	1 Smart device/subject	Low

The second requirement of *sensors should accurately identify and locate physical and cyber entities*, requires an IPS to accurately position subjects to prevent unwarranted viewing of sensitive information. An exact measurement of location is ideal as greater error measurements may result in false positives or negatives [11]. Of the remaining IPSs, stereovision cameras are ideal as they have 10 cm accuracy and are not costly or easily circumventable due to having no tags or devices to be worn or carried.

The third requirement of *logical access control should be enforced in a physical space*, requires a location-based access control model to make use of location information to prevent unwarranted viewing of information with minimal disruption to employees. If a subject enters a virtual perimeter, the actions in policy must be executed to prevent viewing, for example, by dimming of screens, removal of windows from view [14], or prevention of access to a room containing documents in plain view. Each approach should be tailored to the nature and topology of the space to be effective.

Considering the above, the only approaches that satisfy the three requirements are the two camera-based approaches under the pattern recognition category. Approaches such as [14] use line-of-sight cameras to perform access control at end points. Stereo camera research is currently an active study field with approaches such as [21] using depth sensing for localization of entities for use in environmental mapping and room supervision. Technologies on the market such as [22] provide a native means to produce 3D maps of the environment using only the depth-sensing camera and image processing.

Even with these advances, there is still a need to apply such technologies for use in access control. Coupled with the use of biometrics and image processing technologies, computer vision systems provide a unique opportunity to automate intelligent access control systems that require minimal administrative control which reduces the infringement on privacy that cameras generally pose. Biometric systems such as gait recognition can provide greater amounts of information about the context of entities within the environment such as the direction a subject is facing or, when coupled with learning algorithms, if a subject is behaving suspiciously [23]. Such advances provide abilities that the other non-computer-vision-based approaches do not provide in one package.

It is therefore the view of the researchers that the integration of stereo camera systems into access control can be further explored for virtual perimeter access control.

7 Conclusion

This research presents a virtual perimeter as an indoor analog to a geo-fence to protect information that can be given a physical form. The researchers identified the need for accurate location from indoor positioning systems (IPSs) to determine if a virtual perimeter is crossed so that access control can be performed. To enable this, a set of

security requirements were extracted from the scenario and IPSs were compared with respect to their accuracy, usability, and cost. Pattern recognition systems emerged as a contender to satisfy the requirements as they are inexpensive and do not require tags to be carried by subjects making them more difficult to circumvent. This leads them to be ideal candidates for integration into a virtual perimeter access control model.

Furthermore, the researchers identified that stereo camera technologies can accurately localize entities within a physical space while also providing additional context information about the subject such as posture and direction from gait recognition. The researchers therefore conclude that stereoscopic cameras can be a viable technology for location-based access control to be performed in a physical space, wherein logical information may be given a physical form and still requires access to be controlled.

Future work entails an evaluation of computer vision-based IPSs for access control.

References

1. O. Vermesan, P. Friess (eds.), *Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems* (River Publishers, 2013)
2. C. Tsigkanos, L. Pasquale, C. Ghezzi, B. Nuseibeh, On the interplay between cyber and physical spaces for adaptive security. *IEEE Trans. Dependable Secure Comput.* **15**(3), 466–480 (2018)
3. R.T. Fernandez, S. Birse, U.S. Patent 9,646,477, U.S. Patent and Trademark Office, Washington, DC, 2017
4. E. Young-Hyun, C. Young-Keun, S. Cho, B. Jeon, *FloGeo: A Floatable Three-Dimensional Geofence with Mobility for the Internet of Things* (2017)
5. A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M.A. Al-Ammar, H.S. Al-Khalifa, Ultra wideband indoor positioning technologies: analysis and recent advances. *Sensors* **16**(5), 707 (2016)
6. J. Duque Domingo, C. Cerrada, E. Valero, J.A. Cerrada, Indoor positioning system using depth maps and wireless networks. *J. Sensors* **2016** (2016)
7. R.S. Sandhu, P. Samarati, Access control: principle and practice. *IEEE Commun. Mag.* **32**(9), 40–48 (1994)
8. A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry, Challenges for securing cyber physical systems, in *Workshop on Future Directions in Cyber-Physical Systems Security*, vol. 5 (2009)
9. N. Skandhakumar, J. Reid, E. Dawson, R. Drogemuller, F. Salim, An authorization framework using building information models. *Comput. J.* **55**(10), 1244–1264 (2012)
10. N. Fallah, I. Apostolopoulos, K. Bekris, E. Folmer, Indoor human navigation systems: a survey. *Interact. Comput.* **25**(1), 21–33 (2013)
11. R.A. Jarvis, A perspective on range finding techniques for computer vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 122–139 (1983)
12. B. Greaves, M. Coetzee, W.S. Leung, Access control requirements for physical spaces protected by virtual perimeters, in *International Conference on Trust and Privacy in Digital Business* (2018), pp. 182–197
13. J.L. Hernández, M.V. Moreno, A.J. Jara, A.F. Skarmeta, A soft computing based location-aware access control for smart buildings. *Soft Comput.* **18**(9), 1659–1674 (2014)
14. C.D. Jensen, K. Geneser, I.C. Willemoes-Wissing, Sensor enhanced access control: extending traditional access control models with context-awareness, in *IFIP International Conference on Trust Management*, June 2013 (Springer, Berlin, Heidelberg, 2013), pp. 177–192

15. G. Deak, K. Curran, J. Condell, A survey of active and passive indoor localisation systems. *Comput. Commun.* **35**(16), 1939–1954 (2012)
16. L. Ciabattoni, G. Foresi, A. Monteriù, L. Pepa, D.P. Pagnotta, L. Spalazzi, F. Verdini, Real time indoor localization integrating a model based pedestrian dead reckoning on smartphone and BLE beacons. *J. Ambient Intell. Hum. Comput.* 1–12 (2017)
17. K.W. Chiang, J.K. Liao, S.H. Huang, H.W. Chang, C.H. Chu, The performance analysis of space resection-aided pedestrian dead reckoning for smartphone navigation in a mapped indoor environment. *ISPRS Int. J. Geo-Inf.* **6**(2), 43 (2017)
18. L.M. Ni, Y. Liu, Y.C. Lau, A.P. Patil, LANDMARC: location sensing using active RFID. *Wireless Netw.* **10**(6), 701–710 (2004)
19. E. Brassart, C. Pegard, M. Mouaddib, Localization using infrared beacons. *Robotica* **18**(2), 153–161 (2000)
20. L. Priya, S. Anand, Object recognition and 3D reconstruction of occluded objects using binocular stereo. *Cluster Comput.* 1–10 (2017)
21. A. Burbano, M. Vasiliu, S. Bouaziz, 3D cameras benchmark for human tracking in hybrid distributed smart camera networks, in *Proceedings of the 10th International Conference on Distributed Smart Camera* (ACM, 2016), pp. 76–83
22. Stereolabs Homepage. <https://www.stereolabs.com/zed/>. Accessed 18 Dec 2018
23. A.S. George, E. Roy, A. Antony, M. Job, *An Efficient Gait Recognition System for Human Identification using Neural Networks* (2017)

q -LMF: Quantum Calculus-Based Least Mean Fourth Algorithm



Alishba Sadiq, Muhammad Usman, Shujaat Khan, Imran Naseem,
Muhammad Moinuddin and Ubaid M. Al-Saggaf

Abstract Herein, we propose a new class of stochastic gradient algorithm for channel identification. The proposed q -least mean fourth (q -LMF) is an extension of the least mean fourth (LMF) algorithm and it is based on the q -calculus which is also known as Jackson's derivative. The proposed algorithm utilizes a novel concept of error correlation energy and normalization of signal to ensure a high convergence rate, better stability, and low steady-state error. Contrary to conventional LMF, the proposed method has more freedom for large step sizes. Extensive experiments show significant gain in the performance of the proposed q -LMF algorithm in comparison to the contemporary techniques.

Keywords q -Calculus · System identification · q -LMF

A. Sadiq · I. Naseem
COE, PAF-KIET, Karachi, Pakistan
e-mail: alishba.sadiq@pafkiet.edu.pk

I. Naseem
e-mail: imrannaseem@pafkiet.edu.pk

M. Usman · S. Khan (✉)
FEST, Iqra University, Karachi, Pakistan
e-mail: shujaat@iqra.edu.pk

M. Usman
e-mail: musman@iqra.edu.pk

M. Moinuddin · U. M. Al-Saggaf
Center of Excellence in Intelligent Engineering Systems (CEIES),
King Abdulaziz University, Jeddah, Saudi Arabia
e-mail: mmsansari@kau.edu.sa

U. M. Al-Saggaf
e-mail: usaggaf@kau.edu.sa

Electrical and Computer Engineering Department,
King Abdulaziz University, Jeddah, Saudi Arabia

1 Introduction

The modern wireless communication systems provide a reasonable trade-off between performance parameters. They provide high throughput with mobility while maintaining efficient utilization of limited bandwidth. Such cost-effective developments come with a number of stumbling blocks and invigorating challenges. Channel estimation is one such essential technique which is effectively used to enhance the performance of the modern wireless communication systems. It is a widely used technique specifically in mobile wireless network systems as the wireless channel shows significant variations over time, and generally, these variations are caused by a number of reasons such as transmitter or receiver being in motion at high speed. Multi-path interference from surroundings, such as highlands, buildings, and other hindrances also affect mobile wireless communication. In order to offer consistency, accuracy and high data rates at the receiver, accurate estimates of the time-varying channel are the requirement. Linear models provide reasonable estimates with reduced bit error rate (BER), thereby improving the capacity of the system [12]. Adaptive learning methods are widely used to estimate the characteristics of the communication medium. Due to their simplicity and ease of implementation, the least square-based methods are considered to be widely used optimization techniques for adaptive systems. The technique has been applied in diversified applications such as function approximation [10], detection of elastic inclusions [1], noise cancelation [9], nonlinear system identification [11], ECG signal analysis [20], elasticity imaging [6], and time series prediction [15]. Adaptive filters are used to extract the desired components from a signal containing both desired and undesired components. The least mean square (LMS) is a popular choice for designing adaptive filters. However, it has a slow convergence rate [7]. Besides these variants, various definitions of gradient have also been used to derive improved LMS algorithms [3, 13]. The algorithm is derived using Riemann–Liouville fractional derivative for high convergence performance. In [2, 14], some adaptive schemes were proposed for maintaining stability through adaptive variable fractional power. The FOC variants are, however, not stable and diverge if the weights are negative or the input signal is complex [16, 18, 21].

Recently, Jackson's derivative q -steepest descent algorithm is proposed that computes the normal to the cost function to achieve a higher convergence rate [5]. The q -LMS algorithm has also been used for a number of applications, such as system identification, and designing of whitening filter [4]. In [17], adaptive frameworks are proposed for q -parameter. In this research, we propose a modified variant of the stochastic gradient descent method by utilizing the q -steepest decent approach. The proposed method is an extension of the least mean fourth (LMF) algorithm which minimizes the fourth power of the instantaneous estimation error. The conventional LMF can achieve higher convergence compared to the LMS algorithm [22]. However, it is inherently prone to instability due to the cubic power of the error signal in its update rule. In the proposed q -calculus-based LMF (q -LMF) algorithm, the performance of the conventional LMF can be improved while maintaining the stability of the algorithm by using an additional controlling term q .

1.1 Research Contributions and Paper Organization

Following are the main contributions:

- A novel adaptive learning algorithm is derived for the identification of linear systems. In particular, Jackson’s derivative-based variant of LMF is derived using the q -gradient descent method [5].
- The reactivity of the proposed q -LMF algorithm is analyzed for q controlling parameter.
- An interesting application of XE-NLMF filter is presented, and a time-varying normalization technique is designed.
- Performance of the LMS and the LMF algorithms are compared to the q -LMF.
- The *Big Oh* complexity is also analyzed. The q -LMF achieves significantly improved performance at the cost of very low computational overhead.
- Performance claims are validated through computer simulations for a linear channel estimation task.

We organize the paper as follows. A detailed overview of the q -calculus is explained in Sect. 2. In Sect. 3, the description of the proposed algorithm is provided, followed by the experimental findings in Sect. 4. The paper is concluded in Sect. 5.

2 Basics of q -Calculus

Quantum calculus is also referred to as the calculus without a limit. It has been successfully used in various areas including number theory, adaptive filtering, operational theory, mechanics, and the theory of relativity [8]. In q -calculus, the differential of a function is defined as $d_q(p(x)) = p(qx) - p(x)$. The derivative therefore takes the form $D_q(p(x)) = \frac{d_q(p(x))}{d_q(x)} = \frac{p(qx)-p(x)}{(q-1)x}$, and when $q \rightarrow 1$, the expression becomes the derivative in the classical sense. For the form x^n , the q -derivative of a function is given as:

$$D_{q,x}x^n = \begin{cases} \frac{q^n - 1}{q - 1}x^{n-1}, & q \neq 1, \\ nx^{n-1}, & q = 1. \end{cases} \tag{1}$$

The q -gradient of a function $p(x)$ for n number of variables, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ is given as $\nabla_{q,w}p(x) \triangleq [D_{q1,x1}p(x), D_{q2,x2}p(x), \dots, D_{qn,xn}p(x)]^T$, where $q = [q_1, q_2, \dots, q_N]^T$ [19]. There is also a rule similar to the chain rule for ordinary derivatives. Let $g(x) = cx^k$. Then, $D_qp(g(x)) = D_q^k(f)(g(x))D_q(g)(x)$.

3 Proposed q -Least Mean Fourth (q -LMF) Algorithm

By utilizing the idea of steepest descent with the following weight update rule, the conventional LMF algorithm is obtained

$$\mathbf{w}(i+1) = \mathbf{w}(i) - \frac{\eta}{4} \nabla_w J(\mathbf{w}), \quad (2)$$

where η is the step size, $J(\mathbf{w})$ is the cost function for the q -LMF algorithm and is defined as $J(\mathbf{w}) = e^4(i)$, where $e(i)$ is the estimation error which is the deviation between the output signal at the i th instant and the desired response $d(i)$, i.e., $e(i) = d(i) - \mathbf{w}^\top(i)\mathbf{x}(i)$. Here, $\mathbf{x}(i)$ is the input signal vector defined as $\mathbf{x}(i) = [x_1(i), x_2(i), \dots, x_M(i)]^\top$, and $\mathbf{w}(i)$ is the vector which consists of weights given as $\mathbf{w}(i) = [w_1(i), w_2(i), \dots, w_M(i)]$, where M is the length of the filter.

The q -LMF utilizes the Jackson derivative method [5], and it takes larger steps (for $q > 1$) toward optimal solution. To derive q -LMF algorithm, conventional gradient in (2) can be replaced with the q -gradient, we get

$$\mathbf{w}(i+1) = \mathbf{w}(i) - \frac{\eta}{4} \nabla_{q,w} J(\mathbf{w}). \quad (3)$$

The q -gradient of the cost function $J(\mathbf{w})$ for the k th weight is defined as $\nabla_{q,w_k} J(\mathbf{w}) = \frac{\partial_{q_k} J(\mathbf{w})}{\partial_{q_k} e} \frac{\partial_{q_k} e(i)}{\partial_{q_k} y} \frac{\partial_{q_k} y(i)}{\partial_{q_k} w_k(i)}$. Solving partial derivatives using the Jackson derivative defined in Sect. 2 gives $\frac{\partial_{q_k} J(\mathbf{w})}{\partial_{q_k} e} = \frac{\partial_{q_k} (e^4(i))}{\partial_{q_k} e} = \frac{q_k^4 - 1}{q_k - 1} e^3(i) = (q_k^3 + q_k^2 + q_k + 1)e^3(i)$, where $J(\mathbf{w}) = e^4(i)$, we employ the instantaneous error term and it is given as $e(i) = d(i) - y(i)$. Substituting $\frac{\partial_{q_k} y(i)}{\partial_{q_k} w_k(i)} = \mathbf{x}_k(i)$, and $\frac{\partial_{q_k} e(i)}{\partial_{q_k} y} = -1$ gives $\nabla_{q,w_k} J(\mathbf{w}) = -(q_k^3 + q_k^2 + q_k + 1)e^3(i)\mathbf{x}_k(i)$. Similarly, for $k = 1, 2, \dots, M$,

$$\begin{aligned} \nabla_{q,w} J(\mathbf{w}) = & -(q_1^3 + q_1^2 + q_1 + 1)e^3(i)x_1(i), (q_2^3 + q_2^2 + q_2 + 1)e^3(i)x_2(i), \\ & \dots, (q_M^3 + q_M^2 + q_M + 1)e^3(i)x_M(i). \end{aligned} \quad (4)$$

Consequently, Eq. (4) can be written as $\nabla_{q,w} J(\mathbf{w}) = -4E[\mathbf{G}\mathbf{x}(i)e^3(i)]$, where η is the learning rate and \mathbf{G} is a diagonal matrix $\text{diag}(\mathbf{G}) = [(\frac{q_1^3+q_1^2+q_1+1}{4}), (\frac{q_2^3+q_2^2+q_2+1}{4}), \dots, (\frac{q_M^3+q_M^2+q_M+1}{4})]^\top$. Due to the ergodicity of the system $\nabla_{q,w} J(\mathbf{w})$ results in $\nabla_{q,w} J(\mathbf{w}) \approx -4\mathbf{G}\mathbf{x}(i)e^3(i)$. Substituting $\nabla_{q,w} J(\mathbf{w})$ in (2) renders the learning method of the q -LMF algorithm by

$$\mathbf{w}(i+1) = \mathbf{w}(i) + \eta \mathbf{G}\mathbf{x}(i)e^3(i). \quad (5)$$

3.1 Formulation of the q XE-LMF

The least mean fourth (LMF) algorithm shows better performance in non-Gaussian environment compared to the LMS [22]. For its stability, normalized versions of the LMF algorithm are proposed that enhanced its performance and stability in Gaussian noisy environments. For example, the newly developed normalized LMF (XE-NLMF) algorithm is normalized by the error powers and mixed signal, and fixed mixed-power parameter is used to give it some weight [23]. As quantum calculus-based algorithms show faster convergence, it is thought-provoking to expand the concept of normalization for the proposed q -LMF algorithm. The proposed q XE-NLMF algorithm is expressed by the following equation:

$$\mathbf{w}(i+1) = \mathbf{w}(i) + \eta \mathbf{G} e^3(i) \mathbf{x}(i), \quad (6)$$

where $\text{diag}(\mathbf{G}) = 1/(\delta + \alpha \|\mathbf{x}\|^2 + (1 - \alpha) \|e\|^2)$, δ is a small value added to avoid indeterminate form, the notation $\|\cdot\|^2$ shows the squared Euclidean norm of a vector and α is the mixing power parameter.

3.2 Computational Complexity

Computational complexity is an important performance measure of the learning algorithms. We analyze the computation cost of the LMS, the LMF, and the proposed q -LMF algorithm. For each iterations, the proposed q -LMF requires $3M + 3$ multiplications and $2M$ additions, which is $M + 2$ and M multiplications expensive compared to the LMS and LMF algorithms, respectively. Here, M denotes the number of unknown weight parameters. The additional multiplications are required due to the presence of diagonal matrix \mathbf{G} . Additionally, in initialization step, the proposed q -LMF requires $3M$ multiplications and $3M$ additions only once to compute the \mathbf{G} matrix.

4 Experiments

For a system identification scenario, the performance of the proposed q -LMF algorithm is examined in this section. Channel estimation, for instance, is a widely used method in communication systems to estimate the characteristics of an unknown channel.

The mathematical model of the system is defined as $y(t) = h_1 x(t) + h_2 x(t-1) + h_3 x(t-2) + h_4 x(t-3) + d(t)$, where the input and output of the system are denoted by $x(t)$ and $y(t)$, respectively, and $d(t)$ is the disturbance which is taken to be white uniform noise in this case. For this experiment, $\mathbf{x}(t)$ is chosen to be 1×10^4 randomly

generated samples obtained from Gaussian distribution of zero mean and variance of 1. For the simulation purpose, the signal-to-noise ratio (SNR) is set to be 10 dB. The experiments are repeated for 1000 Monte Carlo independent runs and mean results are reported. Impulse response of the system $h_{round\#}$ is in each simulation round set randomly, and weights of the adaptive filter were reset to zero. Following are the objectives of our simulations:

- To show the performance gain of q -LMF over LMS in non-Gaussian environment. In particular, we compare the performance of the proposed algorithm with the conventional LMS when q -LMF is operating as conventional LMF filter (i.e., $q = 1$).
- To test the reactivity of the q -LMF algorithm over q parameter.
- To evaluate the performance of the proposed algorithm when q -LMF operating as XE-NLMF.

For the performance analysis, the normalized weight difference (NWD) between the actual and the estimated weights is calculated. In particular, we define $NWD = \frac{\|\mathbf{h} - \mathbf{w}\|}{\|\mathbf{h}\|}$, where \mathbf{w} is the obtained weight-vector and \mathbf{h} is the actual impulse response of the channel.

4.1 The Proposed q -LMF as the LMF Algorithm

In this experiment, we compare the performance of q -LMF algorithm with the conventional LMS algorithm. It is well established in [22], that when operating in a non-Gaussian environment the LMS perform inferior to the LMF. For the evaluation of the proposed method in non-Gaussian environment, we choose the step size $\eta = 1e^{-3}$ for both the LMS and the proposed q -LMF with a fixed value of q , i.e., 1. It can be seen from Fig. 1 (first column) that when operating in non-Gaussian noise environment q -LMF also outperforms the LMS algorithm in steady-state error and convergence rate measures. The proposed method converges at 2000th iteration, whereas the convergence of the LMS algorithm is achieved at 4000th iteration at a higher steady-state error compared to q -LMF. In steady-state error, the proposed q -LMF algorithm outperformed the LMS algorithm by a margin of approximately 0.6 dB. The steady-state error for LMS is -18.239 dB while for the proposed q -LMF it is relatively smaller, i.e., -18.827 dB. q -LMF exhibits results in comparison to the traditional LMS.

4.2 Reactivity Study of the Proposed q -LMF

We observe the change in response of the q -LMF algorithm with the change in q parameter. Specifically, we conducted the simulations for a system identification problem and compare the normalized weight difference (NWD) learning graphs of

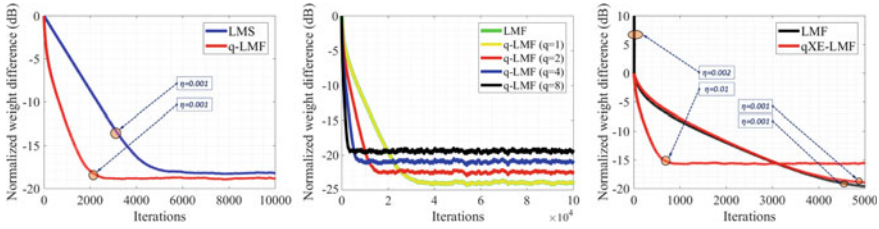


Fig. 1 (first column) NWD behavior for the proposed q -LMF and the LMF algorithm. (second column) Sensitivity of q -LMF. (third column) Comparison of the conventional LMF and the proposed q XE-LMF

the proposed q -LMF algorithm on various values of q (see Fig. 1 (second column)). We considered four different values of q , i.e., $q = 1, q = 2, q = 4$, and $q = 8$. Figure 1 (second column) shows that for $q = 1$ the proposed q -LMF behaves exactly like the conventional LMF algorithm. Note that for higher values of q , the proposed q -LMF shows faster learning accompanied with a larger final error. The q -LMF took the largest number of iterations for $q = 1$, i.e., at 28,000 while for greater values of q it took lesser iterations such as for $q = 2, q = 4$, and $q = 8$ it took 14,000, 6000, and 2500 iterations to converge.

4.3 The Proposed q -LMF as XE-LMF

Stability of the LMF algorithm is difficult to achieve, since the cubic power of the error (e^3) in the LMF gradient vector can create overwhelming initial uncertainty. To solve this problem we propose to use the q -LMF in XE-NLMF mode, this will help q -LMF to operate at higher values of step size on which the conventional LMF diverges and the proposed algorithm can achieve better performance. In this section, we show the comparison of the LMF and the proposed q -LMF algorithm for large and small values of step size. In Fig. 1 (third column), when operating at $\eta = 0.001$ it can be noticed that both the algorithms show slow convergence. At a higher convergence rate, we simulated the LMF algorithm for same simulation setup with 2 times greater value of step size, i.e., $\eta = 0.002$ at which it shows the divergence while the proposed q -LMF when operating in XE-NLMF mode (derived in (6)) can operate at even higher convergence speed (i.e., $\eta = 0.01$). Overall, the proposed q -LMF algorithm when operating in XE-NLMF mode can achieve higher convergence while maintaining the stability.

5 Conclusion

In this research, we proposed a q -calculus-based LMF algorithm called q -LMF. The proposed algorithm provides additional control over the convergence and steady-state performances through q parameter. The standard LMS and LMF algorithms are compared to the proposed q -LMF for a problem of channel estimation in non-Gaussian environment. The algorithms are compared on the basis of steady-state error, convergence rate, and computational complexity. The simulations were repeated for 1000 independent Monte Carlo simulation rounds at 10 dB SNR value. Overall, the proposed q -LMF algorithm comprehensively outperformed the LMS, and the LMF algorithms, achieving the best steady-state error and convergence rate performances.

References

1. T. Abbas, S. Khan, M. Sajid, A. Wahab, J.C. Ye, Topological sensitivity based far-field detection of elastic inclusions. *Results Phys.* **8**, 442–460 (2018)
2. J. Ahmad, M. Usman, S. Khan, I. Naseem, H.J. Syed, RVP-FLMS: a robust variable power fractional LMS algorithm, in *2016 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (IEEE, 2016)
3. J. Ahmad, S. Khan, M. Usman, I. Naseem, M. Moinuddin, FCLMS: fractional complex LMS algorithm for complex system identification, in *13th IEEE Colloquium on Signal Processing and its Applications (CSPA 2017)* (IEEE, 2017)
4. U.M. Al-Saggaf, M. Moinuddin, A. Zerguine, An efficient least mean squares algorithm based on q -gradient, in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 891–894
5. U.M. Al-Saggaf, M. Moinuddin, M. Arif, A. Zerguine, The q -Least Mean Squares algorithm. *Signal Process.* **111**(Suppl. C), 50–60 (2015)
6. H. Ammari, E. Bretin, J. Garnier, H. Kang, H. Lee, A. Wahab, *Mathematical Methods in Elasticity Imaging* (Princeton University Press, 2015)
7. S.C. Douglas, A family of normalized LMS algorithms. *IEEE Signal Process. Lett.* **1**(3), 49–51 (1994)
8. T. Ernst, *A Comprehensive Treatment of q -Calculus*, 1st edn. (Springer, Basel, 2012)
9. J.M. Górriz, J. Ramírez, S. Cruces-Alvarez, C.G. Puntonet, E.W. Lang, D. Erdogmus, A novel LMS algorithm applied to adaptive noise cancellation. *IEEE Signal Process. Lett.* **16**(1), 34–37 (2009)
10. S. Khan, I. Naseem, R. Togneri, M. Bennamoun, A novel adaptive kernel for the RBF neural networks. *Circuits Syst. Signal Process.* **1–15**, (2016)
11. S. Khan, J. Ahmad, I. Naseem, M. Moinuddin, A novel fractional gradient-based learning algorithm for recurrent neural networks. *Circuits Syst. Signal Process.* **1–20**, (2017)
12. S. Khan, N. Ahmed, M.A. Malik, I. Naseem, R. Togneri, M. Bennamoun, FLMF: fractional least mean fourth algorithm for channel estimation in non-Gaussian environment, in *International Conference on Information and Communications Technology Convergence 2017 (ICTC 2017)* (Jeju Island, Korea, October 2017)
13. S. Khan, M. Usman, I. Naseem, R. Togneri, M. Bennamoun, A robust variable step size fractional least mean square (RVSS-FLMS) algorithm, in *13th IEEE Colloquium on Signal Processing and its Applications (CSPA 2017)* (IEEE, 2017)

14. S. Khan, M. Usman, I. Naseem, R. Togneri, M. Bennamoun, VP-FLMS: a novel variable power fractional LMS algorithm, in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN) (ICUFN 2017)* (Italy, Milan, July 2017)
15. S. Khan, I. Naseem, M.A. Malik, R. Togneri, M. Bennamoun, A fractional gradient descent-based RBF neural network. *Circuits Syst. Signal Process.* **1–22**, (2018)
16. S. Khan, I. Naseem, A. Sadiq, J. Ahmad, M. Moinuddin, Comments on “Momentum fractional LMS for power signal parameter estimation”. arXiv preprint [arXiv:1805.07640](https://arxiv.org/abs/1805.07640) (2018)
17. S. Khan, A. Sadiq, I. Naseem, R. Togneri, M. Bennamoun, Enhanced q -least mean square. arXiv preprint [arXiv:1801.00410](https://arxiv.org/abs/1801.00410) (2018)
18. S. Khan, A. Wahab, I. Naseem, M. Moinuddin, Comments on “Design of fractional-order variants of complex LMS and NLMs algorithms for adaptive channel equalization”. arXiv preprint [arXiv:1802.09252](https://arxiv.org/abs/1802.09252) (2018)
19. J. Koekoek, R. Koekoek, A note on the q -derivative operator. *ArXiv Mathematics e-prints* (1999)
20. N.V. Thakor, Y.S. Zhu, Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Trans. Biomed. Eng.* **38**(8), 785–794 (1991)
21. A. Wahab, S. Khan, Comments on “Fractional extreme value adaptive training method: fractional steepest descent approach”. arXiv preprint [arXiv:1802.09211](https://arxiv.org/abs/1802.09211) (2018)
22. E. Walach, B. Widrow, The least mean fourth (LMF) adaptive algorithm and its family. *IEEE Trans. Inf. Theor.* **30**(2), 275–283 (2006)
23. A. Zerguine, M.K. Chan, T.Y. Al-Naffouri, M. Moinuddin, C.F. Cowan, Convergence and tracking analysis of a variable normalised LMF (XE-NLMF) algorithm. *Signal Process.* **89**(5), 778–790 (2009)

Process Driven Access Control and Authorization Approach



John Paul Kasse, Lai Xu, Paul de Vrieze and Yuewei Bai

Abstract Compliance to regulatory requirements is key to successful collaborative business process execution. The review of the EU General Data Protection Regulation (GDPR) brought to the fore the need to comply with data privacy. Access control and authorization mechanisms in workflow management systems based on roles, tasks, and attributes do not sufficiently address the current complex and dynamic privacy requirements in collaborative business process environments due to diverse policies. This paper proposes process driven authorization as an alternative approach to data access control and authorization where access is granted based on a legitimate need to accomplish a task in the business process. Due to vast sources of regulations, a mechanism to derive and validate a composite set of constraints free of conflicts and contradictions is presented. An extended workflow tree language is also presented to support constraint modeling. An industry case pick and pack process is used for illustration.

Keywords Compliance · Collaborative business process · Verification and validation

The original version of this chapter was revised: The author's name has been corrected to "Paul de Vrieze". The correction to this chapter is available at https://doi.org/10.1007/978-981-15-0637-6_45

J. P. Kasse (✉) · L. Xu · P. de Vrieze
Computing and Informatics Faculty of Science and Technology, Bournemouth University,
Poole BH12 5BB, UK
e-mail: jkasse@bournemouth.ac.uk

L. Xu
e-mail: lxu@bournemouth.ac.uk

P. de Vrieze
e-mail: pdvrieze@bournemouth.ac.uk

Y. Bai
Industry Engineering of Engineering College, Shanghai Polytechnic University,
Jinhai Road 2360, Pudong, Shanghai, P. R. China
e-mail: ywbai@sspu.edu.cn

1 Introduction

Compliance requires strict adherence to policies, norms, and regulations by an organization's business processes which translate into products and services, e.g., products must meet quality standards, systems must observe data privacy etc. Non-compliance is punishable with monetary fines or litigations. Business processes aim to achieve business objectives, yet compliance objectives provide a form of controls that constrain the business process and overall operations.

To achieve a balance between objectives and compliance requirements, a compliance by design approach is adopted where both business and compliance requirements are designed into the process. Data privacy management is a key driver made mandatory by the EU's General Data Protection Regulation (GDPR). It requires privacy by design, which in the business process context impacts the entire engineering process. Separation of duty (SoD) and binding of duty (BoD) [1, 2] are other forms of constraints restricting process behavior from Sarbanes Oxley Act and Basel II.

Business processes are constrained by both company internal and external policies. As policies restrict valid executions of processes (or combinations of processes), these restrictions could lead to deadlocks in the process where the process is incapable of meeting the policy requirements [3]. For example, in a complex process with multiple restrictions, the four eyes principle could lead to a problem where there is only one authorized person that meets the other restrictions. This makes the need for verification of process behavioral conformance with constraints legitimate.

Existing compliance frameworks do not address conflict checking among regulatory requirements [4, 5]. In a collaborative environments where different policies apply, an illustration of how to achieve a composite policy set and verifying it against contradictions, inconsistency, and inaccuracy is desirable.

To address constraint modeling and validation problem in the context of regulatory requirements, an extended workflow tree language with constructs like OR, loops, and time is presented. Using a constrained process model, we illustrate process driven authorization as a data access control mechanism with the case study introduced in [6].

The rest of the paper is organized as follows. Section 2 presents the motivating use case; Sect. 3 presents the proposed language illustrating application of extended constructs while Sect. 4 illustrates how to achieve a composite policy set and its verification. Section 5 discusses how to achieve process driven authorization. Section 6 presents related work and Sect. 7 is conclusion and outlook.

2 Use Case

Pick and pack process is based on actual industry use. It is collaborative and designed for use in international corporations (Europe and parts of Asia).

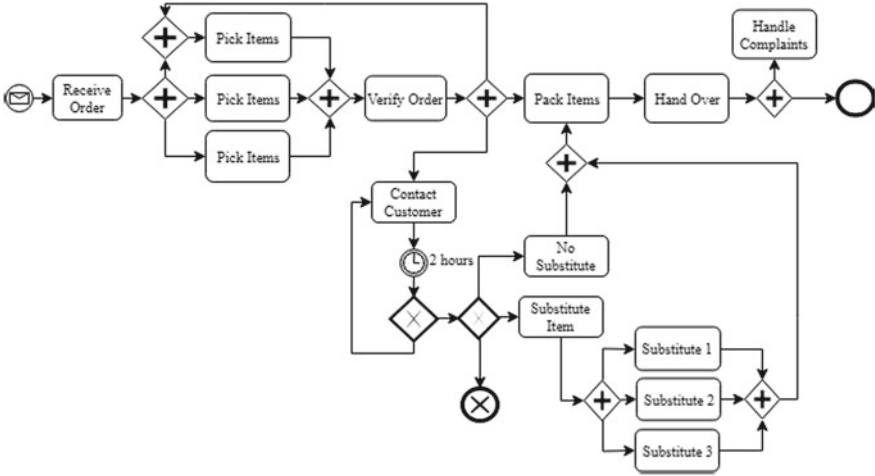


Fig. 1 A BPMN representation of the pick and pack business process

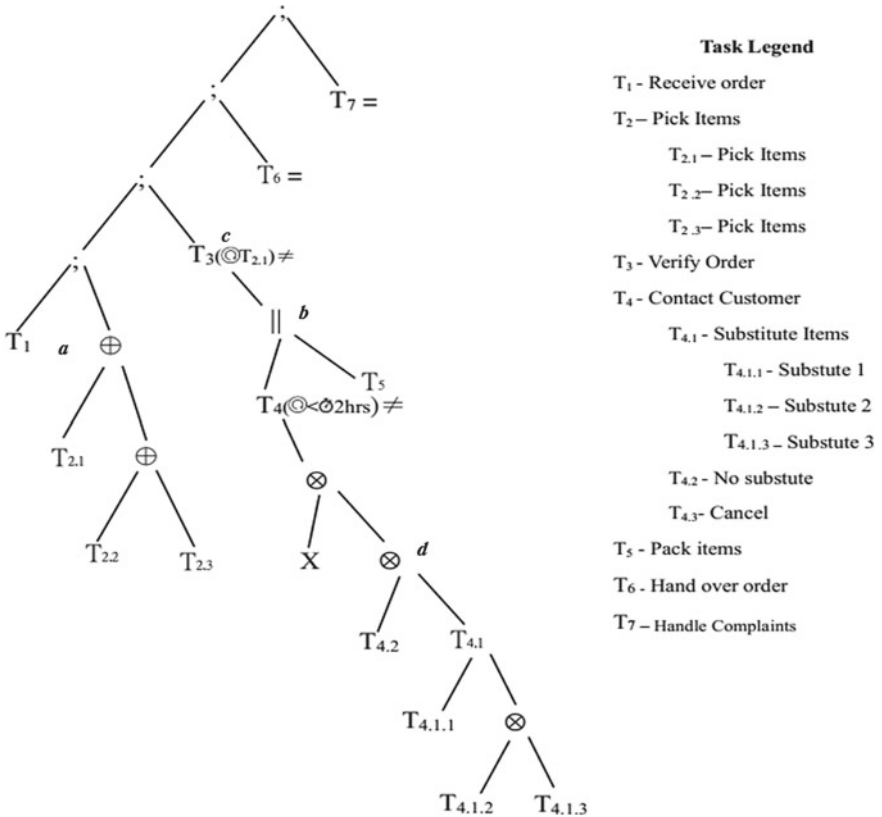


Fig. 2 A workflow tree representing a constrained pick and pack business process

To create orders (Figs. 1 and 2), customers register online. Once order is received, the customer and the store are notified. The store staffs check order details and proceed to pick and pack the order. Before handover, the order is verified to match with order details. For items that may be out of stock, the order is suspended for a period until stock is available or customer is contacted to seek opinion either to proceed without the item, substitute it, or cancel the order. Delayed orders can be canceled by customers; ready ones are picked or delivered by the delivery team. Individual stores may vary the process to fit specific contexts. Consequently, a family of process variants is created with different implications on the control flow and data resource allocations.

3 Workflow Tree Language

Several formal approaches are used in process modeling with BPMN being a standard from [7]. BPMN limitations like inability to expressively support intuitive and in-depth analysis of business process models involving simulation, validation, and verification [8]. To support this analysis, models are enhanced with annotations, e.g., security and safety [8–10], model verification [11], etc. BPMN may not be the best formalism to model and verify compliance constraints because annotations come with associated complexity.

To this effect, a workflow tree language (WTL) is proposed. WTL is a popular approach in process modeling and validation. In Nikovski and Akihiro [15], WTL is used to represent processes in a way that facilitates process mining in models where parallelism is not explicitly recorded. Crampton and Gutin [16] use WTL to express workflow model constraints to facilitate means to extend and solve the workflow satisfiability problem. The study however omits important constructs like the OR, loops, and time which are relevant for current business processes. These constructs form part of our extension as represented by symbols in Fig. 2 to support the modeling and analysis of processes which are collaborative, adaptive, and declarative [12] as well as expression of constraints relating to privacy, SoD, BoD, and need to know (Table 1).

Workflow trees provide a natural hierarchical representation of processes. In an ordered tree, the process tasks and functional units are represented by leaves and internal nodes, respectively. For instance, $\odot T_5$ represents a loop back to T_5 in a workflow. The X symbol is a cancel or termination, e.g., customer cancels the order

Table 1 Symbols and their meaning

Symbol	Name	Symbol	Name
	Parallel	•	Sequence
⊗	XOR	⊕	Inclusive OR
⊙	Loop	X	Cancel

due to delay. WTL extension is intended to support verification; (1) among constraints to identify conflicts and inconsistencies, and (2) between model and the constraints. Using compliance attributes in Kasse et al. [6], we illustrate to achieve modeling and compliance verification for process models.

3.1 Constraint Expression with WTL

Constraints limit the behavior of the business process in terms of task ordering, resource assignment, and data flow. WTL facilitates constraints expression in a manner useful to analyze and identify properties necessary to support their verification. Figure 2 is a WTL pick and pack process model with SoD (\neq) and BoD ($=$) constraints symbols adopted from [13]. Constraints expression over models yields complexity and task redundancy. This necessitates model verification to guarantee soundness.

In [6], useful compliance attributes in relation to branching and temporal constructs are suggested which we adopt to express compliance constraints over a WTL model. Figure 2 illustrates expression of serialization (*a*), parallelism (*b*), looping (*c*), XOR (*d*), and choice (*e*) constructs as segments of the use case.

The OR is likely to introduce redundancy in the workflow tree. For example, if three tasks are represented on a single node. The nesting of tasks or use of similar labels for two nodes that have parent/child relationship should be avoided to retain a sound workflow tree. Simply, add a child node to the current node. All nodes must have two children or otherwise be eliminated [14, 15]. Time-based constraints specify temporal requirements defined as absolute time or relative time, e.g., task durations, deadlines, task waiting time, resource availability, data access, and authorization schedules. These compound into total process duration, e.g., the total order processing duration is 6 h from submission time. Delays cause process costs or trigger exception handling tasks, e.g., when customers reject delayed orders, it leads to a cancellation.

4 Optimal Policy Derivation and Validation

A mechanism to achieve a composite set of policies from internal and external policies and their validation is described. Policies change overtime directly impacting on existing processes. Changes must be propagated to all areas where it has effect.

4.1 Optimal Policy Derivation

Regulations are specified in natural language without implementation specifics. Natural language can be a source of ambiguity. External regulations have a direct influence over internal policies and the two should not contravene, otherwise a violation

results in the business processes. Mapping internal and external policies has associated complexity or requires skills not common to compliancy officers. A mechanism to derive an optimal composite policy set is a step to solve the complexity and facilitate non-expert users.

Internal Policy set: composed of polices to regulate processes behavior. For instance, parties a, b, c collaborate on a business process each with individual internal policies. $IP_{\text{internal}.a} = \{P_{1.a}, \dots P_{n.a}\}$, $IP_{\text{internal}.b} = \{P_{1.b}, \dots P_{n.b}\}$, $IP_{\text{internal}.c} = \{P_{1.c}, \dots P_{n.c}\}$

Contractual Policy set: an integration of non-contradicting policies from IP_{internal} to form a set $P_{\text{contractual}}$ binding all parties. If there other relevant policies outside of the IP_{internal} , they are co-opted as P_{other} . Therefore,

$$P_{\text{contractual}} = \sum_{i=a}^c \bigcup_{P_{\text{internal}.i}}^{P_{\text{other}}} \quad (1)$$

Global Policies: composed of industry wide policies $P_{\text{global}} = \{R_{g1}, R_{g2} \dots R_{g3}\}$

Composite set: composed of global and contractual sets. Therefore

$$P_{\text{Composite}} = \sum \left(P_{\text{contractual}}, P_{\text{global}} \right) \quad (2)$$

The composite set should be complete to include all relevant policies.

4.2 Validation of the Derived Optimal Equation

To validate the composite policy set, we define and formalize consistency and completeness equations to support formal reasoning to identify potential errors.

Consistency—equation with at least one solution. The composite set is composed of all non-repeating policies compounded in contractual and global sets.

$$\exists(P_{\text{contractual}}, P_{\text{global}}) \rightarrow P_{\text{composite}} \quad (3)$$

Simple consistency—a composite policy set is consistent if and only if there is no policy ρ in Φ such that a policy ρ and its negation exists in the same set, otherwise Φ is inconsistent. No policy should allow and disallow actions at the same time, e.g., no resource assignment can be SoD and BoD at the same task otherwise a deadlock results.

$$\exists\rho(P_{\text{composite}}) \leftrightarrow \nexists\neg\rho(P_{\text{composite}}) \quad (4)$$

Maximum consistency—a composite policy set is maximally consistent if and only if for every policy ρ is part of the set.

$$\forall \rho \in (P_{\text{composite}}) \leftrightarrow \exists \rho \in (P_{\text{contractual}}) \quad (5)$$

Completeness: $P_{\text{composite}}$ should include all relevant policies from the internal, contractual, and global sets, otherwise it is incomplete.

$$\forall \rho \in (P_{\text{contractual}}) \supseteq (P_{\text{composite}}) \quad (6)$$

For all policies sets in contractual set are superset of the composite set.

From space limitation, it is not possible to illustrate the mechanism with the use case.

5 Toward Process Driven Authorization (PDA)

With a consistent composite policy set, PDA mechanism aims to control access to data based on legitimate and legalized purpose for which it is required in the process. Access is granted with respect to time and history of task executions in the workflow.

5.1 Constraint Formalization

Preliminary workflow W definitions are concerned with user—task assignment (u, t), user—role assignment ($UR \in UXR$), and role—permission assignment ($RP \in RXP$).

i. SoD constraint δ for two workflow tasks T_1 and T_2 is a tuple expressed as

$$\delta_{t_1 \in T_1, t_2 \in T_2} \rightarrow \neg \exists u \in U \{(u, t_1), (u, t_2)\} \subseteq W \quad (7)$$

δ Constraint is satisfied iff there exists different users assigned to tasks t_1 or t_2 in W

ii. BoD constraint, β a user is assigned to execute two conjoint tasks t_1 and t_2

$$\beta_{t_1, t_2 \neq t_1} \rightarrow \forall u \in U [(u, t_1) \subseteq W \rightarrow ((u, t_2) \subseteq W \wedge \neg \exists u' \neq u [(u', t_2) \in W])] \quad (8)$$

β Constraint is satisfied if there exists a user assigned to execute tasks t_1 and t_2 in W , e.g., tasks ‘pack items’ and ‘verify order’ are executed by different users

iii. *Need to Know (N2K) constraint* η assigns special permission to execute task and access necessary data

- iv. Authorization policy \wp over a workflow is a triple composed of constraints SoD, BoD and need to know.

$$\wp \rightarrow (\beta, \delta, \eta) \quad (9)$$

Workflow history includes past executed task instances relevant to future user task assignment (UT). This makes the element of temporal constraint relevant. Temporal constraints assignment applies to the user, object, action to be executed and the intention to allow or deny access, i.e.,

$$tm = (I, U, A, +/-) \quad (10)$$

where I is the period interval, U is the subject or user, A is the action to be taken (e.g., read) and $+/-$ permissions to allow or deny time-based access. These variables fit well with the proposed time-based compliance attributes in Kasse et al. [6], e.g., AllowBefore, AllowAt, DenyBefore, DenyAt, etc. Since the user is already part of the task assignment, it is withdrawn to retain the formula as

$$tm = (I, A, +/-) \quad (11)$$

Therefore, an authorization policy with temporal constraint is

$$\wp \rightarrow (\beta, \delta, \eta, tm) \quad (12)$$

Additionally, access under PDA is granted with respect to history h executions. A valid constrained workflow model is one that satisfies the authorization policy in reference to the execution history. The history is important during execution to check whether a previous user has right to access current task in reference to SOD and BOD. Formally, a constrained workflow model CW with a history is;

$$CW \rightarrow (\wp, h) \quad (13)$$

where h is workflow history. An execution of a workflow model satisfying all constraints is an authorized model under PDA. PDA is achieved as a service at runtime which is contacted whenever a task is to execute. The authorization engine checks the assignments and grants or denies access.

6 Related Work

The consistency of task-based constraints is addressed in [5] where the authors derive a consistent constrained workflow schema. However, the study did not consider temporal constraints which we have addressed in this paper. Crampton and Gutin address

workflow satisfiability problem using constraint expression in [16] and refine it in [17]. Compliance of a workflow to specified constraints is considered a workflow satisfiability problem which they provide solution to. Like the previous study, temporal constraints were ignored. In Basin [13], an approach for deriving an optimal workflow aware authorization is presented as an NP hard problem and solved as a parameter tractable problem. We did not take that direction though it is a future plan. In [18 and 19], a tool is implemented to automate the enforcement of privacy policies and requirements on personal data used in organization systems. The tool disregards other forms of compliance based on business process perspectives. In all, the studies are relevant to the subject of compliance to regulations. However, none of them specifically supports optimal policy derivation as well as its validation.

7 Conclusion and Outlook

This paper presents an explicit mechanism to compose and validate policies that originate from different sources. By presenting a mechanism to integrate, validate and verify different policy sets for consistency and completeness we contribute to the subject. The concept of process driven authorization as an access control mechanism to achieve compliance to data privacy and other regulations has been introduced along with a WTL. Using an industry use case, the concept has been illustrated. Currently, we are working on theorem proofs and lemmas to make the concept more concrete.

Acknowledgements This research has been sponsored by EU H2020 FIRST project (Grant No. 734599, FIRST: vF Interoperation supportIng buSiness innovaTion) and National Key R&D Program of China (2017YFE0118700).

References

1. E. Bertino, C. Bettini, E. Ferrari, P. Samarati, An access control model supporting periodicity constraints and temporal reasoning. *ACM Trans. Database Syst.* **23**(3), 231 (1998)
2. E. Bertino, E. Ferrari, V. Atluri, The specification and enforcement of authorization constraints in workflow management systems. *ACM Trans. Inf. Syst. Secur.* **2**(1), 65–104 (1999)
3. G. Karjoth, Aligning security and business objectives for process-aware information systems, in *Proceedings 5th ACM Conference Data Applied Security Privacy—CODASPY'15* (2015) pp. 243–243
4. S. Sadiq, G. Governatori, Managing regulatory compliance in business processes. *Handb. Bus. Process Manag.* **2**, 159–175 (2010)
5. K. Tan, J. Crampton, C.A. Gunter, The consistency of task-based authorization constraints in workflow systems, in *Proceedings 17th IEEE Computer Security Foundations Workshop*, (2004) pp. 155–169
6. J.P. Kasse, L. Xu, P.T. de Vriese, The need for compliance verification in collaborative business processes (2018)

7. O.M.G. Omg, Business Process Model and Notation (BPMN) Version 2.0, in *Business*, vol. 50 (2011), p. 170
8. M. Salnitri, F. Dalpiaz, P. Giorgini, Modeling and verifying security policies in business processes, in *Lecture Notes in Business Information Processing*, vol. 175 (LNBIIP, 2014), pp. 200–214
9. G. Monakova, A.D. Brucker, A. Schaad, Security and safety of assets in business processes, in *Proceedings of the 27th Annual ACM Symposium on Applied Computing—SAC'12* (2012) p. 1667
10. J. Müller, Security mechanisms for workflows in service-oriented architectures (2015)
11. G. Koliadis, Verifying semantic business process models in inter-operation, in *IEEE International Conference on Services Computing* (2007)
12. J.P. Kasse, L. Xu, P. de Vrieze, A comparative assessment of collaborative business process verification approaches, vol. 506 (2017)
13. D. Basin, E.T.H. Zurich, Optimal workflow-aware authorizations, in *Proceedings of the 17th ACM Symposium Access Control Models and Technologies ACM* (2011) pp. 93–102
14. A.M. Awad, A Compliance Management Framework for Business Process Models. Ph.D. thesis (2010)
15. D. Nikovski, B. Akihiro, Workflow trees for representation and mining of implicitly concurrent business processes, in *ICEIS 2008—Proceedings of the 10th International Conference on Enterprise Information Systems (ISAS)*, vol. 2 (2008), pp. 30–36
16. J. Crampton, G. Gutin, Constraint expressions and workflow satisfiability, in *Proceedings of the 18th ACM Symposium Access Control Models and Technologies ACM* (2013), pp. 73–84
17. D.R. dos Santos, S.E. Ponta, S. Ranise, Modular synthesis of enforcement mechanisms for the workflow satisfiability problem, in *Proceedings of the 21st ACM Symposium Access Control Models and Technologies—SACMAT'16* (2016), pp. 89–99
18. M.C. Mont, R. Thyne, Privacy policy enforcement in enterprises with identity management solutions. *J. Comput. Secur.* **16**(2), 133–163 (2008)
19. M.C. Mont, R. Thyne, A systemic approach to automate privacy policy enforcement in enterprises, in *International Workshop on Privacy Enhancing Technologies* (2006), pp. 118–134

Comprehensive Exploration of Game Reviews Extraction and Opinion Mining Using NLP Techniques



Stefan Ruseti, Maria-Dorinela Sirbu, Mihnea Andrei Calin, Mihai Dascalu, Stefan Trausan-Matu and Gheorghe Militaru

Abstract Sentiment analysis and opinion summarization have become an important research area with the increase of available data on the Web. Since the Internet started containing more and more opinions and reviews for different products, individual users and companies saw the benefits of a priori evaluations based on other users' experiences; thus, automated analyses centered on customer impressions and experiences emerged as crucial marketing instruments. Our aim is to create a scalable and easily extensible pipeline for building a custom-tailored sentiment analysis model for a specific domain. A corpus of around 200,000 games reviews was extracted, and three state-of-the-art models (i.e., support vector machines, multinomial Naïve-Bayes, and deep neural network) were employed in order to classify the reviews into positive, neutral, and negative. Current results surpass previous experiments based on word counts applied on a similar game reviews dataset, thus arguing for the adequacy of the proposed workflow.

Keywords Sentiment analysis and opinion mining · Game reviews · Natural language processing

S. Ruseti · M.-D. Sirbu · M. A. Calin · M. Dascalu (✉) · S. Trausan-Matu · G. Militaru
University Politehnica of Bucharest, Splaiul Independenței 313, 60042 Bucharest, Romania
e-mail: mihai.dascalu@cs.pub.ro

S. Ruseti
e-mail: stefan.ruseti@cs.pub.ro

M.-D. Sirbu
e-mail: maria.sirbu@cti.pub.ro

M. A. Calin
e-mail: mihnea.calin@gmail.com

S. Trausan-Matu
e-mail: stefan.trausan@cs.pub.ro

G. Militaru
e-mail: gheorghe.militaru@upb.ro

M. Dascalu · S. Trausan-Matu
Academy of Romanian Scientists, Splaiul Independenței 54, 050094 Bucharest, Romania

1 Introduction

Internet has become a common practice and this trend leads to an increased influence in marketing and buying decisions. For example, online comments play a major role in the popularity of a game, movie, or any kind of media product. Thus, the popularity of a brand increases or decreases depending on the thoughts expressed by different people. Gaming is one of the most thriving industries in which players are highly influenced by user reviews. As game rankings are mostly based on aggregated user or critic scores, it is essential to extract and consider the actual value of individual reviews. Full-text reviews provide important benefits for users who can make informed decisions, as well as for game companies which can use online marketing mechanisms to take appropriate decisions (e.g., promotions for low evaluated games).

A sentiment refers to attitudes, emotions, and opinions conveyed with regards to a given entity. Multiple points of views or ideas can be found just by analyzing online reviews as these originate from different social groups, genders, professions; therefore, this provides a better overview of the perceived impressions. Sentiment analysis (or opinion mining) focuses on the task of extracting human feelings and opinions from texts written in natural language [1, 2]. Sentiment analysis models are widely employed in different domains including marketing, business, education, sociology, psychology, and economics [2]. These automated models frequently use natural language processing (NLP), information retrieval, and data mining [3] techniques. An important factor for the increased efficiency of machine learning and NLP methods resides in the huge amount of information available online for training purposes [4]. In addition, we can perceive sentiment analysis as a simplification of in-depth discourse analysis and semantics as our aim is to automatically extract global features (e.g., positive or negative sentiments and their corresponding labels) [5].

Sentiment analysis methods entail various processing steps, out of which two are frequently encountered: text preprocessing, followed by the automated classification. The first stage uses NLP techniques such as: tokenization, stop-words, numbers and punctuation removal, and lemmatization. Second, a wide range of classifiers can be employed consisting of two major categories—machine-learning techniques and lexicon-based statistics—as well as a combination of them (i.e., hybrid methods). Pang and Lee [6] envision an optimal solution for opinion-mining as a machine which processes the text for a given item, creates a list of product features, and aggregates all opinions for the given entity.

The aim of this paper is to compare multiple state-of-the-art models capable of classifying game reviews as positive, negative, or neutral. The next section describes frequently employed methods for opinion-mining and sentiment analysis. The third section introduces the used corpora (consisting of around 200,000 game reviews gathered for over 3000 games), followed by a description of the proposed methods. The fourth section presents the results and a comparison of our selected models, followed by conclusions and the future work.

2 State of the Art

Two major NLP-based methods for sentiment analysis compete, for which representative models are presented in this section: lexicons and machine-learning techniques, each with its own limitations. Machine-learning techniques require a large dataset of human examples for training, which is opposite to lexicon-based approaches that do not capture context-sensitive semantics because they rely on isolated word occurrences [7]. Moreover, context is important for expressing word meanings because some words can have multiple valences in different contexts. However, most approaches rely on the *bag-of-words* assumption in which word order is disregarded. Thus, the discourse structure is completely disregarded, and, for example, the following two phrases will have the same polarity, even though they express opposite positions: “*You are right, I do not like this ice cream*” versus “*You are not right, I like this ice cream.*”

Sentiment lexicons are lists of words which express polarities for different dimensions. These vectors contain information about semantic valences (e.g., intensification or negation) [8], potential parts of speech [9], and can be divided into two main categories: domain-independent word lists (i.e., general dictionaries that grant a global overview) and domain-dependent vectors (i.e., accurate for certain categories like books, tourism, shopping, gaming, or movies review, but potentially irrelevant for other domains). Multiple dictionaries and tools have been developed, out of which the most representative open-source ones are: Affective Norms for English Words (ANEW) [10], SenticNet, The General Inquirer (GI) [11] including the Lasswell [12] dictionary, Geneva Affect Label Coder (GALC) [13], and EmoLex [14]. In addition, several approaches build on top of individual word lists by combining them into meaningful components—for example, SEANCE [15] or the method introduced by Sirbu, Secui, Dascalu, Crossley, Ruseti, and Trausan-Matu [16].

Machine-learning approaches mostly rely on supervised-learning algorithms to classify opinions into positive and negative classes, such as: Naïve-Bayes [17], maximum entropy [18], multinomial logistic regression [19], support vector machines [20] or deep neural networks [20]. The most representative methods were selected for our pipeline and are presented in detail in the following section. Some NLP libraries, like Stanford CoreNLP [21], also include a sentiment analysis component. Socher, Perelygin, Wu, Chuang, Manning, Ng and Potts [22] use a recursive neural network applied on the constituency parsing tree of a sentence. Each node in the tree is labeled with a polarity score with seven possible values.

In addition, Kim, Ganesan, Sondhi, and Zhai [23] point out reasons why opinion-mining has not yet been used by major industries when developing products. Unfortunately, this subject is still in research and the solutions are not as accurate as possible. The accuracy of the solutions should incorporate better understanding of the text and make a more generalized solution so that it can be used in different fields. The downside of using more features to improve the accuracy means that scalability decreases. This can be mitigated by introducing parallel processing and streamlined workflows. Another potential problem is the lack of common and public

datasets. Most researches are performed on private data and other researches cannot have easy access to it. Moreover, besides standardization, quality control [24] needs to be enforced—only in this manner, valuable opinions are taken into consideration for researches and projects [25].

3 Method

3.1 Corpora

Our dataset consists of 201,552 games reviews crawled from Metacritic. Crawler4j [26] was used to extract reviews from over 3000 games. Games and reviews were indexed in Elasticsearch [27], whereas Kibana [28] was used to create interactive visualizations and statistics over our dataset. Figure 1 displays the number of reviews for each user rating.

In Fig. 2, we present a comparison between the users' score and critics' metascore as we want to analyze the accuracy of the user ratings. All scores were rounded to the closest integer in order to ease the follow-up classifications. We consider that the metascores are the best available reference for quantifying the subjectivity and objectivity within a user's opinion. The data shows that critics usually give higher scores compared to user reviews; however, there are no major differences in the assignments of games in specific intervals when using their corresponding metascores or the users' scores.

The initial dataset was split into three categories depending on the rating given by users, as follow: negative—reviews with rating between 1 and 5, neutral—reviews with rating between 6 and 8, and positive—reviews with rating 9 and 10. In order to have a balanced dataset, the same number of reviews was sampled from each



Fig. 1 Number of reviews per score

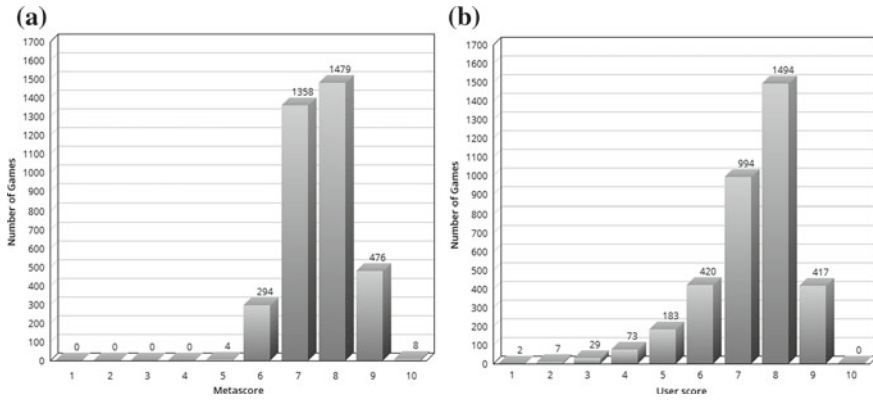


Fig. 2 Comparison between **a.** Metascores (critics) and **b.** Average user score of the considered games

Table 1 Classes and reviews number for sentiment prediction pipeline

	Negative reviews	Neutral reviews	Positive reviews
Training	33,000	33,000	33,000
Validation	3000	3000	3000
Test	3000	3000	3000

class. Table 1 contains the distribution of the three types of reviews in the training, validation, and test partitions.

3.2 Extensible Architecture

Figure 3 presents our extensible sentiment analysis pipeline. First, we extracted a large number of games reviews from Metacritic using Crawler4j [26]. Second, all games reviews were indexed into Elasticsearch [27]. Third, we applied a text preprocessing pipeline which includes: stop-words removing (stop-words like preposition,

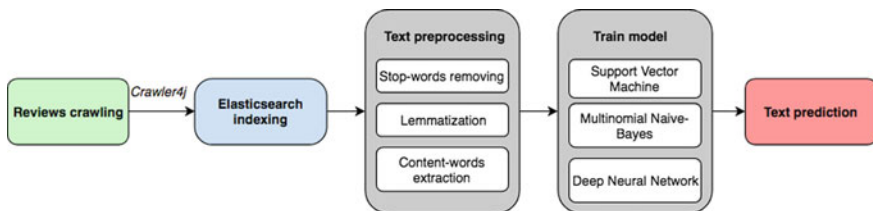


Fig. 3 Building a scalable sentiment analysis pipeline

conjunctions, pronoun, etc., are eliminated), lemmatization (bringing words to the base form, i.e., verbs are transformed to infinitive form), content-words extraction (all models were trained only on content-words because they express sentiments and contain valuable information).

In the last step, we evaluated three classifiers on the preprocessed texts: support vector machine, multinomial Naïve-Bayes, and deep neural networks (DNN). The three classifiers use the words as features and predict the score of the review. The hyper-parameters of each model were tuned using the validation partition and the reported results are obtained on the test partition by training on both training and validation sets.

In general, a neural network-based text classifier uses an encoder that computes a representation for the text followed by some fully connected layers that produce the probabilities for each output class. In this case, we used the deep averaging network (DAN) and transformer versions of the Universal Sentence Encoder [29] from TensorFlow Hub¹. These models were pre-trained on several text prediction tasks which require a general representation of the text and show good results when transferred on other tasks.

The SVM and multinomial Naïve-Bayes were applied on the bag-of-words representation of the text. We have also tested a Tf-Idf weighting of the words. The scikit-learn python machine-learning library² was used for preprocessing and model training.

4 Results

The accuracies obtained by each evaluated model are presented in Table 2. The DNN models performed better than the other classifiers, but did not achieve a significantly higher accuracy. One possible explanation for this limitation is the fact that the pre-trained encoders should capture the meaning of a text, which might not be very useful for sentiment analysis applied on a specific dataset.

Table 2 Accuracy of different models on the test set

Model	Representation	Accuracy (%)
SVM	BoW	66
SVM	Tf-Idf	61
Multinomial NB	BoW	65
Multinomial NB	Tf-Idf	65
DNN	DAN	66
DNN	Transformer	67

¹<https://www.tensorflow.org/hub/modules/google/universal-sentence-encoder/2>.

²<http://scikit-learn.org>.

The obtained results surpass previous lexicon-based analyses performed on a similar game reviews dataset that considered emerging PCA components [16, 30].

5 Conclusions and Future Developments

This paper analyses the accuracy of different classifiers for predicting the rating of game reviews extracted from Metacritic. Our goal was to create a scalable and easily extensible pipeline for building a custom-tailored sentiment analysis model for a specific domain.

In the future, we want to improve the training dataset and create a complete pipeline for Romanian language that can be employed onto different domains and industries, including books and film reviews. However, the first step is to extract a large collection of reviews written in Romanian. Moreover, we observed a lot of writing mistakes in reviews, which can potentially influence the accuracy of the classifiers. This is especially important in the case of bag-of-words inputs in which any wrongly spelled word acts as a completely different feature. One way to avoid this is to use an automated spell-checker in the preprocessing step. In addition to the previous classifiers, we will also consider additional potential methods (e.g., k -nearest neighbors, random forest) in order to create a strong baseline.

Additional experiments should be conducted on this dataset using other DNN architectures. The dataset is large enough to train an encoder on it, instead of using a pre-trained one. In a sentiment analysis task, the adjectives and adverbs from the text are more important than the nouns and verbs, which are usually central elements for computing text similarity. By training directly on this dataset, the encoder should be capable of learning what parts of speech to focus on.

Acknowledgements This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI—UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0689/“Lib2Life—Revitalizarea bibliotecilor si a patrimoniului cultural prin tehnologii avansate”/“Revitalizing Libraries and Cultural Heritage through Advanced Technologies”, within PNCDI III.

References

1. B. Liu, *Sentiment Analysis and Opinion Mining* (Morgan & Claypool Publishers, San Rafael, CA, 2012)
2. C.J. Hutto, E. Gilbert, Vader: a parsimonious rule-based model for sentiment analysis of social media text, in *8th International AAAI Conference on Weblogs and Social Media* (AAAI Press, Ann Arbor, MI, 2014), pp. 216–225
3. Z. Hailong, G. Wenyan, J. Bo, Machine learning and lexicon based methods for sentiment classification: a survey, in *2014 11th Web Information System and Application Conference (WISA)* (IEEE, 2014) pp. 262–265
4. B. Pang, L. Lee, Opinion mining and sentiment analysis (foundations and trends (R) in Information Retrieval). Now Publishers Inc. (2008)

5. B. Liu, Sentiment analysis and opinion mining. *Synth Lect. Hum. Lang. Technol.* **5**(1), 1–167 (2012)
6. B. Pang, L. Lee, Opinion mining and sentiment analysis. *Found. Trends Inf. Retrieval* **2**(1–2), 1–135 (2008)
7. O.K.M. Cheng, R.Y.K. Lau, Probabilistic language modelling for context-sensitive opinion mining. *Sci. J. Inf. Eng.* **5**(5), 150–154 (2015)
8. J.G. Shanahan, Y. Qu, J. Wiebe, *Computing Attitude and Affect in Text: Theory and applications*, vol. 20 (Springer, Berlin, 2006)
9. A. Hogenboom, F. Boon, F. Frasinca, *A statistical approach to star rating classification of sentiment*, Management Intelligent Systems (Springer, 2012), pp. 251–260
10. M.M. Bradley, P.J. Lang, *Affective Norms for English words (ANEW): Stimuli, Instruction Manual and Affective Ratings*, (The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, 1999)
11. P. Stone, D.C. Dunphy, M.S. Smith, D.M. Ogilvie, *Associates: The General Inquirer: A Computer Approach to Content Analysis* (The MIT Press, Cambridge, MA, 1966)
12. H.D. Lasswell, J.Z. Namenwirth, *The Lasswell Value Dictionary* (Yale University Press, New Haven, 1969)
13. K.R. Scherer, What are emotions? And how can they be measured? *Soc. Sci. Inf.* **44**(4), 695–729 (2005)
14. S.M. Mohammad, P.D. Turney, Crowdsourcing a word–emotion association lexicon. *Comput. Intell* **29**(3), 436–465 (2013)
15. S. Crossley, K. Kyle, D.S. McNamara, *Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis*. *Behavior Research Methods* (in press)
16. M.-D. Sirbu, A. Secui, M. Dascalu, S.A. Crossley, S. Ruseti, S. Trausan-Matu, Extracting gamers' opinions from reviews, in *18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2016)* (IEEE, Timisoara, Romania, 2016), pp. 227–232
17. A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in *LREC 2010* (Valletta, Malta, 2010)
18. A. Go, R. Bhayani, L. Huang, Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, vol. 1(2) (Stanford, 2009)
19. P. Melville, W. Gryc, R.D. Lawrence, Sentiment analysis of blogs by combining lexical knowledge with text classification, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2009), pp. 1275–1284
20. T. Mullen, N. Collier, Sentiment analysis using support vector machines with diverse information sources, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (2004)
21. C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S.J. Bethard, D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (ACL, Baltimore, MA, 2014), pp. 55–60
22. R. Socher, A. Perelygin, J.Y. Wu, J. Chuang, C.D. Manning, A.Y. Ng, C.P. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in *Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)* (ACL, Seattle, WA, 2013)
23. H.D. Kim, K. Ganesan, P. Sondhi, C. Zhai, Comprehensive Review of Opinion Summarization (2011)
24. B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, in *Mining Text Data* (Springer, 2012), pp. 415–463
25. L. Zhuang, F. Jing, X.-Y. Zhu Movie review mining and summarization, in *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (ACM, 2006), pp. 43–50
26. Y. Ganjisaffar, Crawler4j–Open Source Web Crawler for Java, Google Scholar (2012)

27. C. Gormley, Z. Tong, *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine* (O'Reilly Media, Inc. California, 2015)
28. Y. Gupta, *Kibana Essentials*, Packt Publishing Ltd (2015)
29. D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R.S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Universal Sentence Encoder. arXiv preprint (2018), [arXiv:1803.11175](https://arxiv.org/abs/1803.11175)
30. A. Secui, M.-D. Sirbu, M. Dascalu, S.A. Crossley, S. Ruseti, S. Trausan-Matu, Expressing sentiments in game reviews, in *17th International Conference on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2016)* (Springer, Varna, Bulgaria, 2016), pp. 352–355

MHAD: Multi-Human Action Dataset



Omar Elharrouss, Noor Almaadeed and Somaya Al-Maadeed

Abstract This paper presents a framework for a multi-action recognition method. In this framework, we introduce a new approach to detect and recognize the action of several persons within one scene. Also, considering the scarcity of related data, we provide a new data set involving many persons performing different actions in the same video. Our multi-action recognition method is based on a three-dimensional convolution neural network, and it involves a preprocessing phase to prepare the data to be recognized using the 3DCNN model. The new representation of data consists in extracting each person's sequence during its presence in the scene. Then, we analyze each sequence to detect the actions in it. The experimental results proved to be accurate, efficient, and robust in real-time multi-human action recognition.

Keywords Human action recognition · Multi-human action recognition · Convolutional neural network (CNN) · Video surveillance

1 Introduction

In the past twenty years, a big number of videos are captured by different types of cameras, including surveillance cameras, phone cameras, and filming crew cameras. Videos that recorded, uploaded, and transferred via Internet contain a massive amount of data to be analyzed [1], which in turn implies the need for systems and technics that can analyze the content. Analyzing these videos represents the main goal in all computer vision tasks. One of the most important tasks is the video summarization that aims to reduce the browsing time to find a particular information. Another important task in computer vision is the recognition of human action, which is the

O. Elharrouss (✉) · N. Almaadeed · S. Al-Maadeed
Department of Computer Science and Engineering, Qatar University, Doha, Qatar
e-mail: elharrouss.omar@gmail.com

N. Almaadeed
e-mail: n.alali@qu.edu.qa

S. Al-Maadeed
e-mail: S_alali@qu.edu.qa

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_28

first step for many other tasks like summarization [2]. Therefore, many datasets were created and collected to be used for testing different human action recognition approaches.

Existing human action datasets provide a massive number of videos that represent human actions in many situations. According to the purpose of acquisition or collection of these videos, datasets can be generally classified into two categories: surveillance datasets and Web-movie datasets. Surveillance datasets are captured by a fixed camera where the objects (human bodies) acting in a static scene. This category includes many datasets, such as KTH [3], Weisman [4], IXMAS [5], UCF-ARG [6], and PETS series [7–9]. In each dataset many actors with different situations and appearances are acting an action in the scene; herein, each video contains one action made by a single actor only. Because of the simplicity of the videos in terms of camera's position and illumination variation; the recognition rate is very high (up to 90%) for almost all methods. So, these datasets can be used for surveillance purposes to recognize human actions.

On the other hand, the Web-movie datasets contain videos from Internet like YouTube and movie clips that are captured by filming crew. Scenes in these kinds of videos can change with time and while cameras move. Under this category, we can find the following datasets: Hollywood [10], Hollywood2 [11], UCF sport [12], UCF50 [13], UCF101 [14], HMDB51 [15], HMDB [16], and ActivityNet [17]. These are called realistic action datasets because they contain real-life videos from YouTube and real movies. The diversity of clips content (scale of the actors, objects change positions, etc.), the variation in camera's motion and background obsolete analysis of this category of dataset. Table 1 presents all the aforementioned datasets in details.

Recently, many approaches of detecting and recognizing human action in video sequence have been proposed. The robustness of these methods depends upon the features used and the technologies implemented. Recent studies employ deep learning models to improve the performance of action recognition process [18]. Existing deep learning models, based on 2D convolutional networks, are extended into 3D domain to be suitable for temporal information [19]. Authors in [20] use the motion in consecutive frames to extract information for action recognition; they use two-stream ConvNet architecture that incorporates spatial and temporal networks. In the same context, Wang et al. [21] built up a spatiotemporal CNN model that recognizes multi-action using Motion History Image (MHI) as input for the deep learning model named MHI-CNN. Moreover, in [22], trajectory-pooled CNNs were designed by fusing Improved Trajectories into CNNs architecture, while adaptive recurrent convolutional hybrid networks were proposed consisting of a data module and a learning module. Other methods use infrared images with CNN architecture to recognize the human action [23]. The use of infrared images is a good feature to be used with a CNN because of the simplicity of information in images.

Surveillance systems record videos that are being analyzed after; however, many tasks require real-time analysis. Herein, human actions can provide important information for the monitoring systems. Thus, the recognition of human action provides a good assistance to these systems and to security agents to act in any expected event. In real scenes, human can do many actions during his presence in the scene

Table 1 Description of existing human activity recognition datasets

Category	Dataset	Background	Year	Action	#videos	Resolution
Surveillance	KTH [3]	Static	2004	6	2391	160 × 120
	PETS'04 [7]	Static	2004	–	28	384 × 288
	Weisman [4]	Static	2005	10	90	180 × 144
	IXMAS [5]	Static	2006	11	33	–
	PETS'07a [8] [9]	Static	2007	–	7	768 × 576
	UCF-ARG [6]	Static	2010	10	4320	960 × 540
Web-Movie	Hollywood [10]	Dynamic	2008	8	129	Vary
	Hollywood2 [11]	Dynamic	2009	12	3669	Vary
	UCF sport [12]	Dynamic	2009	9	200	720 × 480
	UCF50 [13]	Dynamic	2010	50	6676	Vary
	UCF101 [14]	Dynamic	2012	101	13,320	320 × 240
	HMDB51 [15]	Dynamic	2011	51	6849	320 × 240
	HMDB [16]	Dynamic	2011	51	7000	Vary
	ActivityNet [17]	Dynamic	2015	203	27,801	1280 × 720

monitored by a surveillance camera. Each person's actions can be recognized and summarized to be used where needed. Researchers working on human action recognition for videos surveillance systems found a lack of datasets that contain videos with many actors acting many actions in the same time. Therefore, in this work, we propose a new human action dataset that represents the common human action which can be found in surveillance videos. Also, we present a 3DCNN-based approach for multi-human action recognition.

2 Multi-Human Action Dataset (MHAD)

MHAD¹ attempts to provide a new dataset that contains many actions made by many actors in the same video, as the naming reflects. In one hand and related to video surveillance needs, each actor can act many actions during his presence in the scene. That can be a good data for many tasks in computer vision including video summarization methods based on human action, motion detection and tracking methods, people detection and recognition, and people counting. On the other hand, many persons can be found in the same video in action. Compared to the existing video surveillance dataset (that contains moving objects in the scene but with one

¹<https://drive.google.com/open?id=1pfnansy4VAejLRKNhCA8fn9IABarwz>.

Table 2 Characteristics of MHAD dataset

Actions	10
Actors	3–5
Number of videos	4
Duration of each video	2–3 min
Duration of each action	2–3 s

action like walking), our dataset provides many persons acting different actions in the same video.

The proposed dataset can help computer vision researchers, especially those working on video summarization, motion detection and tracking, real-time human action recognition, and many related tasks. By the following, we present the characteristics of the proposed dataset in details.

2.1 Dataset Characteristics

The proposed dataset includes a set of human actions representing usual human activities. MHAD composed of ten actions, including boxing, walking, running, hand waving, hand clapping, jogging, carrying, standing, backpack carrying, and two persons fighting.

Generated videos contain from three to five persons acting in the scene. In addition, three of the videos are outdoor videos and one is an indoor. The background is generated for each video, and annotations of each moving actor are provided. Table 2 illustrates the characteristics of the MHAD dataset. Also, Fig. 1 represents many frames from each video with also the background.

3 Proposed Method

To recognize multi-action at the same time in a video, the silhouette of human bodies should be detected and extracted to form a new sequence for each one of them. This new presentation of data consists of the extraction of sequence of each human body during his presence in the scene to form a new sequence. The new sequence generation based on the human body detection using motion detection method in [24], then tracking it using kernel-correlation-filter-based method in [25]. Figure 2 illustrates the sequence generation of each of existing persons in the scene. The background model of each video allows using it for detecting the moving objects [26, 27]. So, using the background subtraction technique, we form the sequence of detected person to be used in action recognition.

After that, with the proposed 3DCNN deep learning model, each action will be recognized using the new representation of video inputs. The 3DCNN-proposed



Fig. 1 MHAD dataset. First row: background images, rest rows: frames from the dataset

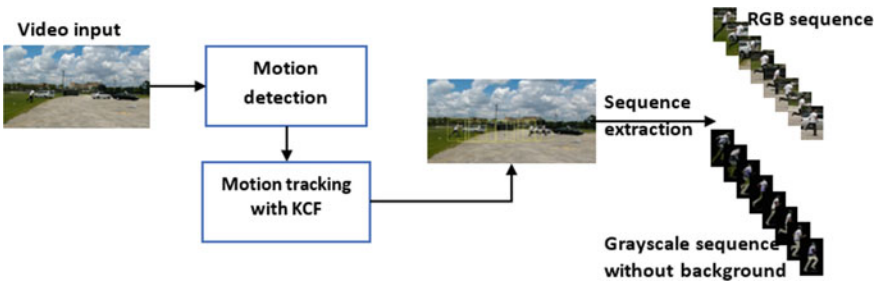


Fig. 2 Sequence extraction from a video from UCF-ARG dataset

architecture is an extended version of the CNN model presented in [23, 28]. We use the same model and transferred into three-dimensional-based model. After the recognition of each action, the result will be presented in the original video.

The architecture of our model, as illustrated in Fig. 3, composes of two 3D convolution-pooling units, with two convolutional layers and two MaxPooling layers, one flattened layer and two fully connected layers. The output layers comprise of ten neurons that represent the number of actions. We introduce 3D convolution as Conv (x,y,d,f) and pooling Mpool (x,y,d,k) where x and y are video dimension d the temporal depth, f number of channels, and k number of kernels. Using the notations, the proposed 3DCNN model can be described as follow:

conv(32, 32, 10, 1), Mpool(10, 10, 3, 64), conv(10, 10, 3, 64), Mpool(3, 3, 1, 64), flatten(576), FC(128).

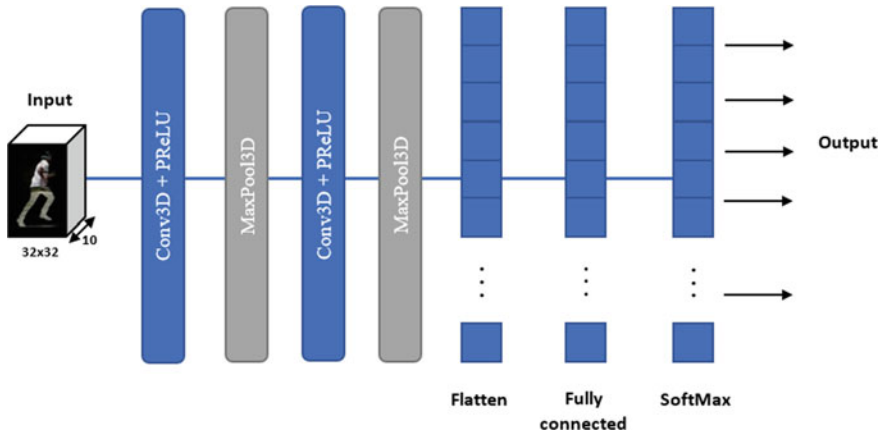


Fig. 3 Proposed 3D CNN architecture

The input of the system is a video of the background subtraction results with a resolution of 32×32 pixels with a temporal depth of 10. For the training and testing, we use the preprocessed data from KTH Weizmann and UCF-ARG datasets.

4 Experimental Results

Experimental results were obtained by employing the proposed datasets as well as the proposed 3DCNN methods for human action recognition.

Each video in our dataset contains four people acting many actions from the range of ten actions, mentioned earlier, during his presence in the scene. In each video, and using motion detection and tracking method, the annotations that represent the coordinates of each person in the scene are provided. The location of the acting person in the scene helps us to generate a sequence for each one, in order to be used for recognition. Here, we extracted two sequences for each actor, the first sequence contains RGB images, and the second sequence contains the gray scale of extracted images after subtracting the background. For the new representation, we can also extract binary videos and other features like Motion History Image MHI, HOG, HOF, etc.

The existing datasets, presented in Table 1, that can be classified into two categories: (1) Surveillance datasets that are captured by a fixed camera with a static background. (e.g., KTH, Weizmann, UCF-ARG, and IXMAX). (2) Datasets collected from Web like YouTube and from movies that are more complex.

The proposed system has been implemented with Python programming language using a laptop with GPU NVIDIA 1070 GTX. The evaluation has been made using the new representation of data of four dataset KTH, Weizmann, UCF-ARG, and our dataset MHAD. For the training phase, we use 80% of data where 10% of data is

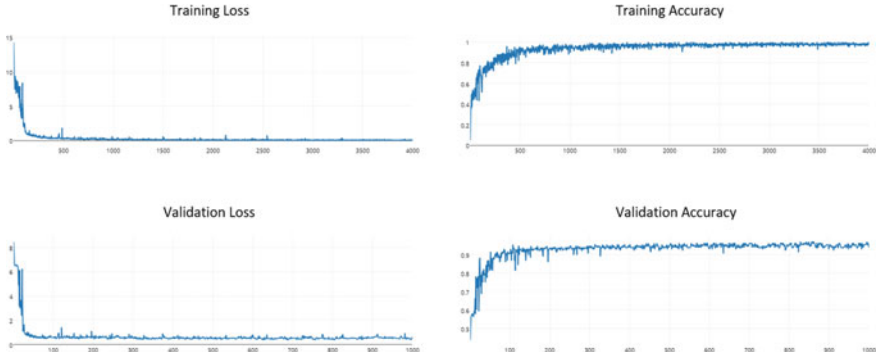


Fig. 4 Training and validation evaluation results

Table 3 Performance comparison with state-of-the-art-methods

Methods	Accuracy (%)
Jin et al. [21]	96 (KTH)
Akula et al. [23]	87.44
Proposed approach	93.75

for validation and 10% for testing. Using 1000 epochs, the accuracy of the proposed 3DCNN-based approach reaches 93.75% for validation and 98.44% for training. Figure 4 illustrates two graphs that represent training and validation lose and training and validation accuracy, respectively.

Compared to state-of-the-art-methods, that use the same datasets represented in Table 3, it can be observed that our proposed method results are improved and more effective. Obtained results are related to the use of the new representation of the data, in addition to the use of spatiotemporal features represented by 3DCNN which is more robust for human action recognition in video.

5 Conclusion

In this work, we provide a multi-human action dataset MHAD. Compared to the existing datasets, our dataset contains different actions made by many persons at the same time in the same video. The dataset allows a real-time human action recognition in public area where people acting simultaneously in the scene. Also, in this paper, we present a human action recognition method based on three-dimensional convolutional neural networks 3DCNN. With the proposed representation of data in preprocessing phase, which we performed prior to the recognition phase, the accuracy of action recognition with proposed architecture reaches 94% which represent a good recognition rate. Also, it can be robust for the real-time multi-human action recognition.

Acknowledgements This publication was made by NPRP Grant# NPRP8-140-2-065 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

1. M. Vrigkas, C. Nikou, I.A. Kakadiaris, A review of human activity recognition methods. *Front. Robot. AI* **2**, 28 (2015)
2. Z.S. Abdallah, M.M. Gaber, B. Srinivasan et al., Activity recognition with evolving data streams: A review. *ACM Comput. Surv. (CSUR)* **51**(4), 71 (2018)
3. C. Schuldt, I. Laptev, B. Caputo, Recognizing human actions: a local SVM approach, in *Proceedings International Conference on Pattern Recognition* (Cambridge, 2004), pp. 32–36
4. M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, in *Proceedings IEEE International Conference on Computer Vision* (Beijing, 2005), pp. 1395–1402
5. D. Weinland, E. Boyer, R. Ronfard, Action recognition from arbitrary views using 3D exemplars (ICCV, 2007)
6. A. Nagendran et al., New system performs persistent wide-area aerial surveillance. <http://spie.org/x41092.xml?ArticleID=x41092>
7. R.B. Fisher, PETS04 Surveillance Ground Truth Dataset (2004). Available at <http://www-prima.inrialpes.fr/PETS04/>
8. R. B. Fisher, Behave: Computer-assisted prescreening of video streams for unusual activities (2007). Available at <http://homepages.inf.ed.ac.uk/rbf/BEHAVE/>
9. R.B. Fisher, PETS07 Benchmark Dataset (2007). Available at <http://www.cvg.reading.ac.uk/PETS2007/data.html>
10. I. Laptev, M. Marszałek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in *CVPR* (2008)
11. M. Marszałek, I. Laptev, C. Schmid, Actions in context, in *CVPR* (2009)
12. M. Rodriguez, J. Ahmed, M. Shah, Action mach: A spatiotemporal maximum average correlation height filter for action recognition, in *CVPR* (2008), <http://server.cs.ucf.edu/vision/data.html>
13. J. Liu, J. Luo, M. Shah, Recognizing realistic actions from videos "in the wild", in *CVPR* (2009), <http://serre-lab.clps.brown.edu/resources/HMDB/>
14. H. Kuehne, H. Jhuang, E. Garrote, et al. HMDB: A large video database for human motion recognition, in *2011 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2011), pp. 2556–2563
15. F. C. Heilbron, V. Escorcia, B. Ghanem, J.C. Niebles, ActivityNet: A large-scale video benchmark for human activity understanding, in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Boston, MA, 2015), pp. 961–970
16. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F. Li, Large-scale video classification with convolutional neural networks, in *Proceedings on 2014 Computer Vision Pattern Recognition* (2014), pp. 1725–1732
17. H. Yang, C. Yuan, J. Xing, et al., Senn: Sequential convolutional neural network for human action recognition in videos, in *2017 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2017), pp. 355–359
18. K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in *Proceedings of Neural Information Processing System* (2014), pp. 568–576
19. H. Rahmani, A. Mian, M. Shah, Learning a deep model for human action recognition from novel viewpoints. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(3), 667–681

20. C. Feichtenhofer, A. Pinz, A. Zisserman, Convolutional two-stream network fusion for video action recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 1933–1941
21. C.B. Jin, S. Li, H. Kim, Real-time action detection in video surveillance using sub-action descriptor with multi-cnn. arXiv preprint (2017) <https://arxiv.org/abs/arXiv:1710.03383>
22. L. Wang, Y. Qiao, X. Tang, Action recognition with trajectory-pooled deep-convolutional descriptors, in *CVPR*, pp. 4305–4314 (2015)
23. A. Akula, A. K., Shah, R. Ghosh, Deep learning approach for human action recognition in infrared images. *Cogn. Syst. Res.* **50**, 146–154 (2018)
24. O. Elharrouss, A. Abbad, D. Moujahid, H. Tairi, Moving object detection zone using a block-based background model. *IET Comput. Vis.* **12**(1), 86–94 (2017)
25. J. F. Henriques, R. Caseiro, P. Martins et al., High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 583–596 (2015)
26. O. Elharrouss, A. Abbad, D. Moujahid, J. Riffi, H. Tairi, A block-based background model for moving object detection. *ELCVIA: Electronic Letters on Computer Vision and Image Analysis.* **15**(3), 0017–31 (2016)
27. O. ELHarrouss, D. Moujahid, S. E. Elkaitouni, H. Tairi, Moving objects detection based on thresholding operations for video surveillance systems, in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–5. IEEE (2015)
28. N. Almaadeed, O. Elharrouss, S. Al-Maadeed, A. Bouridane, A. Beghdadi, A novel approach for robust multi human action detection and recognition based on 3-Dimensional convolutional neural networks. arXiv preprint (2019) <https://arxiv.org/abs/1907.11272>

Machine Learning-Based Approaches for Location Based Dengue Prediction: Review



Chamalka Seneviratne Kalansuriya, Achala Chathuranga Aponso
and Artie Basukoski

Abstract Dengue is a fast-spreading viral disease which has no preventive medicine. Due to this infectious disease, almost half of the global population is at risk. Consequently, much research has been conducted using various medical as well as computational methods in order to prevent this menace. The main aim of this paper is to review machine learning approaches to this problem and to identify the most suitable method to predict the spread of this disease for distinctive geographical areas of countries like Sri Lanka. We consider environmental factors such as climate and vegetation data, dengue case data along with the population of a specific geographic area for the disease outbreak predictions. Specifically, this paper consists of the following sections: (i) A brief description of the disease and the factors affecting the spread; (ii) review the pattern of the environmental and population factors affecting the spread; (iii) a review and comparison of machine learning algorithms for prediction of the spread of the disease (SVM, decision tree, neural network, and random forest).

Keywords Dengue · Climate · Vegetation · Population · Dengue case data · Dengue endemic countries · Machine learning · SVM · Decision trees · Random forest · ANN

C. S. Kalansuriya (✉) · A. C. Aponso
Informatics Institute of Technology, Colombo, Sri Lanka
e-mail: chamalka246@gmail.com

A. C. Aponso
e-mail: achala.a@iit.ac.lk

A. Basukoski
University of Westminster, London, UK
e-mail: A.Basukoski@westminster.ac.uk

1 Introduction

Dengue is the most prevalent vector-borne viral disease in the world which can even cause deaths in its extreme form [1, 2]. This mosquito-borne disease is transmitted mainly through the bites of *Aedes aegypti* and *Aedes albopictus* mosquitoes with the serotypes of virus DENV-1, DENV-2, DENV-3, and DENV-4 [3]. According to the World Health Organization (WHO), the occurrence of dengue has increased in 30 folds when compared with previous 50 years by affecting almost 50 percent of global population [4, 5]. Today, it heralds an era which desperately needs control and prevention from dengue [6]. Due to the non-availability of prevention medicine, little success is obtained to control this menace; therefore, lots of researches are carried throughout the world in order to eradicate this life-threatening disease [4]. Dengue disease is prevalent in more than 120 countries of tropical and subtropical country areas. This disease takes place worldwide in countries like Australia, US-affiliated pacific islands, US Virgin Islands, Caribbean, and Asian countries [7]. Europe is also in threat of possible disease outbreak with the first transmission reported in 2010 in France [8]. Dengue has become a huge health burden to Asians including countries like Sri Lanka [5]. In 1989, a major outbreak was reported in Sri Lanka, and since then, epidemics are present regularly with the increase of dengue statistics [9]. World Health Organization stated that in Sri Lanka, the statistics for 2017 occurrence has increased in 4.3 folds compared to the average of dengue case statistics for the years from 2010 to 2016.

Dengue has become a hazard to Sri Lankans massively. Substantial proportion of Sri Lankan budget for health is utilized for dengue-associated expenditures and the management of crisis in Sri Lanka, but limited impact could be seen in the prevention purpose [10]. The only reliable method to prevent from this global burden of health is through vector control measures [11]. The vector causing the disease is the mosquitoes; therefore, mosquito eradication and control methods should be carried out to minimize the spread. Through this rapid spreading menace, lot of valuable lives and money are destroyed. Therefore, there is an urgent requirement for timely accurate early dengue outbreak detection for a specific area with the use of new computational methods in order to control this deadly disease.

Previous studies showed different factors which affect the disease spread.

2 Factors Affecting the Disease Spread

The epidemiology of dengue is inextricable from the ecology of vector [5]. The disease transmission and spread show interactions with climate, population, and mosquito disseminating virus with distinct patterns [5]. For the expansion of vectors by facilitating dengue spread, climate change along with different environmental changes are considered as the major factors [12]. The circulation of this burden of disease is not only influenced by climate factors but also through socio-environmental

factors like land use and population density [13]. Through identification of pattern of the factors affecting the disease spread of a specific area, the model can detect future dengue outbreaks.

2.1 Dengue Occurrence with Climate, Vegetation, and Population

Dengue has a significant correlation with the climate change [13]. In lots of studies, it is mentioned about the associations on dengue occurrences with climate factors [14, 15]. The temperature climate factor has a remarkable influence on the life cycle of the mosquitoes [15, 16]. The temperature increase and rainfall which result in climate variation may trigger outbreaks of dengue [15, 16]. According to studies, climate and vegetation variables are remarkably related to dengue incidence. Vegetation plays an important role in the explanation of dengue occurrences and also with the future outbreak location identification [17]. A research done in Sao Paulo mentioned that different vegetation coverage has different impacts on the dengue incidence [18].

Population is also one important factor affecting the disease spread [19]. The population growth is a key factor for the increase of the spread of this menace [20]. According to a study, human population is an important factor which gives better predictive power when modeling high-risk areas of dengue [21]. It is vital to emphasize the relation of environmental and population factors in order to minimize the number of dengue occurrence statistics [20, 21]. Dengue data is also important to identify the dengue occurrence in an area and also to detect the count and impact of dengue for the related area [13].

As discussed in this paper, dengue data along with the environmental and population data plays a crucial role in the disease spread and is important for the predictions. A prediction model with the identification of a distinct pattern of dengue for a specific area will be of immense capability and save lots of valuable human lives. Due to the severity of this global health issue, lots of research methods are carried out in the world to prevent and predict the disease spread [13]. With the advancement of technology, several computational model development attempts have been carried out globally for the disease prevention purpose [22]. In this paper, the machine learning approach for prediction is discussed with better evaluation.

3 Machine Learning Approach

Machine learning is a computational field which delivers the ability for computers to learn without explicit programming [23]. Machine learning has become omnipresent and crucial for resolving complicated problems of any sciences especially in the field

of medicine [24]. With the use of technical advancement, machine learning algorithms will provide the ability to predict, identify different diseases in medical field like experts by handling numerous complicated data [24]. Machine learning is considered as one of the best research methods for disease prediction. Different previous researches done using machine learning are included in this paper to emphasize its importance. Machine learning approach is used by some research done in UK for predicting cardiovascular risks, and the results show that it gives more accurate results from this approach compared to other methods [25]. Some researchers of USA and Canada did a research on the machine learning usage for heart failure predictions and classification [26]. Multiple researches with the application of machine learning for different disease prediction could be found [26–28]. Through these researches, the invaluable outcome from the machine learning-approached predictions to the health sector is emphasized. Research done using this machine learning approach could be seen in substantial amount for dengue also.

Different types of machine learning algorithms could be used for the disease prediction. Various types of dengue prediction researches carried out in dengue prevalent areas are critically evaluated in this research paper. In Table 1, various types of machine learning algorithm approaches are evaluated with the related work for the research.

In lots of dengue prediction research classification, machine learning approach comparisons are carried out to find the best approach [40]. In a research carried out for dengue prediction, different machine learning techniques like SVM, KNN, and decision tree method of random forests were used and compared [41]. First, the data was acquired then after processing and feature selection, the training algorithms were used for the predictions. In this research, the random forests method outperformed the other classification algorithms in the performance. It is emphasized that the random forests method produced best accurate results by using random decision trees with the discovering of other effective features which would else be overshadowed through the prominent features [41]. Decision tree algorithm is an algorithm which has a hierarchical flow with a tree-like structure where nodes, branches, and leaves present the features, rules, and outcomes [42]. A random forest is a collection of decision trees. A research for dengue early detection through machine learning approach used different algorithms to detect the disease early [40]. In this research, with high classification accuracy, the method of the random forests is selected due to the superior performance comparatively to the others [40].

This prediction methodology with the use of random forests of decision trees gives solutions to the dimensionality issues which are present in ANN and KNN algorithms. The ANN and SVM algorithms are complex, comparatively decision tree methods are easy to understand and interpret. Through the reviewed research cases, the Random forests with decision trees provide high accuracy with better visual understanding of pattern identifications, provide priority to each contributory factor, and also superior predictions are given. Different types of supervised classification algorithms of machine learning approach are discussed in Table 2—algorithm analysis with their pros and cons. Many studies emphasize the accuracy of the prediction of the disease is dependent on the data set and the mechanisms [29]. These algorithms

Table 1 Existing systems

System (name)	Variables used and area researched	Description	Evaluation
Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico-2018 [29]	<ol style="list-style-type: none"> 1. Population size 2. Climate Data <ul style="list-style-type: none"> • Area-Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico(USA) 	Successful results were obtained by using ANN with genetic algorithms to predict dengue cases for the research area. The predictive power was above 70% [29]	ANN is highly versatile and deals well with nonlinear data [30]
Kernel-Based Machine Learning Models for the Prediction of Dengue and Chikungunya Morbidity in Colombia-2017 [31]	<ol style="list-style-type: none"> 1. Dengue Case Data • Area-Colombia 	Kernel Ridge Regression and Gaussian Processes were used to predict the future outbreaks [31]	Gaussian kernel outperforms for linear regression but kernel ridge regression does not have a simple interpretation though it achieves more accuracy [32]
Prediction of Dengue Cases in Paraguay Using Artificial Neural Networks-2017 [30]	<ol style="list-style-type: none"> 1. Climate data 2. Dengue case data 3. River level <ul style="list-style-type: none"> • Area-Paraguay 	Use Artificial Neural Networks(ANN) along with regression method comparisons [30]	Have the ability of accuracy maintenance when some data are missing ANN also produces acceptable results when data are noisy and incomplete [33]
Using C-support Vector Classification to Forecast Dengue Fever Epidemics [34]	<ol style="list-style-type: none"> 1. Climate Data 2. Dengue Fever case data <ul style="list-style-type: none"> • Area-Taiwan 	Uses classification of object in data set based on C-SVM kernel with the use of RBF and linear kernels, grid search method is used for the selection of hyper parameters [34]	Good results were obtained from SVM but according to result comparison of some researches more accurate outcomes could be obtained through other methods also. [35]
Classification Rules Using Decision Tree for Dengue Disease [36]	<ol style="list-style-type: none"> 1. Climate Data 2. Dengue Case Data <ul style="list-style-type: none"> • Area- India 	Uses unsupervised clustering as well as supervised model of prediction is achieved through decision tree classification model [36]	Overall performance can be considered good possession of the suitability in order to discover rules in data mining [37]
Analysis of Significant Factors for Dengue Infection Prognosis Using the Random Forest Classifier [38]	<ol style="list-style-type: none"> 1. Dengue Case Data <ul style="list-style-type: none"> • Area- India 	With the use of RF classification tree to determine and visualize factors which are significant to identify the dengue patients in order to improve stability and accuracy [38]	High accurate results could be obtained from Random Forest compared to other techniques of statistical models [39]

Table 2 Algorithm analysis

Name of algorithm	Brief description	Benefits	Drawbacks
Support Vector Machine (SVM)	Classifier method which defines decision boundaries through the decision plane concept	<ul style="list-style-type: none"> • Efficiency of training and generalization • Solving problems in nonlinear classification and regression • Can be used when limited number of data is available • With the mapping data heightens the accuracy and efficiency of analysis [43] 	<ul style="list-style-type: none"> • Parameter selection is quite difficult • When the input training data is huge may take long running time. [43]
Artificial Neural Networks(ANN)	Computational algorithm which is based on functions and structure of neural network of biology. Works in the manner of human brain [43]	<ul style="list-style-type: none"> • Capability in extraction of nonlinear and complex relationships in systems • Reliability when huge number of data is input for training • Higher accuracy • High tolerance to data which are noisy [28, 43] 	<ul style="list-style-type: none"> • Give reliable results only when massive training data is taken as input • Over fitting [43]
Decision trees	A decision tree is a tree-like structure [44]	<ul style="list-style-type: none"> • Easiness to understand • Ability to handle nominal and categorical data • Higher Accuracy [45] 	<ul style="list-style-type: none"> • Over sensitivity to training set [44]
Random forests	A classifier with a collection of tree-structured classifiers	<ul style="list-style-type: none"> • Higher accuracy • Rather robust to outliers and noise • Easiness to interpret and understand [46] 	<ul style="list-style-type: none"> • Can be subjected to over fitting and under fitting specially when the data set is small [46]

function in pros and cons according to different situations and variable effect with data. The data collection accuracy as well as the data set will have a huge impact on the final output accuracy of the classification [28]. The best classification algorithm methodology selection through comparisons and with the use of accurate data sets will give the best solution model for these location-based dengue predictions.

4 Conclusion

Dengue global health burden with the factors affecting the dengue spread is critically evaluated in this paper. Then, the relationship of the factors affecting the disease spread with the importance of identification of different patterns of those factors with dengue incidence is included. Different computational approaches are used for the disease predictions. In this paper, different researches carried out for disease prediction with the use of computational method of machine learning are critically evaluated. Finally, machine learning-approached dengue research models with different classification algorithms with their pros and cons which will be useful for the research are mentioned. One main thing which emphasizes on the functionality and accuracy of these models is the data set. Data is collected manually, and human error could be a major limitation on accuracy of the model. With the use of proper data set along with the most suitable machine learning approach, accurate and efficient location-based dengue prediction models can be developed. In this review, the decision tree methods and random forests of machine learning approach are emphasized due to the superior performance. The contributory factors affecting the diseases change from the location to location so with best accurate approach, variable selections, and accurate data sets will provide best location-based models. These types of location-based predictive models will be really useful for dengue-endemic countries like Sri Lanka which is in huge threat due to dengue occurrences. These models will help the government, public, and the decision makers of health sector to prevent and minimize future dengue outbreaks. Through that millions of valuable human lives could be saved as well as the money utilized for disease treatment could also be saved.

References

1. WHO | Dengue (WHO, 2018)
2. I.A. Rather, H.A. Parray, J.B. Lone, W.K. Paek, J. Lim, V.K. Bajpai, Y.-H. Park, Prevention and control strategies to counter dengue virus infection. *Front. Cell. Infect. Microbiol.* **7**, 336 (2017)
3. L.B. Carrington, C.P. Simmons, Human to mosquito transmission of dengue viruses. *Front. Immunol.* **5**, 290 (2014)
4. Defeating dengue with GM mosquitoes [Online], University of Oxford, (2016). Available: <http://www.ox.ac.uk/research/research-impact/defeating-dengue-gm-mosquitoes>.

Accessed 29 Oct 2018

5. M.C. Castro, M.E. Wilson, D.E. Bloom, Program on the Global Demography of Aging at Harvard University. Working Paper Series. Disease and economic burdens of dengue Series. Dengue 1, Disease and economic burdens of dengue (2017)
6. T. Pang, T.K. Mak, D.J. Gubler, Prevention and control of dengue—the light at the end of the tunnel. *Lancet. Infect. Dis.* **17**(3), e79–e87 (2017)
7. S.H.W. Tyler, M. Sharp, J. Perez-Padilla, Dengue—Chapter 3—2018 Yellow Book| Travelers' Health| CDC [Online] (2018). Available: <https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/dengue>. Accessed 18 Dec 2018
8. GlobalData Healthcare, Dengue in Europe: is there an outbreak threat in new areas? (2018) [Online]. Available <https://www.hospitalmanagement.net/comment/dengue-in-europe/>. Accessed 18 Dec 2018
9. G.N. Malavige, S. Fernando, D.J. Fernando, S.L. Seneviratne, Dengue viral infections. *Postgrad. Med. J.* **80**(948), 588–601 (2004)
10. M.C. Weerasinghe, D.S. Sinha, It's time to review strategies for dengue in Sri Lanka. *BMJ* (2017). [Online]. Available <https://blogs.bmj.com/bmj/2017/08/06/time-to-review-strategies-for-dengue-in-sri-lanka/>. Accessed 30 Oct 2018
11. N.L. Achee, F. Gould, T.A. Perkins, R.C. Reiner, A.C. Morrison, S.A. Ritchie, D.J. Gubler, R. Teysou, T.W. Scott, A critical assessment of vector control for dengue prevention. *PLoS Negl. Trop. Dis.* **9**(5), e0003655 (2015)
12. P. Sirisena, F. Noordeen, H. Kurukulasuriya, T.A. Romesh, L. Fernando, Effect of Climatic Factors and Population Density on the Distribution of Dengue in Sri Lanka: A GIS Based Evaluation for Prediction of Outbreaks. *PLoS One* **12**(1), e0166806 (2017)
13. D.N. Pham, S. Nellis, A.A. Sadanand, A. Jamil, J.J. Khoo, A literature review of methods for dengue outbreak prediction, in *eKNOW 2016 : The Eighth International Conference on Information, Process, and Knowledge Management* no. c (2016), pp. 7–13
14. S. Díaz-Castro, M. Moreno-Legorreta, A. Ortega-Rubio, V. Serrano-Pinto, Relation between dengue and climate trends in the Northwest of Mexico. *Trop. Biomed.* **34**(1), 157–165 (2017)
15. H. Lee, J.E. Kim, S. Lee, C.H. Lee, Potential effects of climate change on dengue transmission dynamics in Korea. *PLoS ONE* **13**(6), e0199205 (2018)
16. Y.-H. Lai, The climatic factors affecting dengue fever outbreaks in southern Taiwan: an application of symbolic data analysis. *Biomed. Eng. Online* **17**(S2), 148 (2018)
17. A. Stanforth, M.J. Moreno-Madriñán, J. Ashby, N. El-Sheimy, Z. Lari, A. Moussa, D.R. Mishra, D.G. Goodin, X. Li, P.S. Thenkabail, Remote sensing exploratory analysis of dengue fever niche variables within the río magdalena watershed (2016)
18. R.V. Araujo, M.R. Albertini, A.L. Costa-da-Silva, L. Suesdek, N.C.S. Franceschi, N.M. Bastos, G. Katz, V.A. Cardoso, B.C. Castro, M.L. Capurro, V.L.A.C. Allegro, São Paulo urban heat islands have a higher incidence of dengue than other urban areas. *Brazilian J. Infect. Dis.* **19**(2), 146–155 (2015)
19. C.J. Struchiner, J. Rocklöv, A. Wilder-Smith, E. Massad, Increasing dengue incidence in Singapore over the Past 40 Years: Population growth, climate and mobility. *PLoS ONE* **10**(8), e0136286 (2015)
20. A. Wilder-Smith, D.J. Gubler, S.C. Weaver, T.P. Monath, D.L. Heymann, T.W. Scott, Epidemic arboviral diseases: priorities for research and public health. *Lancet. Infect. Dis.* **17**(3), e101–e106 (2017)
21. J.F. Obenauer, T. Andrew Joyner, J.B. Harris, The importance of human population characteristics in modeling *Aedes aegypti* distributions and assessing risk of mosquito-borne infectious diseases. *Trop. Med. Health* **45**, 38 (2017)
22. P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, G. Luo, Z. Li, J. He, Y. Zhang, W. Ma, Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl. Trop. Dis.* **11**(10), e0005973 (2017)
23. V.K. Damodar Reddy Edla, P. Lingras, *Advances in Machine Learning and Data Science*, vol. 705 (2018)
24. Z. Obermeyer, E.J. Emanuel, Fast-track zika vaccine development. *N. Engl. J. Med.* **375** (2016)

25. S.F. Weng, J. Repts, J. Kai, J.M. Garibaldi, N. Qureshi, Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE* **12**(4), e0174944 (2017)
26. P.C. Austin, J.V. Tu, J.E. Ho, D. Levy, D.S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. *J. Clin. Epidemiol.* **66**(4), 398–407 (2013)
27. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015)
28. A.A. Annakutty, A.C. Aponso, Review of brain imaging techniques, feature extraction and classification algorithms to identify alzheimer's disease. *Int. J. Pharma Med. Biol. Sci.* **5**(3), 178–183 (2016)
29. A. Laureano-Rosario, A. Duncan, P. Mendez-Lazaro, J. Garcia-Rejon, S. Gomez-Carro, J. Farfan-Ale, D. Savic, F. Muller-Karger, Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of Yucatan, Mexico and San Juan, Puerto Rico. *Trop. Med. Infect. Dis.* **3**(1), 5 (2018)
30. V. Ughelli, Y. Lisnichuk, J. Paciello, J. Pane, Prediction of dengue cases in paraguay using artificial neural networks, in *3rd International Conference on Health Informatics Medical Systems* (2017), pp. 130–136
31. W. Caicedo-Torres, D. Montes-Grajales, W. Miranda-Castro, M. Fennix-Agudelo, N. Agudelo-Herrera, Kernel-based machine learning models for the prediction of dengue and chikungunya morbidity in Colombia. *Commun. Comput. Inf. Sci.* **735**, 472–484 (2017)
32. N. Méndez, M. Oviedo-Pastrana, S. Mattar, I. Caicedo-Castro, G. Arrieta, Zika virus disease, microcephaly and Guillain-Barré syndrome in Colombia: Epidemiological situation during 21 months of the Zika virus outbreak, 2015–2017 (2015)
33. B. Jongmuenwai, S. Lowanichchai, S. Jabjone, Prediction Model of Dengue Hemorrhagic Fever Outbreak using Artificial Neural Networks in Northeast of Thailand (2018)
34. D. Rahmawati, Y.P. Huang, Using C-support vector classification to forecast dengue fever epidemics in Taiwan, in *2016 IEEE International Conference on System Science and Engineering (ICSSE)*, (2016)
35. G. Zhu, J. Hunter, Y. Jiang, Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy, in *Proceedings on 2016 IEEE International Conference Internet Things (iThings); IEEE Green Computing and Communications (GreenCom); IEEE Cyber, Physical and Social Computing CPSCom; IEEE Smart Data Smart Data* (2016), pp. 828–832
36. N.K.K. Rao, G.P.S. Varma, D. Rao, P. Cse, Classification rules using decision tree for dengue disease. *Int. J. Res. Comput. Commun. Technol.* **3**(3), 2278–5841 (2014)
37. R. Babu, Decision tree model for dengue data analysis, *Int. J. Res. Sci. Comput. Eng.* **3**(1) (2017)
38. A.S. Fathima, Analysis of significant factors for dengue infection prognosis using the random forest classifier. *Int. J. Adv. Comput. Sci. Appl.* **6**(2), 240–245 (2015)
39. T.M. Carvajal, K.M. Viacrusis, L.F.T. Hernandez, H.T. Ho, D.M. Amalin, K. Watanabe, Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infect. Dis.* **18**(1), 183 (2018)
40. N. Rajathi, S. Kanagaraj, R. Brahmanambika, K. Manjubarkavi, Early detection of dengue using machine learning algorithms
41. A. Macrae, C. Schiano De Colella, E. Sebastian, CS229 project: classification of dengue fever outcomes from early transcriptional patterns
42. I. Jenhani, N. Ben Amor, Z. Elouedi, Decision trees as possibilistic classifiers. *Int. J. Approx. Reason.* **48**(3), 784–807 (2008)
43. R. Gholami, N. Fakhari, Support vector machine: principles, parameters, and applications, in *Handbook of Neural Computation* (Elsevier, 2017), pp. 515–535
44. C. Petri, *Decision Trees* (2010)

45. L. Tanner, M. Schreiber, J.G.H. Low, A. Ong, T. Tolfvenstam, Y.L. Lai, L.C. Ng, Y.S. Leo, L. Thi Puong, S.G. Vasudevan, C.P. Simmons, M.L. Hibberd, E.E. Ooi, Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl. Trop. Dis.* **2**(3), e196 (2008)
46. L. Breiman, *Machine Learning* (Kluwer Academic Publishers, Dordrecht, 2001)

Critical Evaluation of Different Biomarkers and Machine-Learning-Based Approaches to Identify Dementia Disease in Early Stages



Gayakshika Gimhani, Achala Chathuranga Aponso
and Naomi Krishnarajah

Abstract Dementia is the loss of cognitive functioning and behavioural abilities to some extent. This is a major neurocognitive disorder which is a group of symptoms caused by other different conditions. Alzheimer's disease has considered as the most common type of dementia. Apart from that, vascular dementia (VD), Lewy body dementia (DLB), frontotemporal dementia (FTD), Parkinson's disease dementia, normal pressure hydrocephalus (NPH), Creutzfeldt–Jakob disease and syphilis are under Dementia. The cure for this disease is yet to be found, hence recognizing the disease in early stages and delaying the progress is a highly important fact. So, the investigation of this disease will remain as an open challenge. The aim of this paper is to review biomarkers and selected machine-learning techniques that can be segregated into early detection of dementia. Various machine-learning techniques such as artificial neural networks, decision trees and support vector machine will be discussed in this paper to find a better approach to identify Dementia in early stages. Especially this paper is consisting of following sections: (i) A brief description of Dementia and each type and the global statistics; (ii) A review of various type of medical techniques to identify dementia (MRI, CT, SPECT, fMRI, PET, EEG and CSF); (iii) Pre-processing signals; (iv) A review of machine-learning techniques.

Keywords Machine learning · Dementia detection · Alzheimer's disease · Vascular dementia (VD) · Lewy body dementia (DLB) · Frontotemporal dementia (FTD) · Normal pressure hydrocephalus (NPH) · Parkinson's disease dementia · Syphilis and creutzfeldt–jakob disease · Artificial neural networks · Decision trees · Support vector machine · MRI · CT · SPECT · FMRI · PET · EEG · CSF

G. Gimhani (✉) · A. C. Aponso · N. Krishnarajah
Informatics Institute of Technology, University of Westminister, Colombo, Sri Lanka
e-mail: kmggimhani@gmail.com

A. C. Aponso
e-mail: achala.a@iit.ac.lk

N. Krishnarajah
e-mail: naomi.kr@iit.ac.lk

1 Introduction

Among older people worldwide, Dementia is one of the major causes of dependency and disability. It not only is an impact on people with dementia but also to the society; mentally, physically, socially and economically. But this is not a normal part of ageing, though it mainly affects older people. According to the “Dementia fact sheet December 2017” of World Health Organization [WHO] [1] globally, there are around 50 million people with dementia and nearly ten million new cases reported every year. Estimated proportion of the general population with dementia is between five and eight per 100 people. Also, forecasted figures note that the number of demented people will mark a 204% increase from 2018 to 2050, which shall be 50 million in 2018 increased to 82 million in 2030 and to 152 million in 2050. World’s Alzheimer’s Report 2015 [2] states that East Asia is the region with the most people living with dementia (9.8 m). As per the “Dementia UK: Second edition” [3], the evaluated total price of dementia in UK is £26.3 billion. It should be noted that one out of every three people born in 2015 will develop Dementia during their lifetime [4]. Nevertheless, there are 209,600 numbers of new cases per year in UK and 74,000 of men and the rest of 135,000 are women [5]. It has been found that Dementia as the first cause of death of women and third of men in the UK [6]. A detailed report of the survey can be found in [7].

The most common type of dementia is Alzheimer’s disease (AD). It accounts 60–80% of cases [8]. Most rarely this can affect who are under age 65 and it is called “early-onset Alzheimer’s”. This happens due to the changing in the brain by building up two proteins called amyloid and tau. Although the researchers have not confirmed what triggers Alzheimer’s, they suggest that these both proteins are involved in progressing the disease and when it does, more nerve cells become damaged and then it leads to the symptoms of the Alzheimer’s [9]. The second most common type of dementia is vascular dementia (VD) [8]. This occurs due to the damages of blood vessels in the brain which has the body’s richest network of blood vessels. So, the inadequate blood flow can be damaged and eventually it kills cells anywhere in the body. This can cause memory and thinking problems [9]. Mild stage of this is called vascular cognitive impairment (VCI) and vascular brain changes often coexist with the changes to the other types of dementia [8]. Dementia with Lewy bodies (DLB) is the third most common type of dementia is [8]. This is caused by small round clumps called Lewy bodies of proteins such as alpha-synuclein which damage the way nerve cells work and communicate and it affects the thinking, memory and movement. This is also due to the movement problems in Parkinson’s disease and the changes these show are typical of Alzheimer’s which makes hard to distinguish DLB from Alzheimer’s [9]. Clinical diagnosis is using for DLB and it diagnoses when dementia symptoms appear within one year after movement symptoms. As DLB, Parkinson’s dementia has plaques and tangles. This affects movements first and then gradually affects mental function such as memory and the ability to pay attention [8]. Frontotemporal dementia (FTD) is caused due to the brain cell damaged by building up of proteins such as tau, TDP-43 and FUS

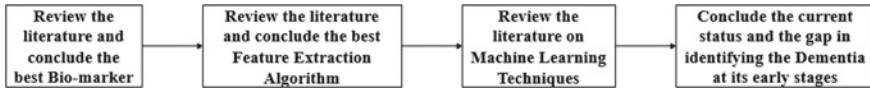


Fig. 1 Motivation and contribution of this study

in temporal and frontal lobes of the brain. This affects the personality, emotions and behaviour, speech and understanding of language [9]. In the diagnosis of the FTD, brain scans such as glucose positron emission scans and magnetic resonance imaging (MRI) are very useful [8].

According to the “Prevent Study” which is conducted by National Institute for Health Research of UK, says that even though the disease has arrived at an early stage of an individual’s life, it will stay dormant until the symptoms will surface after about 20–30 and the average survival of patients’ after clinical diagnosis is about 5–8 years [10]. It should be noted that no cure has been found for those affected with the captioned illness but some treatments at the early stages of dementia might be helpful to delay the evolution of dementia [11, 12]. Hence, recognizing the disease in early stages and delaying the progression is a highly important fact for managing the consequences of this illness [13]. However, as discussed in the preceding sections, having common symptoms similar to which a person will get due to ageing, the medical diagnosing of dementia is difficult. As a solution to this, various machine-learning techniques have been vastly investigated to identify individuals with potential dementia. However, according to the systematic review carried by Pellegrini et al. [14], still, there are gaps which make these machine-learning solutions, refrain from the routine usage. As such, there is a need for a suitable technical solution with a higher percentage of accuracy and eligibility for routine usage, which can be helpful to early diagnosis of dementia. Therefore, the motivation and contribution of this study are to discuss and review the current status of diagnosing dementia at early stages (see Fig. 1).

2 Biomarkers for Detecting Dementia

Dementia is occurring due to the affection with the disease of the brain or due to cognitive impairment [15]. In the past years, biomarkers such as biochemical, genetic, neuroimaging and neurophysiological biomarkers played a great role in revealing the early stages of dementia [16–19].

2.1 Biochemical Marker

Cerebrospinal fluid (CSF) and serum are the two main types for detecting dementia under this category [23, 28]. Several types of research have been done addressing the

development in protein analysis of total tau (T-tau), amyloid β ($A\beta$) and hyperphosphorylation tau (P-tau) in CSF and though this is specific for AD, Paraskevas et al. [20], provided a research where CSF was used to investigate the potential contribution for the differential diagnosis between AD, MCI and vascular dementia. T-tau and $A\beta_{42}$ measurements in CSF are used to identify the MCI and $A\beta_{42}$ and P-tau can assist in detect VD or FTD [21]. But the disadvantage of using the CSF and serum as the biomarkers is the limits of sensitivity and the specificity of these tests [22].

2.2 Genetic Biomarkers

A blood test can be helpful to evaluate the changes in thinking and memory. Thyroid-stimulating hormone (TSH), complete blood count (CBC), rapid plasma reagin (RPR), vitamin B12, comprehensive metabolic panel (CMP), and human immunodeficiency virus (HIV) are some of the most common blood tests [22]. Though genetic biomarkers provide an indication for developing dementia, it also needs the other biomarkers such as neuroimaging and chemical [18, 19, 22].

2.3 Neuroimaging

This can be classified into Structural scans and Functional scans [17].

2.3.1 Structural Scans

Computed tomography (CT) scan and magnetic resonance imaging (MRI) scan are under this type [17]. These are used to detect the affected area and the type of atrophy or vascular damage. While MRI is righteous at contrast resolution, CT is righteous at spatial resolution [23].

Rentoumi et al. [24] have presented a framework to predict dementia and cognitive impairment using brain MRI images concerning its automatic segmentation of grey and white matter regions as anatomical features. It is mentioned that the changes in the cerebral structure of demented patients can be influenced by the changes in the size or volume of these regions and also the thickness of the cortex helps to determine the severity level of the disease. As Nayaki and Varghese [25] mentioned in their research paper, grey matter (GM) of the brain is segmented and local patterns are extracted using GM and they have used the GM changes in NCI, MCI and AD patients and have found that MCI progresses than AD.

2.3.2 Functional Scans

Positron emission tomography (PET), Functional MRI (fMRI) and single-photon emission computed tomography (SPECT) are considered under this. These can measure brain metabolism parameters, such as regional cerebral glucose metabolism and regional cerebral blood flow are good at indicating AD and VD before morphological changes occur [17, 26].

Xia et al. [27] have used PET data of 86 subjects and with the proposed non-sparse infinite-kernel learning machine (NS-IKLM) recognition he was able to differentiate AD from normal people. By using 310 FDG-PET studies with 100 AD, 110 HC and 100 MCI, Lu et al. [28] have proposed an unsupervised and semi-supervised model to differentiate AD, MCI and NC (normal controls). PET scan is capable of showing activities of tissues of the brain and it can detect whether there is an increment of amyloid protein which is a sign of AD [27]. Geller et al. [29] have provided a research based on SPECT scan with quantile-based classification to differentiate AD, FTD patients and normal people. SPECT scan is capable of showing blood flows through arteries in the brain and affected areas show diminished activity [17]. Tripoliti et al. [30] have used fMRI which has the capability of showing vivid pattern activation on the brain and has differentiated the AD from normal people. His dataset consists of 40 gender and age-matched subjects; MCI and AD patients. Moreover, fMRI indirectly reflects neuronal activity and identify the brain activities which are correlated with cognitive tasks. This also can be used to measure the brain function over time based on blood oxygen level at rest [30]. Huang and Chen [31] proposed an fMRI-based immersive tool based on arterial spin labelling images to diagnose the severity of the dementia disease using 350 patients of including AD, MCI and non-cognitive impairment (NCI) patients.

Though there is a high spatial resolution for anatomical details, neuroimaging techniques have limited temporal resolution. Incapability of differentiating the stages within the brain distribution network in series or in parallel activation is another disadvantage [32]. Additionally, CT and MRI may be affected by fluid imbibition after brain injury in some cases, thus becoming incapable of detecting the best risk changing or becoming inadequately sensitive to detect dementia in its early stages [33].

2.4 Neuropsychology

Clinical biomarkers, such as EEG, quantitative electroencephalography and Vagus nerve stimulation are under this category [30]. EEG has shown significant growth in the research interest to detect dementia as the full investigation of neurodynamic time-sensitive biomarker [33–35]. EEG is a widely available faster than other imaging devices and it is a noninvasive method which is capable of detecting dementia early as well as to classify the severity degree at lower cost for mass screening [36]. Al-naimi et al. [23] provided in their research, where EEG is used on two EEG datasets

with over 65 years old AD and HC samples. They have suggested that the differences in EEG amplitudes are a successful biomarker by quantifying the slowing of EEG in the time domain. Also, Cejnek et al. [37] have done a research with 110 dementia patients' EEG test reports, based on novelty detection. Kapoor et al. [38] have used multivariate Fourier decomposition method with EEG samples to extract the features of brain signals. Fiscon et al. [39] have done a research using EEG tests on 86 patients with Fourier transformation and wavelet transformation [40].

Brain signals are non-stationary and a higher brain-imaging technique in both spatial and temporal resolution is essential to select. According to the above discussion, it is clear that EEG is the most capable technique to diagnose dementia at its early stage.

3 Pre-processing Signal

Many artefacts can be produced and added to the original signal when getting brain signals via EEG brain-imaging techniques [41]. According to Teplan [42], electromyography (EMG), electrocardiogram (ECG), mirror body movements, eye movements and sweating can be categorized under patient-related artefacts and EEG machine capable movements, battery issues, damaged wires and too much jelly-contained electrodes are under technical artefacts. In order to extract the vital signals without losing any important information, reduce the implementation complexity and information processing cost, these artefacts should be omitted. A variety of methods such as wavelet transform (WT), fast Fourier transform (FFT), frequency distributions (TFD), auto regressive method (ARM) and eigenvector methods (EM) have been vastly used for this purpose [43]. Comparison of main feature-extraction algorithms has shown in Table 1.

Fiscon et al. [39], have used fast Fourier transform (FFT) but wavelet transform is suggested for higher performance. Fiscon et al. [40], have proved in his research paper that wavelet transform has outperformed FFT. Al-Qazzaz et al. [30] have used a wavelet transform in his research. Suleiman et al. [44] have stated that, since EEG is a non-stationary signal, FFT provides less accurate results. As Suleiman et al. [44] mentioned in his research, he had to use an appropriate window with Fast Fourier function, because of its incapability of representing the non-stationary signals. So, according to the above discussion and the comparison in Table 1, it is clear that the wavelet feature-extraction algorithm is more capable of handling the EEG brain signals. After the pre-processing, the data needs to be trained and classified by applying classification algorithms to diagnose the disease.

Table 1 Comparison of feature-extraction algorithms

Algorithm	Advantages	Disadvantages
Wavelet transform	(1) Good to analyse non-stationary signals (2) Consists of both time and frequency information (3) Adopt window size according to the frequency Broad—Low Frequency Narrow—High Frequency (4) Good to analyse impermanent signal changes	(1) A proper mother wavelet and decomposition level should be selected
Fast Fourier Transform	(1) Faster than other methods	(1) Not suitable for non-stationary signals (2) Doesn't have a good spectral estimation (3) Incapable to analyse short EEG signals
Autoregressive	(1) Suitable for short segments (2) Give best frequency resolution (3) Reduce the loss of spectral problems (4) Provide improved frequency resolution	(1) Model order must be selected correctly. (2) Poor spectral estimation can be given due to incorrect model ordering (3) Hugely lack of consistency (4) Vulnerable to heavy biases
Time-frequency distribution	(1) Capable to analyse non-stationary signals (2) Feasible to analyse continuous segments of EEG signals	(1) Slow due to gradient ascent computation (2) Windowing process needs to be completed in the pre-processing phase (3) It is possible to depend on the extracted features on each other

4 Machine-Learning Techniques

As much as this disease has become a major problem for the society, various researches have been carried over the past years using different machine-learning techniques. To accomplish this, it has to use one or more classification algorithms based on their characteristics. The structure of the dataset and the number of data records have a huge impact on the flexibility and accuracy of each algorithm. Cejnek et al. [37] have performed an analysis on AD and HC samples to differentiate demented from non-demented with a 90% accuracy. The obtained increment of prediction error and adaptive weights during the prediction of EEG channels has been used to evaluate the novelty of EEG signals and has used gradient descent adaptation with linear dynamic neuron as the predictor. Another research has been done by

Houmani et al. [45], to perform an automated EEG diagnosis to differentiate subjective cognitive impairment (SCI), MCI, AD patients with other pathologies. They have shown that both techniques, namely “epoch-based entropy” and “bump modelling” are adequate for efficient differentiation between the above types and have used multi-class probabilistic SVM classifiers. They have gained 91.6% accuracy in differentiating SCI from AD. According to the review done by Arumugam and Aponso [41], random forest classifier and decision trees give better results than SVM in diagnosing dementia. Due to the lack of transparency of the output, complexity of the algorithm and the selection of the kernel, choosing SVM can be caused for a less percentage of accuracy in diagnosing dementia at early stages. Even in the random forest classifier, the need of high amount of data and the slowness can be major defects when diagnosing the illness [46].

Bansal et al. [47] have done a comparative analysis on J48 which is a C4.5 decision tree, random forest, Naïve-Bayes and multilayer CFSSubsetEval for attribute reduction to diagnose dementia and have shown that J48 has outperformed random forest, Naïve-Bayes and multilayer CFSSubsetEval. Jin and Deng [48] have performed the analysis on HC, AD, MCI and MCI converter (cMCI) using NCA features with boosting tree model and also has compared with sequential feature selection (SFS) and principal component analysis (PCA) with SVM classification and showed NCA feature-extraction with boosting tree model outperforms the other tested methods. They could gain 67.5 and 80% for HC and AD, respectively, but 27.5 and 50% accuracy for differentiating MCI and cMCI, respectively. Fiscon et al. [39] have used SVM, decision trees and rule-based classifiers to differentiate AD patients from health control (HC) and has stated that decision tree methods showed higher performances than the other two methods with 86, 88 and 83% accuracy for AD, HC and MCI, respectively. As Fiscon et al. [40] state in his second research, the decision tree is a good choice for the requirement and has outperformed SVM and all rule-based, function-based, Bayesian-based and naive-based classifications with 79, 83 and 92% of accuracy for differentiating MCI from AD, AD from HC, HC from MCI, respectively. The capability of handling high dimensional data is one of the reasons for this high accuracy of decision trees.

On the other hand, Liu et al. [49] have proposed an ensemble-learning framework based on artificial neural networks to differentiate AD, MCI and HC. In his research, they have shown in that this framework which was built using neural network has outperformed linear regression (LR), SVM, Naive-Bayes classifier (NB), logistic regression (LGR), Mmultimodal multitask (M3T) and high-order graph matching (HOGM). Lu et al. [50] have suggested a novel deep neural network to differentiate AD, MCI stages from HC with an accuracy of 82.4% in identifying MCI. Also, Ishfaqe et al. [46] have proven that the neural network is more capable of classifying brain signals than decision trees. This can be due to the high tolerance on noisy data, less amount of training and flexibility. Also, neural networks are faster than other methods because it has structured mimicking the brain processing.

According to the above analysis, it is noticeable that, decision trees and neural network have better characteristics. Also, Nanni et al. [51] have suggested an ensemble classification method which is acquired as a combination of SVM trained using

various clusters of data to improve the performance and outperformed all standalone classifiers such as SVM, random subspace of Adaboost (RS_AB), GaussianProcess-Classifier (GPC) and Random Subspace of Rotation Boosting (RS_RB). They have stated that ensemble has the great advantage of performing well than standalone classifiers. On the other hand, the dataset also has a huge impact on the accuracy of the result.

The main goal of this is to diagnose dementia in early stages so that the therapeutic treatments will be helpful to delay the progression of the condition [12, 52]. Therefore, the most important type which is essential to differentiate is Alzheimer's disease from mild cognitive impairment stages. Even though, as per the above discussion it can be seen that, though there are various machine-learning implementations have been done to diagnose dementia for the past years, most of them have focused on differentiating AD from HC and fewer studies and poor accuracy have been done and gained for differentiating AD from MCI stages. Even in the latest researches done by Ficon et al. [40], Salvatore and Castiglioni [53] and Nanni et al. [51] are not efficient and there are improvements that need to be done. Also, Pellegrini et al. [14] have shown in his systematic review, that there is still a gap in differentiating AD from MCI stages with better accuracy and these machine-learning techniques are still incapable for routine usage. Therefore, still, there is a need of a solution by considering advantages and disadvantages of each machine-learning technique to gain better accuracy and performance in diagnosing dementia in early stages.

5 Conclusion

In this paper, from its inception has considered information pertaining previous studies carried out with regard to the diagnosis of dementia in early stages using machine-learning techniques. Identifying the most suitable brain-imaging technique, the feature-extraction algorithm and classification algorithm are the main concerns that have focused on this paper. EEG as the best brain-imaging technique and wavelet as the best feature-extraction algorithm in diagnosing dementia in early stages has been derived through the comparison that has been done in this paper. Even though decision trees and artificial neural networks have gained the best results so far, it is still not good enough for routine usage due to the limitations of the algorithms. So, finding a proper accurate solution for this problem still remains as an open challenge.

Acknowledgements I would like to express my sincere gratitude to my supervisor Mr. Achala Chathuranga Aponso for providing me continuous support and guidance towards this research. And also, special gratitude for Ahila Arumugam as her research paper on dementia helped me to continue my research work. Special thanks to Informatics Institute of Technology and University of Westminster.

References

1. World Health Organization, WHO (2018) [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/dementia>. Accessed 31 Oct 2018
2. M.J. Prince, World Alzheimer Report 2015: The Global Impact of Dementia (25 Aug 2015). [Online]. Available <https://www.alz.co.uk/research/world-report-2015>. [Accessed 17 Dec 2018]
3. M. Prince, et al., *Dementia UK Second Edition—Overview*, p. 62 (2014)
4. F. Lewis, Estimation of future cases of dementia from those born in 2015 (2015) p. 12
5. F. E. Matthews et al., A two decade dementia incidence comparison from the cognitive function and ageing studies I and II, *Nat. Commun.*, **7** (2016)
6. J. Gallagher, Dementia tops female causes of death, 29 Oct 2014
7. Deaths registered in England and Wales (series DR)—Office for National Statistics. [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2017>. Accessed 17 Dec 2018
8. Alzheimer's Association| Alzheimer's Disease & Dementia Help, Alzheimer's Disease and Dementia (2018). [Online]. Available: <https://alz.org/>. Accessed 31 Oct 2018
9. Alzheimer's Research UK | ARUK, Alzheimer's Research UK (2018)
10. The 'PREVENT' Study | Join Dementia Research News. [Online]. Available <http://news.joindementiaresearch.nihr.ac.uk/prevent-study/>. Accessed 18 Oct 2018
11. J. Cummings et al., Drug development in Alzheimer's disease: the path to 2025. *Alzheimers Res. Ther.* **8**(1), 3 (2016)
12. P. Scheltens et al., Alzheimer's disease. *Lancet* **388**(10043), 505–517 (2016)
13. Why early diagnosis is important—Dementia—SCIE (2018). [Online]. Available: <https://www.scie.org.uk/dementia/symptoms/diagnosis/early-diagnosis.asp>. Accessed 31 Oct 2018
14. E. Pellegrini et al., Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **10**, 519–535 (2018)
15. S. Borson et al., Improving dementia care: The role of screening and detection of cognitive impairment. *Alzheimers Dement.* **9**(2), 151–159 (2013)
16. Diagnosis. [Online]. Available: <https://stanfordhealthcare.org/medical-conditions/brain-and-nerves/dementia/diagnosis.html>. Accessed 18 Dec 2018
17. Tests for Dementia, Memory and Aging Center. (2018). [Online]. Available: <https://memory.ucsf.edu/tests-dementia>. Accessed 31 Oct 2018
18. A. Cedazo-Minguez, B. Winblad, Biomarkers for Alzheimer's disease and other forms of dementia: Clinical needs, limitations and future aspects. *Exp. Gerontol.* **45**(1), 5–14 (2010)
19. H. Hampel et al., Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nat. Rev. Drug Discov.* **9**(7), 560–574 (2010)
20. G.P. Paraskevas et al., CSF biomarker profile and diagnostic value in vascular dementia. *Eur. J. Neurol.* **16**(2), 205–211 (2009)
21. S.V. Frankfort, L.R. Tulner, J.P.C.M. van Campen, M.M. Verbeek, R.W.M.M. Jansen, J.H. Beijnen, Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia: a review of recent literature, *Curr. Clin. Pharmacol.* (2008). [Online]. Available: <http://www.eurekaselect.com/66878/article>. Accesse 26 Aug 2018
22. S.T. DeKosky, K. Marek, Looking backward to move forward: early detection of neurodegenerative disorders. *Science* **302**(5646), 830–834 (2003)
23. A.H. Al-nuaimi, E. Jammeh, L. Sun, E. Ifeachor, Changes in the EEG amplitude as a biomarker for early detection of Alzheimer's disease, in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2016), pp. 993–996
24. V. Rentoumi et al., Automatic detection of linguistic indicators as a means of early detection of Alzheimer's disease and of related dementias: a computational linguistics analysis, in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen* (2017), pp. 000033–000038

25. K.S. Nayaki, A. Varghese, Alzheimer's detection at early stage using local measures on MRI: a comparative study on local measures, in *2014 International Conference on Data Science & Engineering (ICDSE)* (Kochi, India, 2014), pp. 224–227
26. N.K. Al-Qazzaz, S.H.B.M. Ali, S.A. Ahmad, K. Chellappan, M.S. Islam, J. Escudero, Role of EEG as biomarker in the early detection and classification of dementia, *Sci. World J.* (2014). [Online]. Available: <https://www.hindawi.com/journals/tswj/2014/906038/>. Accessed 26 Aug 2018
27. Y. Xia, S. Lu, W. Wei, D.D. Feng, Y. Zhang, Non-sparse infinite-kernel learning for automated identification of Alzheimer's disease using PET imaging, in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)* (Singapore, 2014), pp. 855–860
28. S. Lu, Y. Xia, W. Cai, D.D. Feng, M. Fulham, Cross-cohort dementia identification using transfer learning with FDG-PET imaging, in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC, 2018), pp. 1550–1554
29. D. Geller, G. Platsch, J. Kornhuber, T. Kuwert, D. Merhof, Quantile-based classification of Alzheimer's disease, frontotemporal dementia and asymptomatic controls from SPECT data, in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)* (2016), pp. 1–5
30. E.E. Tripoliti, D.I. Fotiadis, M. Argyropoulou, A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. *Artif. Intell. Med.* **53**(1), 35–45 (2011)
31. W. Huang, G. Chen, A novel functional MRI-based immersive tool for dementia disease severity prediction via spectral clustering and incremental learning, in *2015 International Conference on Orange Technologies (ICOT)* (2015), pp. 169–172
32. P.M. Rossini, S. Rossi, C. Babiloni, J. Polich, Clinical neurophysiology of aging brain: from normal aging to neurodegeneration. *Prog. Neurobiol.* **83**(6), 375–400 (2007)
33. A.L. Schneider, K.G. Jordan, Regional attenuation without delta (RAWOD): a distinctive EEG pattern that can aid in the diagnosis and management of severe acute ischemic stroke. *Am. J. Electroneurodiagnostic Technol.* **45**(2), 102–117 (2005)
34. J. Jeong, EEG dynamics in patients with Alzheimer's disease. *Clin. Neurophysiol.* **115**(7), 1490–1505 (2004)
35. H. Hampel et al., Perspective on future role of biological markers in clinical therapy trials of Alzheimer's disease: A long-range point of view beyond 2020. *Biochem. Pharmacol.* **88**(4), 426–449 (2014)
36. S.M. Snyder, J.R. Hall, S.L. Cornwell, J.D. Falk, Addition of EEG improves accuracy of a logistic model that uses neuropsychological and cardiovascular factors to identify dementia and MCI. *Psychiatry Res.* **186**(1), 97–102 (2011)
37. M. Cejnek, I. Bukovsky, O. Vysata, Adaptive classification of EEG for dementia diagnosis, in *2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)* (Prague, Czech Republic, 2015), pp. 1–5
38. E. Kapoor, V. Johnson, S. Pati, V.K. Chakka, Fourier decomposition method based descriptor of EEG signals to identify dementia, in *2016 IEEE Region 10 Conference (TENCON)* (Singapore, 2016), pp. 2474–2478
39. G. Fison et al., Alzheimer's disease patients classification through EEG signals processing, in *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (2014), pp. 105–112
40. G. Fison et al., Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC Med. Inform. Decis. Mak.* **18**, 35 (2018)
41. A. Arumugam, A. Aponso, Review of Brain Imaging Techniques, Feature Extraction and Classification Algorithms to Identify Alzheimer's Disease (2016)
42. M. Teplan, Fundamental of EEG Measurement, ResearchGate. [Online]. Available: https://www.researchgate.net/publication/228599963_Fundamental_of_EEG_Measurement. Accessed 18 Dec 2018
43. A.S. Al-Fahoum, A.A. Al-Fraihat, Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains, *International Scholarly Research*

- Notices (2014). [Online]. Available: <https://www.hindawi.com/journals/isrn/2014/730218/>. Accessed 18 Dec 2018
44. A. Suleiman, S. Toka, A.-H. Fatehi, Features extraction techniques of EEG signal for BCI applications (Sep, 2018)
 45. N. Houmani et al., Diagnosis of alzheimer's disease with electroencephalography in a differential framework. *PLoS ONE* **13**(3), e0193607 (2018)
 46. A. Ishfaq, A.J. Awan, N. Rashid, J. Iqbal, Evaluation of ANN, LDA and Decision trees for EEG based Brain Computer Interface, in *2013 IEEE 9th International Conference on Emerging Technologies (ICET)* (Islamabad, Pakistan, 2013), pp. 1–6
 47. D. Bansal, R. Chhikara, K. Khanna, P. Gupta, Comparative analysis of various machine learning algorithms for detecting dementia. *Procedia Comput. Sci.* **132**, 1497–1502 (2018)
 48. M. Jin, W. Deng, Predication of different stages of Alzheimer's disease using neighborhood component analysis and ensemble decision tree. *J. Neurosci. Methods* **302**, 35–41 (2018)
 49. J. Liu, S. Shang, K. Zheng, J.-R. Wen, Multi-view ensemble learning for dementia diagnosis from neuroimaging: An artificial neural network approach. *Neurocomputing* **195**, 112–116 (2016)
 50. Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. [Online]. Available: <https://dash.harvard.edu/handle/1/37067787>. Accessed 20 Dec 2018
 51. L. Nanni, A. Lumini, N. Zaffonato, Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment. *J. Neurosci. Methods* **302**, 42–46 (2018)
 52. J. Cummings et al., Drug development in alzheimer's disease: The path to 2025. *Alzheimers Res. Ther.* **8**(1), 39 (2016)
 53. C. Salvatore, I. Castiglioni, A wrapped multi-label classifier for the automatic diagnosis and prognosis of Alzheimer's disease. *J. Neurosci. Methods* **302**, 58–65 (2018)

Interactive Visualization of Ontology-Based Conceptual Domain Models in Learning and Scientific Research



Dmitry Litovkin, Anton Anikin and Marina Kultsova

Abstract The paper presents an approach to knowledge transferring and sharing on the base of the semantic link network (SLN) representing expert knowledge in the explicit form. To provide an efficient SLN understanding, it is represented with a geometric graph which can be interactively visualized using a combination of the appropriate visualization methods. The coupling of these methods allows getting a different level of details of the SLN visualization in accordance with the user needs. The proposed approach is planned being implemented in the knowledge management system for learning and scientific research.

Keywords Semantic link network · Ontology · Graph · Interactive visualization · Semantic zooming

1 Introduction

An ontology is a convenient tool of domain modeling for a wide range of tasks in various subject domains including learning and scientific research. Being useful for automated information processing in the intelligent systems, at the same time, the ontology requires to use of some special visualization techniques which help a human in ontology perception and understanding. In learning and scientific research, the subject domain might have a rather complex structure and cover a huge set of concepts and relations between them. This fact makes complicated studying appropriate information resources by the persons who are not familiar with the domain. So, due to the complexity, it is difficult to overview and understand the conceptual model represented with the ontology. The choice of various approaches and tools for ontology visualization and navigation depends on the user category who works with

D. Litovkin · A. Anikin (✉) · M. Kultsova
Volograd State Technical University, Volograd, Russia
e-mail: anton@anikin.name

M. Kultsova
e-mail: poas@vstu.ru

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_31

the ontology and the use cases. But in most of the cases, the user deals with the range of issues mentioned in [2]. Thus, developing methods and tools for the convenient and informative representation of ontology for its users is still being a relevant task.

2 State-of-the Art and Related Work

In learning and scientific research, there are two main available knowledge sources—document collection and expert knowledge (knowledge of teachers or researchers). A document collection consists of a large number of heterogeneous and contradictory documents. Its main advantage is high availability but the disadvantage is a rather high semantic noise [15, 18]. Semantic communication noise is a type of disturbance in message transmission that interferes with the message interpretation. It is generated by the content or semantic errors and message distortions during their encoding/decoding.

Expert knowledge can be implicit and potentially explicit knowledge. Implicit knowledge can be gained only by connecting people (for example, sharing and discussing the problems and best practices). The main disadvantage of this knowledge source is non-replicability of such knowledge.

To increase the availability of expert knowledge and reduce semantic noise during knowledge transfer, we propose to create knowledge repositories (cognitive information space—CIS [3]) for each user group with a similar profile. Cognitive information space defines a representation of certain subject domain from the point of view of this user group. CIS includes a focus question, user profile, semantic link network (SLN), document collection, and conceptual index. It should be noted that for different cognitive information spaces, the common document collection can be used.

Focus question determines context, main subject, and boundaries of the scope of knowledge being studied [12].

User profile includes knowledge/competencies, which the user already has; user cognitive style; knowledge/competencies that need to be obtained. The user profile should be consistent with the focus question.

Semantic link network (SLN) is a network that represents semantic relations between concepts [20]. The SLN contains domain WHAT-knowledge [7] structured into a single mental model and answers some focus questions. Each SLN item (a concept, a link between concepts, or a set of the concept/link attributes) is assigned a priori importance of this item in terms of SLN sensemaking as a whole. To reduce the semantic noise associated with SLN encoding and decoding, the SLN is presented both in the formal language OWL 2 [13] and the visual language ORM 2 [5]. SLN formal representation is used by a computer, and its visual representation is used by a human.

Conceptual index stores set of the links between SLN items and documents (or documents fragments) in which knowledge about these items is described. Each link is associated with an assessment that reflects document usefulness as a description of the knowledge piece in terms of the user profile and the focus question. The assess-

ment takes into account the quality of knowledge coding for the user group and the semantic noise that occurs when knowledge decoding. As a result, the assessment of the usefulness of the same document will be different for users with different profiles and for different SLNs.

Knowledge transferring and sharing in learning and scientific research is a CIS learning process (Fig. 1) that implies the steps below:

Step 1. The user sets his profile and keywords that describe his knowledge needs. As a result of the computer-aided search, he gets the ranked list of the appropriate

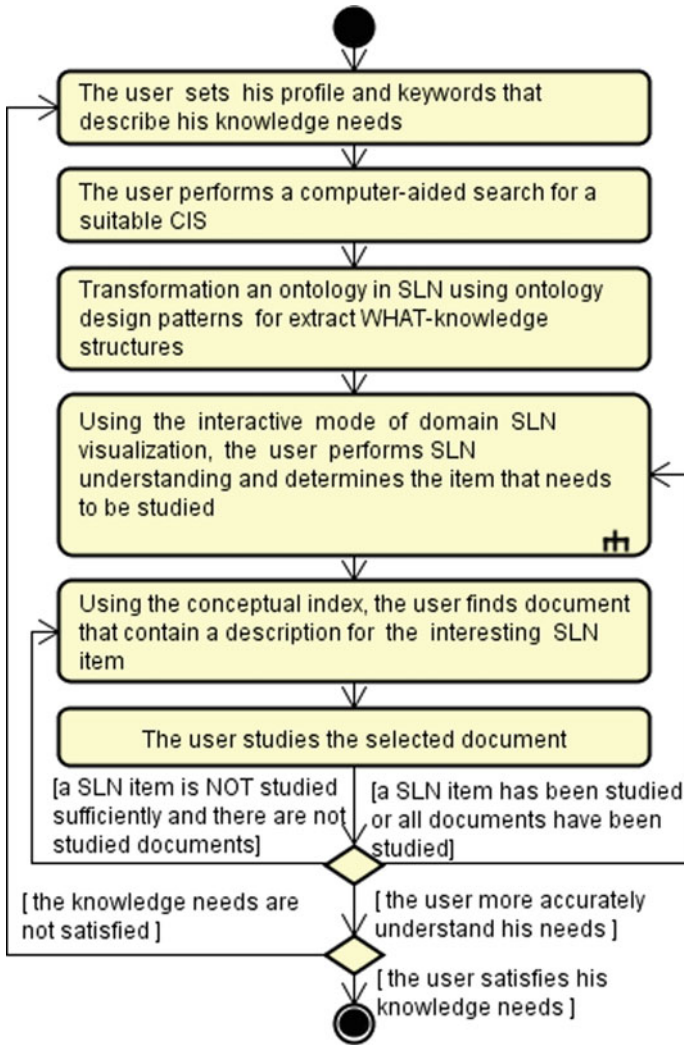


Fig. 1 CIS learning

CISs. Next, the user reviews the CISs (focus question, target user profile, and the SLN) and selects the best one for his needs.

Step 2. Transformation of OWL2-ontology into SLN using ontology design patterns for WHAT-knowledge structures extraction [1].

Step 3. Using the interactive mode of the domain SLN visualization, the user performs a preliminary SLN understanding and determines the item that needs to be studied the first.

Step 4. Using the conceptual index, the user finds documents that contain a description for the SLN item in which he is interested. Looking through the usefulness assessments of the documents, and their content, the user decides which document should be studied the first.

Step 5. After the selected document has been studied, the user returns to Step 4 to continue studying other documents for the given SLN item, or returns to Step 3 to re-understand domain SLN and search some new item for studying.

Step 6. The study of domain SLN and related documents continues until the user satisfies his knowledge needs. On the other hand, after some immersion in the domain, the user can more accurately formulate his needs and return to Step 1.

In this paper, we consider the task of SLN visualization for SLN understanding and determination of the item that needs to be studied (Step 3). For SLN visualization, a node-link paradigm is used that imposed by the visual modeling language ORM 2. As a result, a geometric graph is created, whose nodes are represented by distinct points in general position in the plane and whose edges are drawn as straight line segments, perhaps with crossings. According to ORM 2 notation, the nodes are concepts, attributes, and predicates, and the edges are the links of different types between the nodes. Every node and link is represented as a visualized object of some type. As a visual representation of the SLN is a strongly connected geometric graph with many nodes, edges, and labels used for various purposes, interactive visualization is necessary [9–11].

One of the key features of SLN understanding is its representation as a graph with different levels of the details. Ben Shneiderman formulated the visual information seeking mantra: overview first, zoom and filter, then details on demand [16]. These interactive visualization methods are widely used for large graphs representation [2, 6, 8, 10]:

- Geometric zooming is a technique when the user can only observe different geometrical scales of a visualized object in the geometric graph.
- Filtering, aggregation, and embed focus and context information are techniques that provide the ability to reduce shown graph items (nodes and edges):
 - Filtering removes items, whereas aggregation creates a single new item instead of multiple others and replaces them. Filtering is more straightforward for users to understand. However, users often forget items that have been filtered out, even when they were recently filtered. Aggregation is somewhat safer from the cognitive point of view because the stand-in item contains knowledge about all items that it replaces.

- The focus+context technique uses the paradigm of emphasizing the items in the center (focus), providing at the same time an overview of the global graph structure (context). The goal of embedding focus and context together is to mitigate the potential for disorientation that comes with standard navigation techniques. Focus+context idioms attempt to support orientation by providing contextual information intended to act as recognizable landmarks, using external memory to reduce internal cognitive load.

However, in [4, 14, 19] it notes, that methods for interactive visualization of the ontology graphs are underdeveloped yet. Moreover, most of the existing tools are not able to visualize a significant proportion of OWL 2 [17]; meanwhile, the proposed approach requires to visualize not only owl axioms, but the WHAT-knowledge structures too. So, the development of the approach and software tool for interactive visualization of the ontology-based SLN, containing WHAT-knowledge structures, is very important in the context of large ontologies visual representation.

3 Interactive SLN Visualization for Understanding and Exploration

The main idea of the proposed approach to SLN visualization for understanding and exploration (including determining the items to be studied) is presented in Figs. 2, 3, 4. At the first step, the user performs a high-level SLN overview from the expert point of view. Then, the user explores the SLN to find SLN fragment he is interested in. Finally, he determines the item that needs to be studied. To implement the proposed process, a different level of details of SLN visualization at different times is required:

1. A high-level SLN overview. It is a minimal possible visualization level of the model details, required for the rough (approximate) answer for the focus question. A goal of the high-level overview design is to show all key items in the SLN simultaneously, without any need for navigation, pan, or scroll. The high-level overview is recommended to use at the beginning of the exploration process, to guide users in choosing the points where they can drill down to inspect the conceptual domain model in more detail [10]. The high-level SLN overview should reflect the point of view of the expert.
2. Average details for approximate/preliminary consume knowledge, as well as for exploration, i.e., the transition from one fragment of SLN representation to another one.
3. High or maximum details presentation of knowledge in focal point/region. The focal point indicates SLN item or position on the geometric graph to which the user pays attention as a starting point of the exploration.

The existing interactive methods for graphs and ontology visualization were analyzed, and as a result, the method below was proposed to use for SLN visualization on different levels of the details:

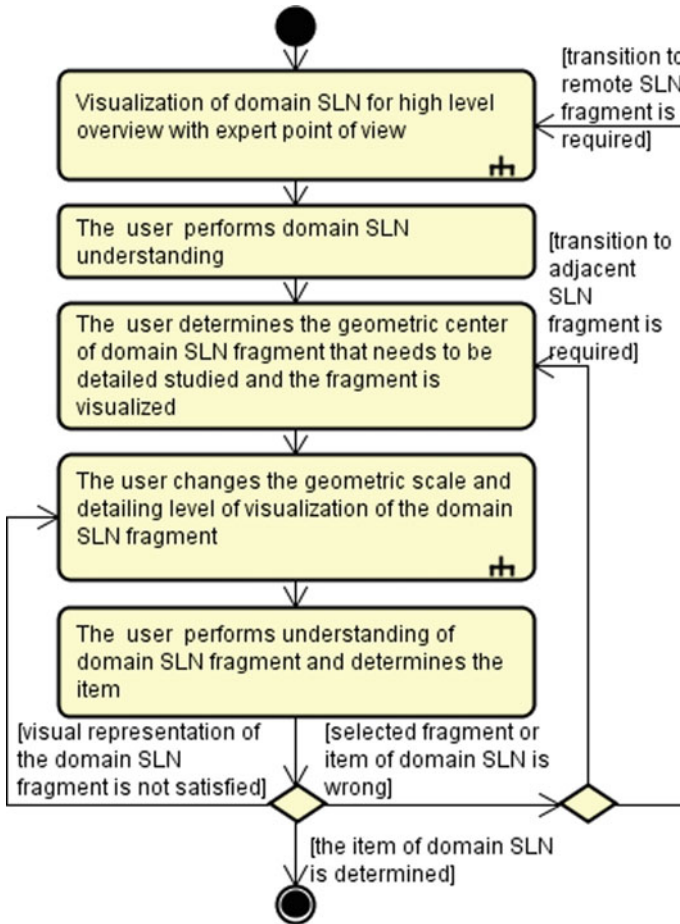


Fig. 2 SLN understanding and determining of the item that needs to be studied

1. Geometric zooming—allows user setting a various size of the visualized object in the geometric graph.
2. Semantic zooming—allows user controlling the level of details of the domain SLN visualization. We use the term “semantic zooming” as it was defined in [19] opposite the definition given in [10]:
 - (a) Changing the set of visible SLN items at different semantic scales. Use of the combination of aggregation, filtering of SLN items and embedding focus information within surrounding context:
 - i. Filtering all non-key SLN items and aggregation trails between key concepts with the creation of stand-in items as simplified links. To define the key and non-key items, DOI-metric [10] is used. The degree of in-

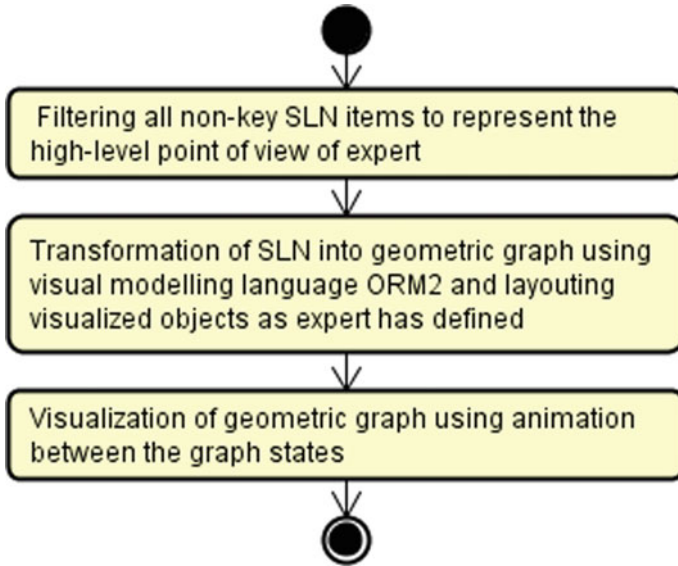


Fig. 3 Representation of the high-level SLN visualization from the expert point of view

terest metric (DOI-metric) for every SLN item should take into account a priori importance, defined by an expert, and distance from item to focus point (target item or position on the geometric graph should be defined). So, only SLN items with DOI-metric is over threshold value are visualized. The threshold value changes depending on the geometric scale. This filtering is enabled by default, but the user can change this to use it on demand. This filtering is enabled by default for the high-level SLN overview.

- ii. Setting the user focus region (with the center in focus point), where the geometric scale distorted upwards (e.g., fisheye lens distortion idiom or magnifying lens distortion idiom is used). In conjunction with methods a) and i), it allows achieving the maximum details representation of knowledge in the focus region.
- iii. Aggregating SLN-fragments into clusters and creating stand-in items as super-nodes and inducing links between super-nodes as well as between super-nodes and concepts. A cluster is a node grouping based on the similarity metric, where nodes (concepts and attributes) within a cluster are more similar to each other than to ones in another cluster [10]. The similarity metric is the degree of a node (i.e., the number of links connected to the node). Moreover, the cluster cannot contain items with high priority importance, so they cannot be stand-in items. SLN clustering is used by default and also to obtain the minimum or average details level. The user can expanse/collapse on demand selected SLN

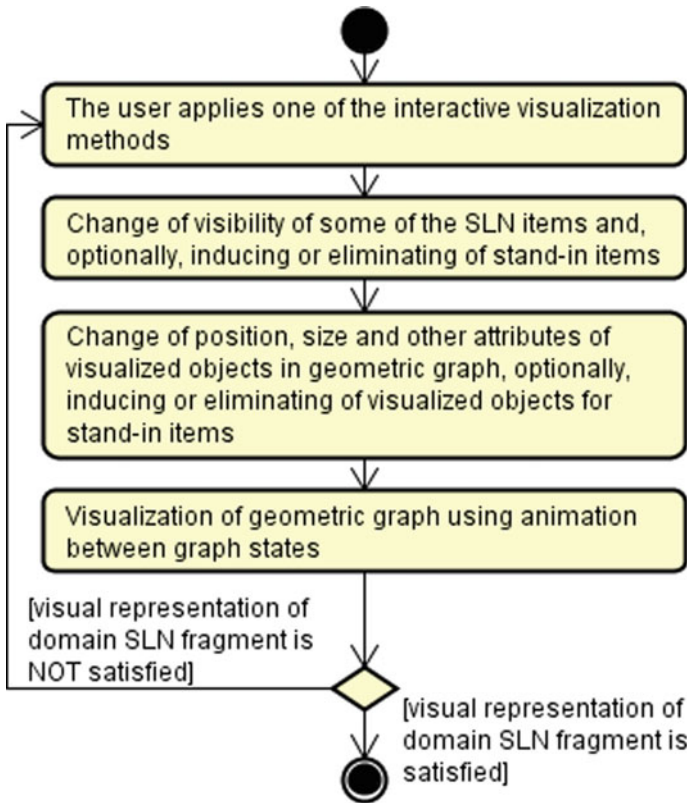


Fig. 4 Changing the detail level of the visual representation of the domain SLN fragment

- cluster. Cluster expansion can be done taking into account or ignoring DOI-metric for items of the cluster.
 - iv. Filtering on demand in SLN all attributes and all links of the selected type taking into account or ignoring DOI-metric.
 - v. Filtering on-demand context for target concept—searching all the adjacent items one or a few hops away from a target concept and eliminating them. The context collapse/expanse is initiated by the user taking into account or ignoring DOI-metric for items of the context.
 - vi. Aggregating multi-links connecting a pair of concepts and inducing the new simplified link. Multi-links collapse/expansion is done by default depending on their DOI-metric; however, the user can collapse/expanse all multi-links on demand.
- (b) Changing the details in the visualized object representation in geometric graph depending on geometrical scale. Wherein it should be taken into account, that the geometric scale in the focus region is magnified. The representation of the visualized object adapts to the number of pixels available

in the image-space region occupied by the object. The representation of the visualized object adapts to the number of pixels available in the image-space region occupied by the object.

3. Changing the visual appearance for the visualized object in the focal point and/or focal region—e.g., color change. Also, the size of the visualized object in the focal point and/or focal region is changed due to the geometric scale distortion.

4 Conclusion and Future Work

In the paper, a process of the SLN understanding and determining the item that needs to be studied is considered in detail. This process is a part of the more general process of knowledge transferring and sharing on the base of cognitive information space in learning and scientific research. One of the key features of SLN understanding is its representation as a graph with different levels of the details. To implement these different levels of the SLN visualization, the specific interactive visualization method was developed as a combination of the methods for graphs interactive visualization. This method implements the high-level SLN overview, average detail for exploration, and maximum detail in focal point/region. The method at the same time takes into account the expert's point of view and allows varying the details of visualization level depending on the user focus point and his explicit requirements. As a future work, we plan to implement the proposed method in SLN visualization software tool based on the ORM 2 notation.

Acknowledgements This paper presents the results of research carried out under the RFBR grants 18-07-00032 and 18-47-340014.

References

1. A. Anikin, M. Kultsova, D. Litovkin, E. Sarkisova, Representation of what-knowledge structures as ontology design patterns, in *2018 9th International Conference on Information, Intelligence, Systems Applications (IISA)*, July 2018
2. A. Anikin, D. Litovkin, M. Kultsova, E. Sarkisova, T. Petrova, Ontology visualization: approaches and software tools for visual representation of large ontologies in learning. *Commun. Comput. Inform. Sci.* **754**, 133–149 (2017)
3. A. Anikin, D. Litovkin, M. Kultsova, E. Sarkisova, Ontology-based collaborative development of domain information space for learning and scientific research, in *Knowledge Engineering and Semantic Web: 7th International Conference (KESW 2016)*, Prague, Czech Republic, ed. by A.C. Ngonga Ngomo, P. Křemen, 21–23 Sept 2016, pp. 301–315. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-45880-9_23
4. M. Dudáš, O. Zamazal, V. Svátek, Roadmapping and navigating in the ontology visualization landscape, in *Knowledge Engineering and Knowledge Management*, ed. by K. Janowicz, S. Schlobach, P. Lambrix, E. Hyvönen (Springer International Publishing, Cham, 2014), pp. 137–152

5. T. Halpin, *ORM 2 Graphical Notation* (Neumont University, Salt Lake City, 2005)
6. I. Herman, G. Melançon, M.S. Marshall, Graph visualization and navigation in information visualization: a survey. *IEEE Trans. Visual. Comput. Graphics* **6**(1), 24–43 (2000). <https://doi.org/10.1109/2945.841119>
7. D. Kudryavtsev, T. Gavrilova, From anarchy to system: a novel classification of visual knowledge codification techniques. *Knowl. Process Manage.* **24**(1), 3–13 (2017). <https://onlinelibrary.wiley.com/doi/abs/10.1002/kpm.1509>
8. T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J.J. van Wijk, J. Fekete, D.W. Fellner, Visual analysis of large graphs: state-of-the-art and future research challenges. *Comput. Graph. Forum* **30**(6), 1719–1749 (2011)
9. T. Munzner, in *Interactive visualization of large graphs and networks*, Ph.D. thesis, Stanford University, June 2000
10. T. Munzner, E. Maguire, *Visualization Analysis and Design*, AK Peters visualization series (CRC Press, Boca Raton, 2015)
11. K. Nazemi, Adaptive semantics visualization, in *Studies in Computational Intelligence*, vol. 646 (Springer, Berlin, 2016). <https://doi.org/10.1007/978-3-319-30816-6>
12. J.D. Novak, A.J. Caas, Theoretical origins of concept maps, how to construct them, and uses in education. *Reflect. Educ.* **3**(1), 29–42 (2007)
13. OWL 2 Web Ontology Language Primer. <https://www.w3.org/TR/owl2-primer/>
14. S. Ramakrishnan, A. Vijayan, A study on development of cognitive support features in recent ontology visualization tools. *Artif. Intell. Rev.* **41**(4), 595–623 (2014). <https://doi.org/10.1007/s10462-012-9326-2>
15. C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423, 623–656 (1948)
16. B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, in *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL'96)* (IEEE Computer Society, Washington, DC, USA, 1996), pp. 336–343
17. G. Stapleton, M. Compton, J. Howse, Visualizing owl 2 using diagrams, in *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Oct 2017, pp. 245–253
18. W. Weaver, Recent contributions to the mathematical theory of communication, in *A Mathematical Theory of Communication* (University of Illinois Press, Champaign, 1949)
19. V. Wiens, S. Lohmann, S. Auer, Semantic zooming for ontology graph visualizations, in *Proceedings of the Knowledge Capture Conference (K-CAP 2017)* (ACM, New York, 2017), pp. 4:1–4:8
20. H. Zhuge, Semantic linking through spaces for cyber-physical-socio intelligence: a methodology. *Artif. Intell.* **175**(5), 988–1019 (2011) (special Review Issue)

A Secure Lightweight Mutual Authentication and Message Exchange Protocol for IoT Environments Based on the Existence of Active Server



Omar Abdulkader, Alwi M. Bamhdi, Vijey Thayananthan, Kamal Jambi, Bandar Al Ghamdi and Ahmed Patel

Abstract Recently, the Internet of Things (IoT) has started to play an important role and one of the states of art solutions to solve various issues in different ICT application domains due to its intrinsic characteristics. However, its security and privacy mechanisms are still not well-tailored and lagging behind. Massive academic and industrial surveys, researches, and studies have been conducted and implemented but the general consensus is that conventional cryptographic methods are not overly suitable for adoption in IoT environments in a straightforward manner without incurring huge operational, computational, storage, and energy costs. Therefore, an alternative is to a lightweight cryptographic method offering high levels of data and system security to mitigate such computational cost, storage capacity, and energy consumption. This paper proposes a lightweight mutual authentication and message exchange scheme between IoT devices via a publically available server based on symmetric and asymmetric hybrid cryptography. The server plays an important role to register and authenticate different IoT devices in a federated environment. Security analysis shows that the proposed scheme satisfies the main security properties and it is resistant against attacks.

O. Abdulkader (✉) · V. Thayananthan · K. Jambi
Department of Computer Science, King Abdul-Aziz University, Jeddah, Kingdom of Saudi Arabia
e-mail: Luk_amri@hotmail.com

V. Thayananthan
e-mail: vthayanathan@kau.edu.sa

K. Jambi
e-mail: kjambi@kau.edu.sa

A. M. Bamhdi
Department of Computer Science, Umm AlQura University, Jazan, Kingdom of Saudi Arabia
e-mail: ambamhdi@uqu.edu.sa

B. Al Ghamdi
Department of ITC, Arab Open University, Jeddah, Kingdom of Saudi Arabia
e-mail: b.alrami@arabou.edu.sa

A. Patel
Universidade Estadual do Ceará, Fortaleza, Brazil
e-mail: whinchat2010@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_32

Keywords IoT · Lightweight · Mutual authentication · Cryptographic · Cybersecurity · Message exchange

1 Introduction

IoT deems as an emerging and promising technology that is promising to convert things into smart things and connect them with each other without or with minimal human interventions. The current studies estimate the total numbers of smart things are growing rapidly, assuming it will reach 50 billion in 2020. So, the challenge appears here is how we will be able to manage and secure this massive number of devices and provide them with the same opportunity to access network while preserving privacy and security for devices generated data [1], also guarantee their QoS requirements. As we know that there are a lot of heterogeneous devices each has its own characteristic and producing different data type which related to their dedicated purpose. So, our concern is whether the existing security approach can achieve IoT privacy preservation and security requirements and resist against different types of attacks (passive and active). Cryptographic still and remains the security methodology that used to achieve security and mutual authentication requirement in different domains [2]. All researchers are agreeing that the existing cryptographic techniques are suffering from high computation overhead and energy consumption. So, it is not adequate to adopt them for IoT environments; this led us for more enhancements and optimization on the existing cryptographic to propose a lightweight cryptographic and mutual authentication that met IoT devices requirements [3, 4]. To produce an optimal cryptographic, mutual authentication, and security preserving, massive of academic research has been conducted. Those researchers aimed to study existing encryption algorithm characteristics and make the necessary improvements or proposed a new lightweight cryptographic and mutual authentication algorithm [5–7]. Our proposed scheme aims to mitigate the long and complicated authentication negotiation and ensure that the messages have been exchanged in a secure manner. However, the proposed model consists of; registration phase, authentication and message exchange phase. Consequently, contribution in this paper can be summarized as follows:

- Registration phase aims to secure machines registration to the system model. In this phase, each device will register with the system in a secure management manner through the disclosure of the server's presence to the associated IoT devices via a publically available address and naming facility like a URL. During this phase in the proposed scheme, the server and associated IoT devices collect and store the necessary registration information (variable) about each other, with the server through its secure management and administration system offering a secure and personalized unique crypto identification for each IoT device that eliminates fraudulent or masquerading registration of IoT devices.
- During the authentication and message exchange phase, the server is an exclusive publically known mediator between IoT devices associated to it and the server

authenticates the IoT devices to each other and ensures that the ability of an adversary to eavesdrop, hack tamper, and interrupt the security of the operation of the proposed modeled system has been eliminated by taking into account the gambit of authentication, authorization, confidentiality, data integrity, randomization, and genuine verification of message exchanges at all times.

The rest of the paper is organized as follows. In Sect. 2, literature reviews have been discussed. The system model discusses in Sect. 3. Section 4 discusses the security analysis. Finally, Sect. 5 concludes the paper.

2 Literature Reviews

Literature review on cryptographic and mutual authentication is a variant, and different solutions have been proposed to address lightweight mutual authentication issues [8, 9]. He et al. [10] propose a lightweight RFID authentication scheme based on elliptic curve cryptosystem. The proposed scheme has been compared with Liao and Hsiao's scheme. Three main criteria (computational cost, storage capacity, and communication cost) have been measured and the proposed scheme overcomes in the computational cost and storage capacity while achieving the same level in communication storage. Meanwhile, it achieves the main security properties and resisting against a different type of attacks. The weakness that appears in the proposed scheme that the tag identifiers are stored in plain text for both sides server and tag which make it easy for tag identifiers disclosure if the adversary success to attack both sides server and tag. Liao et al. [11] propose a secure mutual authentication and Id-verifier transfer scheme based on ECC. It consists of two-phase; setup phase: where preshared variables are acquainted between the server and the tag, and authentication phase: where are authentication process and message exchange will be done through different authenticate challenge. Security analysis shows that the proposed scheme is resisted against different types of attacks with low storage requirements, computational cost, and communication overhead. Devi et al. [12] proposed two authentication approaches for IoT application. The first approach is based on MAC where all related IoT devices MACs are stored on DBMS. Therefore, the authentication will be done through ensuring that any connect request to the server will be checked by comparing the MACs for the connection requesters to those that stores in DBMS. Based on the comparing results, the access is gained or rejected. The other approach is based on the hash function where the requesters provide his IDs to the server and get the one-time password by applying hash function for both server side and user side. The second approach takes much time than the first approach due to the authentication negotiation between the server and users. Lightweight mutual authentication between reader and tags has been proposed by Fan et al. [13]. In their proposed scheme, cache concept has been implemented on the reader side. The cache will store the last keys for the last successful authentication while the DB will store all tags keys. At the first, the reader will check for the tag's key within local cache;

if the key is existing, the hash comparison will be performed and the authentication successful or failed will be determined based on the comparison result. If the tag key does not exist on the reader cache, then the reader will forward the request to the DB. Later on, the DB will check for the tag key and send the result to the reader. If the tag key exists with the DB repository then the reader will update the cache accordingly. It seems that the search method for the tag key considers as computation cost as it is done sequentially on the reader cache or DB repository. As an improvement, the ultralightweight has been introduced. The ultralight mutual authentication applies simple operation such as concatenate, XOR, and it resists against different types of attacks. Tewari et al. [14] propose a mutual authentication scheme based on a one-way hash function and bitwise operation. The main goal is to achieve low storage space and low computational cost on both tag and server side. The mutual authentication is done through four negotiations steps between tags and server. While the proposed scheme proves its capabilities to be secure against various types of attacks, still tango attack, desynchronize, and full disclosure attack do not consider in this scheme.

3 The Proposed Scheme

Figure 1 shows the hardware architecture for the proposed scheme where is comprised of IoT devices and the server. To mitigate the capacity, computation, and energy overhead, the IoT devices only contain the following variables; the device IDs (Id_i, Id_j, \dots, Id_n), symmetric key (S_k), group symmetric key (GS_k), random nonce number (N_i, N_j, \dots, N_n), and server public key (P_s), respectively. While the server contains the following variables: a database for all devices IDs, the symmetric keys

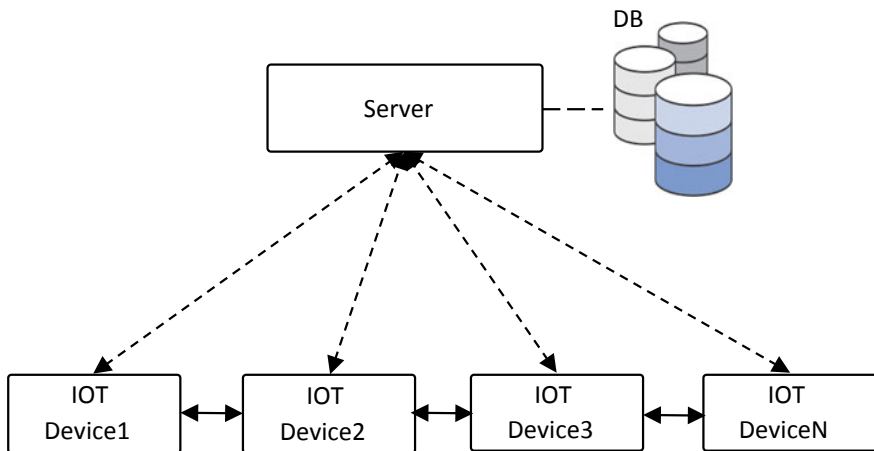


Fig. 1 Hardware architecture for the proposed schemed

for each device, group symmetric keys, server random nonce number (N_s), and his own public (P_s) and private key (P_r), respectively. In the proposed scheme, we aim to provide lightweight mutual authentication between the IoT devices through the server. Well-known that the IoT device is constrained device with low storage capacity, low processing capability, and low energy. While the server considers having full capabilities in terms of storage, processing, computational, and energy. Most of the literature reviews in mutual authentication schemes go through complicated authentication acknowledge messages exchange among devices and server. The long authentication negotiation effect dramatically in the network bandwidth and considered as computational overhead. One more thing this authentication negotiation does not include any messages exchange. Therefore, the messages exchange considered in the next phase will occur after concluding from the authentication phase successfully. The proposed scheme aims to mitigate those long authentication negotiation and message exchange with low computational overhead and considering message exchange within an authentication phase. However, the proposed scheme comprises to two-phase; registration phase, authentication and message exchange phase. The registration phase deemed as an initial phase where preshared keys are shared among IoT devices and the server, assuming that the IoT devices are registered to the server in a secure manner. How to secure the registration phase is beyond current study. While the authentication and message exchange process will be done in the second phase.

3.1 *Lightweight Mutual Authentication*

As mentioned, the proposed scheme aims to provide lightweight mutual authentication between IoT devices through the server while achieving message exchange between IoT devices and the server without bearing any additional hand-check messages or delay message exchange process to the next phase after ensuring that the mutual authentication has been done successfully. With the next subsection, we will describe the proposed scheme in details. Figure 2 shows the mutual authentication and message exchange between the IoT devices through the server.

IoT Device Id_i Side:

1. The IoT device Id_i generated random nonce number N_i .
2. Encrypt the Id_i with P_s .
3. Encrypt the message with GS_k .
4. Calculate the MAC for the plain message M .
5. Send M_1 to the Id_j .

where $M_1 = \{E_{P_s}\{Id_i\} || E_{GS_k}\{M\} || MAC || N_i\}$.

IoT Device Id_j Side:

1. The IoT device Id_j send M_2 to the S

where $M_2 = \{E_{P_s}\{Id_j\}\}$

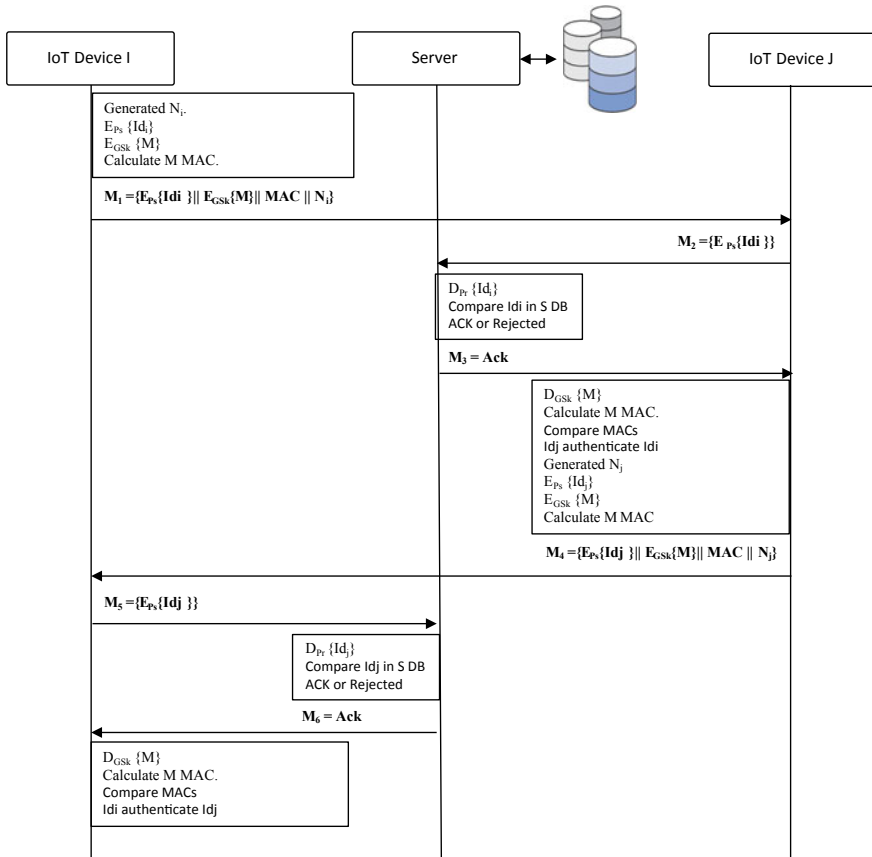


Fig. 2 Mutual authentication and messages exchange

Server Side:

1. The server decrypts the Id_i using its P_r .
2. The server checks the Id_i if it exists within Server SDB.
3. If the Id_i exists, the server will acknowledge it.
4. If the Id_i not exists, the server will not acknowledge it. Further, the server will reject this Id_i and add it to the server block list DB (bDB).
5. Finally, the server sends M_3 to the IoT device Id_j .

where $M_3 = Ack$.

IoT Device Id_j Side:

1. After the S acknowledge the Id_i . The IoT device Id_j decrypts the (CM) using G_{S_k} .
2. After decrypting the (CM) using G_{S_k} . The IoT device Id_j will get the plain text (real message (M)).

3. The IoT device Id_j calculates the (MAC) for the M .
4. Finally, the IoT device Id_j compare the (MACs) if they are equal that means the IoT device Id_i is authenticated and ensure of the message integrity. After that the Id_j will perform the following steps.
5. The IoT device Id_j generated random nonce number N_j .
6. Encrypt the Id_j with P_s .
7. Encrypt the message with GS_k .
8. Calculate the MAC for the message M .
9. Send M_4 to the Id_i .

where $M_4 = \{E_{P_s}\{Id_j\} || E_{GS_k}\{M\} || MAC || N_j\}$

IoT Device Id_i Side:

1. The IoT device Id_i send M_5 to the S

where $M_5 = \{E_{P_s}\{Id_j\}\}$

Server Side:

The server will perform the following steps:

1. The server decrypts the Id_j using its P_r .
2. The server checks the Id_j if it exists within Server SDB.
3. If the Id_j exists, the server will acknowledge it.
4. If the Id_j not exists, the server will not acknowledge it. Further, the server will reject this Id_j and add it to the server block list DB (bDB).
5. Finally, the server sends M_6 to the IoT device Id_j .

where $M_6 = \text{Ack}$.

IoT Device Id_i Side:

1. The IoT device Id_i decrypts the (CM) using GS_k .
2. After decrypting the (CM) using GS_k . The IoT device Id_i will get the plain text (real message (M)).
3. The IoT device Id_i calculates the (MAC) for the M .
4. Finally, the IoT device Id_i compares the (MACs) if they are equal that means the IoT device Id_j is authenticated. Figure 3 shows the flow chart for the proposed protocol.

4 Discussion and Security Analysis

From the aforementioned, we claim that the IoT devices requirements in terms of; computational capability, storage capacity, and energy consumption have been considered and tailored well within the proposed schemes. In addition, different security properties have been achieved. General speaking, we can claim that the following security properties have been guaranteed:

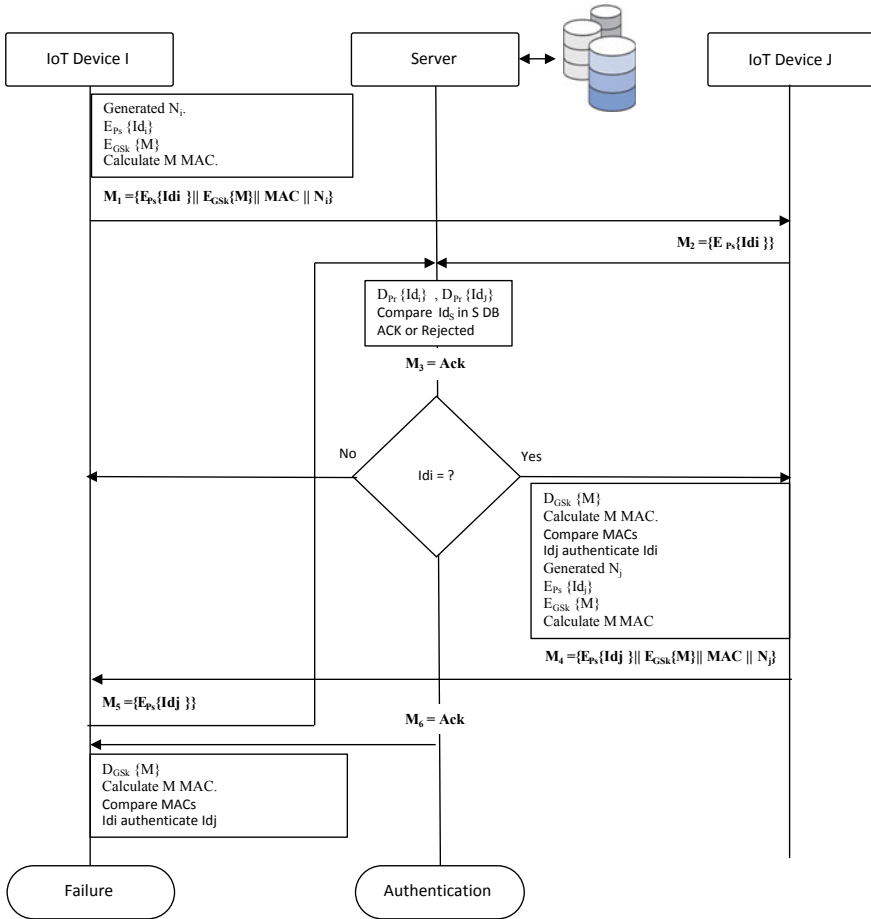


Fig. 3 Flow chart for the mutual authentication and messages exchange

1. Authentication

- a. As the GS_k only known for both IoT devices and the server, therefore, we can claim that the IoT devices authenticate through the server and ensure that the M is encrypted using the group symmetric key (GS_k).

2. Authorization

- a. By decrypting the message with the GS_k , we ensure that the only authorized party can have access to the decrypted message.

3. Confidentiality

- a. The message has been transmitted over the network in ciphertext form, not in plain-text form. Therefore, even the IoT device still not authenticate from the

server. The reception for the transmitted message can not infer or disclosure the content of the message. So, we can claim that confidentiality has been achieved.

4. Data Integrity

- a. By calculating the MAC for both sides in the IoT devices side and comparing the MACs, we can claim that the proposed scheme achieves the data integrity.

5. Replay attack

- a. By generating a random nonce number for each different session. The adversary ability to attach the proposed scheme has been eliminated.

6. Message exchange

- a. To the best of our knowledge, this is the first paper that introduces the idea of transmitting the message within the authentication process. Most of the literature reviews in the domain consider message exchange at next phase after completing the authentication phase. In the proposed scheme, it is clear that the message is transmitted in ciphertext form in plain text to ensure that only the authorized party has the capabilities to decrypt it.

5 Conclusion

In this paper, a secure lightweight mutual authentication and message exchange scheme for IoT devices have been proposed. In the proposed scheme, hybrid cryptographic has been utilized; symmetric cryptographic is adopted in registration phase where asymmetric cryptographic is adopted in authentication and message exchange phase. The server deems as a mediator between IoT devices to authenticate the devices to each other. The main idea is to mitigate the storage capacity, computation cost, and energy consumption for the constrained devices to achieve the main security properties and resist against different types of attacks. The experimental results show that we can achieve mutual authentication and message exchange in one phase without bearing the heavy computational cost, complicated acknowledge messages, and without the need for additional storage capacity. In our future R&D work, we intend to deploy the proposed scheme in a real-life IoT device and server experimental environment to test and validate its actual security levels against a system of criteria for practicability and system-wide performance at both IoT devices and servers ends.

References

1. M. Chen, J. Wan, F. Li, Machine-to-machine communications: architectures, standards and applications (2012)
2. G. Sharma, S. Bala, A.K. Verma, Security frameworks for wireless sensor networks-review. *Proc. Technol.* **6**, 978–987 (2012)
3. X. Xiaokang, D.S. Wong, X. Deng, TinyPairing: a fast and lightweight pairing-based cryptographic library for wireless sensor networks, in *2010 IEEE Wireless Communication and Networking Conference* (IEEE, 2010)
4. O. Delgado-Mohatar, A. Fúster-Sabater, J.M. Sierra, A light-weight authentication scheme for wireless sensor networks. *Ad Hoc Netw.* **9**(5), 727–735 (2011)
5. M. Sangeetha, M. Jagadeeswari, Design and implementation of new lightweight encryption technique. *Int. J. Innov. Res. Sci. Eng. Technol.* (2016)
6. T. Xu, J.B. Wendt, M. Potkonjak, Security of IoT systems: design challenges and opportunities, in *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design* (IEEE Press, 2014), pp. 417–423
7. K.H. Wang, C.M. Chen, W. Fang, W. Tsu-Yang, On the security of a new ultra-lightweight authentication protocol in IoT environment for RFID tags. *J. Supercomput.* **74**(1), 65–70 (2018)
8. P. Gope, R. Amin, S.H. Islam, N. Kumar, V.K. Bhalla, Lightweight and privacy-preserving RFID authentication scheme for distributed IoT infrastructure with secure localization services for smart city environment. *Futur. Gener. Comput. Syst.* **83**, 629–637 (2018)
9. B.B. Gupta, M. Quamara, An identity based access control and mutual authentication framework for distributed cloud computing services in IoT environment using smart cards. *Proc. Comput. Sci.* **132**, 189–197 (2018)
10. D. He, N. Kumar, N. Chilamkurti, J.H. Lee, Lightweight ECC based RFID authentication integrated with an ID verifier transfer protocol. *J. Med. Syst.* **38**(10), 116 (2014)
11. Y.P. Liao, C.M. Hsiao, A secure ECC-based RFID authentication scheme integrated with ID-verifier transfer protocol. *Ad Hoc Netw.* **18**, 133–146 (2014)
12. G.U. Devi, E.V. Balan, M.K. Priyan, C. Gokulnath, Mutual authentication scheme for IoT application. *Indian J. Sci. Technol.* **8**, 26 (2015)
13. K. Fan, Y. Gong, C. Liang, H. Li, Y. Yang, Lightweight and ultralightweight RFID mutual authentication protocol with cache in the reader for IoT in 5G. *Secur. Commun. Netw.* **9**(16), 3095–3104 (2016)
14. A. Tewari, B.B. Gupta, A lightweight mutual authentication approach for RFID tags in IoT devices. *Int. J. Netw. Virtual Organ.* **18**(2), 97–111 (2018)

Monetary Transaction Fraud Detection System Based on Machine Learning Strategies



Lakshika Sammani Chandradeva, Thushara Madushanka Amarasinghe, Minoli De Silva, Achala Chathuranga Aponso and Naomi Krishnarajah

Abstract Fraud is a costly business problem which causes every organization to face huge loss. Fraud may lead to risk of financial loss and loss of the confidence of customers and stakeholders of the company. Cyber security teams and internal audit departments of most of the organizations try to monitor such fraudulent activities using traditional rule-based fraud detection systems. However, with the rapid adaptation of online financial transactions, it is more difficult to identify fraudulent activities by static methods and via data analysis. Further, as traditional rule-based fraud detection systems cannot dynamically adjust the rule set based on the behavioral changes of the fraudsters, there is a high possibility of detecting false positive alerts. The aim of this paper is to review selected machine learning techniques where it can be used to develop a fraud detection system which identifies fraudulent activities in financial transactions.

Keywords Machine learning · Artificial neural network · Convolutional neural network · Recurrent neural network · Bayesian belief network · Hidden Markov model · Support vector machine and decision trees · K -nearest neighbor (K -NN) · Hidden Markov model (HMM)

L. S. Chandradeva (✉) · T. M. Amarasinghe · M. De Silva · A. C. Aponso · N. Krishnarajah
Department of Computing, Informatics Institute of Technology, University of Westminster,
Colombo, Sri Lanka
e-mail: lakshika.2014525@iit.ac.lk

T. M. Amarasinghe
e-mail: thusharaaatm@gmail.com

M. De Silva
e-mail: minoli.d@iit.ac.lk

A. C. Aponso
e-mail: achala.a@iit.ac.lk

N. Krishnarajah
e-mail: naomi@iit.ac.lk

1 Introduction

Financial transaction fraud has been discovered to pose a big threat on financial organizations. Most of the finance sector companies such as banks and other non-banking financial sector companies and institutes presently try to monitor and identify such fraudulent activities by using traditional rule-based fraud detection systems as frauds are not being identified in real time. Typical fraud detection systems use various attributes such as time, account number, card number, transaction type, amount, gender, age, and country for building the rules [1]. However, the main limitations of having rule-based fraud detection system are the active management and monitoring of the rule set required, the rule set being fixed, and rules being based on the human errors and bias. Rule-based fraud detection systems can be effective only when they are monitoring and managing real time, which means fraud detection systems require to undergo scheduled periodic reviews and modifying rules based on the past results. Rule-based fraud detection systems work without any intelligence, and they check only whether the rules criteria are met or not. Those rule-based systems do not contain the ability to dynamically adjust the rules based on the behavioral changes, thus leading to many false positive results. Further, incorrect and poorly defined rules also result in both high fraud rates and high false positive results [2]. Nowadays, fraudsters use anonymous ways to gain access to customers' accounts, customers' personal details, or financial details to enable the criminals to commit fraud. In such cases, rule-based systems and human reviews would fail to block transactions. Frauds are one of the high risks and most common crimes in the world which require monitoring and prevention at an early stage. And expert systems have been found to be the best solution to prevent frauds in the future [3].

2 Background

Due to the rapid increase in ATM transactions, E-commerce transactions, and POS transactions, use of payment cards and online payments have increased and it, in turn, has caused the rapid growth of the financial transaction fraudulent activities. Traditional fraud detection systems which are based on the database systems and customers' knowledge level are usually inaccurate, not real time, and provide delayed results. Then, fraud detection methods like discriminate analysis and regression analysis came into the picture. However, those mechanisms were also not efficient for a large amount of data [4]. Most common intrusion detection systems are based on the signatures of the events. These systems are developed to detect attacks based on the signatures of events where the signatures of the events meet the defined rule criteria. Nevertheless, these systems require periodic reviews and updates on both rules and signatures. Further, these systems generate high percentage of false positive alarms and are unable to scale to gigabit speeds [5]. Effectiveness of using static rules for fraud detection is low. To obtain maximum effectiveness from such a rule-based

system, it should have the ability to depend on any input data. Further, such systems should be user-friendly and able to capture the most complex rules [6]. Furthermore, machine learning has been a tool in resolving most of the important business problems like fraud detection. Fraud management became one of the hardest duties of the banking and finance industry. Machine learning uses several complex algorithms that crawl over large datasets and analyze patterns within the data. Machine learning mechanisms are better than humans at processing large sets of data. When using machine learning mechanisms for fraud detection, not using appropriate data could cause the model to learn the wrong assumptions and make irrelevant assessments on frauds. Further, extracting and training data for accurate predictions is a tough task [7]. Current fraud detection systems have limited purview into datasets, and it causes to limit the ability of taking accurate decisions. Therefore, there is high possibility of detecting false positive events when using machine learning techniques.

3 Rule-Based Techniques

According to Kou et al. [8], credit card fraud can be divided into two types—online and offline frauds. Further, the authors mentioned that offline frauds can happen only by stealing physical card, while online fraud can happen via Web and only the card details are required. Tova Milo, Slava Nogorodov, and Wang-Chiew Tan stated that financial companies try to employ domain experts to manually specify rules that exploit general or domain knowledge to improve fraud detection process. Further, with the time, rules need to be updated and refined to capture the evolving activity patterns of the financial transactions [9]. Bart Baesens mentioned that currently available fraud detection rules can be refined by studying past transactions to detect future cases or financial transactions and trigger and alert when fraud is committed. But such approaches have so many disadvantages. Rule-based systems are expensive to build as they require advanced rule refinements by employing fraud experts and are also hard to maintain. As rule-based detection systems are based on the past events, they cannot guarantee that the new fraud patterns will be alerted. However, frauds are dynamic incidents and need to be traced continuously [10].

4 Machine Learning Techniques

4.1 *Artificial Neural Network*

Amarasinghe and Aponso [11] mentioned that artificial neural network has three layers and weights of each hidden layers have been initialized to numbers which are close to zero. Further, researchers used sigmoid activation function to calculate the probability of transactions being fraudulent or not and artificial neural networks

Table 1 Accuracy measures

Measure	Description	Formula
Accuracy	Represents the accuracy level of the classifier	$(TP + TN)/total$
F1 score	Shows the harmonic mean of precision and recall	$2TP/(2TP + FP + FN)$
Precision	Shows the probability of predicting true false from all positive predictions	$TP/(TP + FP)$
Recall	Shows the true positive rate	$TP/(TP + FN)$

provide a binary output to show whether the transaction is fraudulent or not. They used dropout regularization to reduce the overfitting to address the high variance problem. For evaluation, these researchers used K -fold cross-validation to address the bias variance tradeoff. Various accuracy measures are required due to the unbalancing nature of the data, and Table 1 describes those accuracy measures.

Mishra et al. [12] proposed an architecture for credit card using an artificial neural network type called feed forward neural network. In this research, accuracy is calculated by comparing actual output in the dataset and predicted output of the model after the simulation. Further, they discussed three learning techniques, namely BR, GDA, and LM to train the model for credit card fraud detection and finally concluded that Bayesian regularization (BR) technique is high in accuracy and performance level. Further, they concluded that Bayesian regularization (BR) technique is the better approach to train the multi-layer feed forward back propagation neural network for credit card fraud detection. Gulati et al. [13] identified a method for credit card fraud detection using neural network and geo-location. The proposed system classifies transactions into two types such as suspicious and non-suspicious transactions. First, the system checks the geo-location and pattern of spending money using credit card. Then, if there is any mismatch with the normal behavior of the transactions, then those transactions are subject to a verification process by communicating with the customers. There is a high possibility of getting false positive alarms as the accuracy of the system is based on the geo-location and the pattern. This research used public IP lookup API to verify the geo-location of the customer.

4.2 Convolutional Neural Network

Zhang et al. [14] researched about the model-based convolutional neural network to detect frauds in online transactions, and this constructed model performs a restructuring of raw transaction features to form various convolutional patterns. Further, this model reduces the calculation time of derived variables. Krishna [15] proposed to use LeNET architecture of convolutional neural network as a fraud detection technique in credit card transactions. Further, this researcher mentioned that precision performance in neural networks (NN) is better than CNN, because ability of detecting fraudulent transactions as legitimate transactions is high on neural networks (NN).

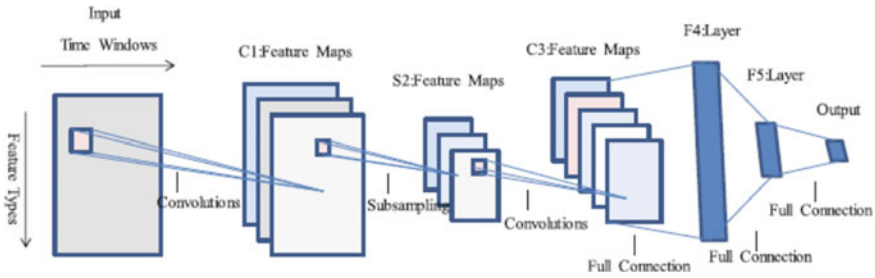


Fig. 1 LeNET architecture

But detecting fraudulent transactions is better on CNN than NN, and it improves recall and the harmonic mean of the precision and recall (Fig. 1).

Chouiekh and Haj [16] found that deep convolutional neural network (DCNN) performed well in fraud detection in comparison with the other traditional machine learning algorithms when considering the accuracy and the training duration. Further, this method can reduce the cost of illegal use of services without payments. This model is developed to identify normal behavior and fraudulent behavior separately.

4.3 Recurrent Neural Network

Ando et al. [17] used recurrent neural network to detect fraudulent behaviors of credit card transactions. In this research, they proposed the approach which uses the structure in Web access logs to detect frauds. This proposed RNN consists of three layers, namely input layer, hidden layer, and the output layer. They used BPTT algorithm of RNN for their research as it can learn fast due to the simple structure. Nevertheless, it is difficult to train the RNN model for large time series data. They used two types of RNN for testing as RNN with LSTM and RNN without LSTM, with sigmoid function. Nevertheless, they found that accuracy of detecting fraudulent behavior is higher in RNN than support vector machines (SVM) and RNN with LSTM meets the required results faster than the RNN without LSTM. Wang et al. [18] used RNN in order to detect session-based fraud in E-commerce transactions. They used RNN to analyze the sequence of clicks in a session for fraud detection in financial transactions. They fed all the clicks of a session into the RNN model based on the time order and derived the risk score at the final click for each session. Then, it shows whether the session is suspicious or not. Further, they also used LSTM for this research to categorize the time dependency of the prediction on the previous clicks. They performed undersampling mechanism in order to balance normal and fraudulent classes. Further, they used TensorFlow module to train the proposed RNN model. These researches evaluate the performance of this RNN model using different embeddings, different RNN structures, and RNN cells.

4.4 Bayesian Belief Network

Maes et al. [19] tested and compared credit card fraud detection using Bayesian belief network (BNN) and artificial neural network (ANN). They used STAGE algorithm for BNN model. Finally, they compared results generated from both the models BNN and ANN and concluded that BNN performs better than ANN in credit card fraud detection. Learning time is also lower in BNN than ANN, and evaluation is high in ANN than BNN.

4.5 Support Vector Machine and Decision Trees

Chen et al. [20] investigated prediction accuracy of questionnaire-responded transaction (QRT) approach by using SVM. Hierarchical SVMs, oversampling, and majority voting techniques also get involved in order to explore their impact to the prediction accuracy. They mentioned that SVM is a strong machine learning mechanism for classifying and doing regression. They used SVM to train and build the classifier model and detect anomalous data. Testing results showed that this QRT approach with SVM is effective and effectiveness can be further increased by combining with mechanisms like hierarchical SVMs, majority voting, weighting, and voting. Nipane et al. [21] used support vector machine (SVM) and decision tree classifier for identifying and categorizing unauthorized users and decision trees were used for fraud handling behavior of the users. SVM is a binary classifier which helps to alert the user whether the transaction is fraudulent or not. In this proposed system, SVM monitors the behavioral patterns of the user and decision trees identify whether the user behavior is anomalous or not. They observed that prediction accuracy showed by SVM gradually decreases when the transaction record increases. Further, they identified that false positive rate is high, and the accuracy level of the proposed system is 59%.

4.6 Hidden Markov Model (HMM)

Dhok and Bamnote [22] proposed hidden Markov model for credit card fraud detection. This researcher mentioned that HMM-based approach reduces the false positive rate which identifies transactions as fraudulent even though they are genuine. HMM detects fraud activities based on the customer's spending time and the number of items purchased with use of further processing. Further, it tries to identify the variance of spending time of the card owner and verify the details like billing address, shipping address, etc. Further, this system calculates total price based on the expenditure history of the cardholder and it compares with the current total price. Also, this HMM model decides prices dynamically based on the clustering algorithms. The proposed model verifies the transactions real time, and it includes two models called

online shopping and fraud detection system. Further, this HMM model also has the capability of handling large sets of transactions. Mhatre et al. [23] also proposed HMM for credit card fraud detection. According to their work, this fraud detection system starts working after the first ten transactions of the cardholder. In this scenario also, when user needs to do a transaction, this model analyzes the spending profile of the user and decides if transactions are fraudulent or not based on the transition probabilistic calculation. Further, hidden Markov model does not need fraud signatures. The proposed model has a database of saved past transactions and unusual transactions. Bhusari and Patil [24] also described the application of hidden Markov model in credit card fraud detection. They also stated that HMM is most probably the easiest model which is able to detect credit card frauds without using the fraud signatures. They further mentioned that the benefit of the HMM model is the reduction of the false positive transaction rate. To analyze this system, these researches defined three price ranges called as high, medium, and low. Those three defined parameters regulate in a training phase using forward–backward algorithm. This algorithm converges all the values starting from initial HMM parameters. Further, this model estimates HMM parameters for every card owner. If the system identified a genuine transaction, then it will consider it for future fraud detection.

4.7 K-Nearest Neighbor (K-NN)

Venkata [25] proposed to use standing outlier detection based on reverse K -nearest neighbors (SODRNN) algorithm for credit card fraud detection. Researcher stated that this proposed method can stop fraudulent activities of stolen credit cards and it detects errors when checking for the credit card validation. Malini and Pushpa [26] stated that K -NN gives better results in credit card fraud detection system using supervised machine learning. K -NN-based credit card fraud detection systems need to assume distance and the similarity measure between two data instances. Further, this method is fast and false positive alert ratio is low.

5 Other Techniques

Zanin et al. [27] showed the possibility of using complex networks for credit card fraud detection by using parenclitic network analysis. Parenclitic network is a reconstruction mechanism that shows the differences between one instance and set of standards. Further, they conclude that this mechanism alone is not sufficient to reach a low classification error. Stolfo et al. [28] used meta-learning for credit card fraud detection. They mentioned that meta-learning is an integration mechanism which works by combining various separately learned classifiers or models. This proposed system will allow banks and other financial organizations to share their fraudulent financial transaction models by exchanging in classifier agents in a secured infrastructure.

6 Algorithmic Analysis

Table 2 describes advantages and disadvantages of algorithms used on selected existing researches.

7 Limitations and Future Scope

During this review money laundering activities, frauds related to loans and lease were not discussed. Financial transactions are considered as sensitive information of financial organizations. Therefore, it is difficult to obtain original data from a financial organization in order to test the model. Hence, modified and masked data will be used to test and evaluate the model. Further, when using machine learning algorithms in order to detect fraudulent activities, there is a high possibility of detecting false positive financial transactions. Most of the currently available systems can be used when only the labeled training data is readily available. But in real-world scenarios, it is hard to find labeled dataset to train the existing systems. Therefore, existing systems can be improved to support unsupervised machine learning fraud detection.

8 Summary

The aim of this paper is to evaluate currently available fraud detection systems and identify the gaps of used mechanisms. Banks and financial institutions have a special concern about financial transaction frauds. Information security teams and internal audit teams have the requirement of having a real-time financial transaction fraud detection system.

9 Conclusion

Fraud is one of the major threats to financial institutes, which may lead to loss to the institutions both financially and reputational. Nevertheless, research has identified that analyzing financial transactions using traditional methods is not a proper solution since. Traditional fraud detection solutions are less accurate, have difficulties in handling large sets of financial transaction logs and incur high operational costs. These factors, together with a lack of skilled IT specialists and increased time consumption for investigations, have a negative impact on efficient and effective financial transaction fraud detection. Then, rule-based fraud detection systems came into the picture. Those too contained limitations such as inability to identify unknown frauds and difficulty to manage and implement complex rules to identify some complex fraudulent

Table 2 Algorithmic analysis

Algorithm	Pros	Cons
Artificial neural network	ANN provides an output to show whether the transaction is fraudulent or not. And, accuracy is high [11–13]	ANN model tuning takes more time [11–13]
Convolutional neural network	According to Zhang et al. [15] and Chouiekh and Haj [16], this approach can save variable derived time. CNN can accurately classify images. Convolutional layer can derive new features based on the existing features	According to Zhang et al. [15] and Chouiekh and Haj [16], performance is slow when compared to BP neural network
Bayesian belief network	As mentioned by Maes et al. [19], training period is short and gives better results concerning fraud detection	Further Maes et al. [19] stated that detection process is slow in this approach
Recurrent neural network	Ando et al. [17] and Wang et al. [18] stated that this approach has simple structure and it can learn fast	And, according to Ando et al. [17] and Wang et al. [18], it is difficult to train the RNN model for huge time series data
Hidden Markov model	False positive rate is low. Easiest model which is able to detect credit card frauds without using the fraud signatures [22–24]	Model should have at least 10 initial transactions to start detecting frauds [22–24]
K-nearest neighbor	This method is fast, and false positive ratio is low [25, 26]	Need to use supervised machine learning to get better results. Supervised machine learning model needs a classified dataset in order to train the model. But classified transaction datasets are hard to find [25, 26]

(continued)

Table 2 (continued)

Algorithm	Pros	Cons
<p>Parenclic network analysis</p>	<p>Zanin et al. [27] mentioned the possibility of using complex networks for credit card fraud detection. Further, network-based model can get better results than a commercial fraud detection system</p>	<p>According to the Zanin et al. [27], features which extracted only from the parenclic networks are not sufficient to reach a low classification error. Further, this algorithm fails for large transactions and less in efficiency</p>
<p>Meta-learning</p>	<p>Stolfo et al. [28] stated that this mechanism produces a good overall performance</p>	<p>Stolfo et al. [28] stated that different machine learning mechanisms need to combined to get better results</p>
<p>Support vector machine and decision trees</p>	<p>Efficiency of this approach is high. SVM is a strong mechanism to classify and do regression [20, 21]</p>	<p>But [20, 21] mentioned that false positive rate is high</p>

activities. Then, the next step of the fraud detection industry became the fraud detection using machine learning techniques. The above literature review identified that there are also some drawbacks. Taking more time for tuning the machine learning model, low performance level, difficulty of training the huge time data series, less efficiency, and the high false positive rate are the gaps that identified from the above. Therefore, current fraud detection field has a requirement of a financial transaction fraud detection system which can have an ability to bear large transactions with the low false positive rate and efficiency, accuracy, and the performance are high.

Acknowledgements I would like to express my sincere gratitude to everyone who were behind me from the beginning to the completion of the paper. Firstly, I am very much thankful to my supervisor Mr. Achala Chathuranga Aponso for the unwavering support, guidance, and insight throughout this period. I acknowledge with thanks for the encouragement received from Ms. Naomi Krishnarajah, Dean of the Informatics Institute of Technology. Special thanks to Mr. Thushara Madushanka Amarasinghe and Ms. Minoli De Silva for all the support and ideas. Last but not least, I would like to express my love and gratitude to all my family members, especially to my father and mother for the constant support toward my studies and for believing me. I am sure that this would have not been possible without the valuable contribution of each one of you.

References

1. A.C. Bahnsen et al., Feature engineering strategies for credit card fraud detection. *Expert Syst. Appl.* **51**, 134–142 (2016). <https://doi.org/10.1016/j.eswa.2015.12.030>
2. K. Bui, *4 Reasons Why Fraud Prevention Needs to Move Beyond Rules Based Engines* (Feedzai, 2016). Available at: <https://feedzai.com/blog/4-reasons-why-fraud-prevention-needs-to-move-beyond-rules-based-engines/>. Accessed: 3 Nov 2018
3. R. Zainal, A. Md. Som, A review on computer technology applications in fraud detection and prevention. *ResearchGate* (2017). Available at: https://www.researchgate.net/publication/323392316_A_REVIEW_ON_COMPUTER_TECHNOLOGY_APPLICATIONS_IN_FRAUD_DETECTION_AND_PREVENTION
4. B. Vijay, J. Swathi, Credit card fraud detection analysis. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **4**, 4 (2016)
5. A. Patcha, J.-M. Park, An overview of anomaly detection techniques: existing solutions and latest technological trends. *Comput. Netw.* **51**(12), 3448–3470 (2007). <https://doi.org/10.1016/j.comnet.2007.02.001>
6. J. McGibney, S. Hearne, An approach to rules based fraud management in emerging converged networks (2004)
7. How Machine Learning Facilitates Fraud Detection? *Maruti Techlabs*, 5 June 2017. Available at: <https://www.marutitech.com/machine-learning-fraud-detection/>. Accessed: 2 Nov 2018
8. Y. Kou et al. Survey of fraud detection techniques, in *IEEE International Conference on Networking, Sensing and Control, 2004. IEEE International Conference on Networking, Sensing and Control, 2004* (IEEE, Taipei, Taiwan, 2004), pp. 749–754. <https://doi.org/10.1109/icnsc.2004.1297040>
9. T. Milo, S. Novgorodov, W.C. Tan, Interactive rule refinement for fraud detection. *OpenProceedings.org* (2018). <https://doi.org/10.5441/002/edbt.2018.24>
10. B. Baesens, V. Van Vlasselaer, W. Verbeke, *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques* (Wiley, Hoboken, NJ, 2015)
11. T.M. Amarasinghe, A.C. Aponso, Fraud detection solution for financial transactions with artificial neural network (2018)

12. C. Mishra, D.L. Gupta, R. Singh, *Credit Card Fraud Identification Using Artificial Neural Networks* (no date), p. 9
13. A. Gulati et al., Credit card fraud detection using neural network and geolocation, in *IOP Conference Series: Materials Science and Engineering*, vol. 263 (2017), p. 042039. <https://doi.org/10.1088/1757-899x/263/4/042039>
14. Z. Zhang et al., A model based on convolutional neural network for online transaction fraud detection. *Secur. Commun. Netw.* (2018). <https://doi.org/10.1155/2018/5680264>
15. Fraud Detection Technique in Credit Card Transactions using Convolutional Neural Network. ResearchGate (no date). Available at: https://www.researchgate.net/publication/321383884-Fraud_Detection_Technique_in_Credit_Card_Transactions_using_Convolutional_Neural_Network. Accessed: 5 Oct 2018
16. A. Chouiekh, E.H.I.E. Haj, ConvNets for fraud detection analysis. *Proc. Comput. Sci.* (2018). <https://doi.org/10.1016/j.procs.2018.01.107>
17. Y. Ando, H. Gomi, H. Tanaka, in *Detecting Fraudulent Behavior Using Recurrent Neural Networks* (2016), p. 6
18. S. Wang et al., Session-based fraud detection in online E-commerce transactions using recurrent neural networks, in *Machine Learning and Knowledge Discovery in Databases*, ed. by Y. Altun, et al. (Springer, Cham, 2017), pp. 241–252. https://doi.org/10.1007/978-3-319-71273-4_20
19. S. Maes, T. Karl, V. Bram, Credit card fraud detection using bayesian and neural networks. ResearchGate (2002). Available at: https://www.researchgate.net/profile/Karl_Tuyls/publication/248809471_Credit_Card_Fraud_Detection_Applying_Bayesian_and_Neural_networks/links/0deec52519708c5f7a000000/Credit-Card-Fraud-Detection-Applying-Bayesian-and-Neural-networks.pdf
20. R. Chen et al., Novel questionnaire-responded transaction approach with SVM for credit card fraud detection (2005)
21. V.B. Nipane et al., Fraudulent detection in credit card system using SVM & decision tree. *Int. J. Sci. Dev Res. (IDS DR)* **1**(5), 5 (2016)
22. S.S. Dhok, D.G.R. Bamnote, Credit card fraud detection using hidden Markov model. *Int. J. Adv. Res. Comput. Sci.* **5**(1), 37–48 (2010)
23. G. Mhatre et al., Credit card fraud detection using hidden markov model. *Int. J. Comput. Sci. Inf. Technol. (IJCSIT)* **5**, 3 (2014)
24. V. Bhusari, S. Patil, Application of hidden markov model in credit card fraud detection. *Int. J. Distrib. Parallel Syst.* **2**(6), 203–211 (2011). <https://doi.org/10.5121/ijdp.2011.2618>
25. Credit card fraud detection using anti k -nearest algorithm. ResearchGate (no date). Available at: https://www.researchgate.net/publication/236962626_credit_card_fraud_detection_using_anti_k-nearest_algorithm. Accessed: 5 Oct 2018
26. N. Malini, M. Pushpa, Analysis on credit card fraud identification techniques based on KNN and outlier detection, in *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (2017), pp. 255–258 <https://doi.org/10.1109/aeecb.2017.7972424>
27. M. Zanin et al., Credit card fraud detection through parenclitic network analysis. *Complexity* (2018). <https://doi.org/10.1155/2018/5764370>
28. S.J. Stolfo et al., Credit card fraud detection using meta-learning: issues and initial results (no date), p. 8

The Impact of Service-Level Agreement (SLA) on a Cloudlet Deployed in a Coffee Shop Scenario



Mandisa N. Nxumalo, Matthew O. Adigun and Ijeoma N. Mba

Abstract The usage of service-level agreement (SLA) to ensure smooth negotiations, easy management, and provision of resources make it a reliable tool to ensure service guarantees in a network. Low-latency Internet connectivity services provided by edge servers such as a Cloudlet deployed in small–medium enterprises (SMEs) can benefit from SLA. This is due to that SMEs are exposed to financial issues and through the use of SLA, SMEs such as a coffee shop can be able to attract consumers plus minimize Cloudlet operational cost. The operational cost is minimized through using the SLA to manage and provide effective resources which eliminate SLA violations in the network. The decrease in SLA violation occurrence has a positive effect on the operational cost SMEs must spend to lease resources and Internet packages from providers. The objective of this study was to demonstrate the effectiveness of SLA in minimizing Cloudlet operational costs using a coffee shop scenario. This was achieved through the use of a CloudSim Plus tool to simulate the SLA in a network. The results show that the management of resources in the network and the monitoring of the agreed SLA eliminate SLA violations and further minimizing operational cost.

Keywords Service-level agreement · Cloudlet · Small–medium enterprises · CloudSim plus tool

M. N. Nxumalo (✉) · M. O. Adigun · I. N. Mba
Department of Computer Science, University of Zululand, Richards Bay, South Africa
e-mail: mandisan.mn@gmail.com

M. O. Adigun
e-mail: profmatthewo@gmail.com

I. N. Mba
e-mail: ijaymba@gmail.com

1 Introduction

The increase in latency-sensitive applications such as face recognition, online gaming, and other virtual reality applications has opened a gap for low-latency high-computation edge servers. Mobile cloud computing (MCC) is the combination of mobile and cloud computing to help resource-limited devices (mobile device) to offload computation on the edge without encountering high latency. The Cloudlet is among the example of edge servers introduced by MCC and is of interest to this study. The deployment of a Cloudlet is within a business premise or other user-focused locations [1]. It is defined as a one-hop, resource-rich computer or cluster of computers that is located in the middle of a three-tier (edge devices–Cloudlet–Cloud) architecture to enable edge devices to offload computation using a wireless local area network (WLAN) connection [2]. Among other exceptional merits possessed by a Cloudlet is easy management [2]. It has a decentralized management, which makes deployment easy and it can operate in environments such as shopping centers, restaurants, coffee shops or other SMEs [3]. A Cloudlet deployed in SMEs is in a form of an access point that is Cloudlet-WiFi enabled. Although, SMEs play an important role in the economy by providing more job opportunities, but they are prone to financial issues [4]. The lack of financial support among SMEs makes them to be more dependent to cost-effective service provisioning and policy making. Achieving a cost-effective service provisioning requires a service provisioning policy that outlines the service type, its parameters, role-players, and their roles measured using a cost function. Similar to Cloud-WiFi the billing of Cloudlet-WiFi consumption by consumers is not direct. This implies that consumers consume Cloudlet-WiFi services for free and the owner is responsible for its operational cost. The operational cost of Cloudlet-WiFi includes but is not limited to leasing resources from Cloud providers and internet packages from Internet service providers (ISP) [5]. In some instances, the owners tend to increase the core product costs in order to adapt to operational costs raised by the above scenario. However, the above choice limits quality of experience (QoE) to consumers and can later have bad influence on the quality of business (QoB). By using a service provisioning policy it is envisaged that SMEs can adapt to Cloudlet operational costs.

SLA is a control service provisioning policy used as a means of communication between involved parties in a service negotiation [6]. It specifies and defines the service, quality of service (QoS) metrics to ensure the stated service delivery, price function, service violations, and violation compensation information. It has two parameters namely, service-level objectives (SLOs) and service-level indicators (SLIs) [7]. The SLOs are referred to as the thresholds of the QoS metrics such as availability, task completion time, task waiting time, and other Cloudlet performance metrics. In consideration of the SLOs, the SLI then measures how much of the service level was provided by the Cloudlet Owner to the Cloudlet consumers on the network. In a case where the results obtained from SLIs shows a violation in SLA, compensation must be done for the affected party as agreed in the SLA. The objective of this paper is to demonstrate the effectiveness of SLA in promoting an optimized

Cloudlet operational cost, elimination of SLA violations while ensuring agreed service guarantees. This work is organized as follows: Sect. 2 describes the scenario of choice. In Sect. 3 the simulation setup is covered. Thirdly, Sect. 4 includes an introduction to the results and also provides a descriptive discussion of the results. Related work to this study is discussed in Sect. 5. Lastly, Sect. 6 concludes with work done.

2 Coffee Shop Scenario

A coffee shop is small-to-medium-sized businesses allocated in busy areas namely, work stations, and institutes. It consists of two role-players: the Cloudlet Owner is an SME owner that has an ability to deploy a Cloudlet (Cloudlet-WiFi) in their premises with a purpose to attract customers and improve core product purchase patterns. The second role-player is a coffee shop customer also referred to as a Cloudlet consumer that is willing to share their browser history data in exchange to gain access to a Cloudlet-WiFi. The data sharing is done to enable a Cloudlet to be context-aware of the consumer’s preference [2, 5]. The sharing is referred to as a compensation for consuming a Cloudlet-WiFi services and is captured on the SLA between the agreed parties. The Cloudlet Consumer only gains access to a Cloudlet-WiFi, if they agree to share the browser history data contained on their device’s browser log file. The figure labeled as Fig. 1 shows that when the Cloudlet Owner lease efficient Cloudlet resources to cater for initiated tasks on the network during peak hours it can lead to a QoB. The QoB is measured by the ability of the Cloudlet Owner to ensure

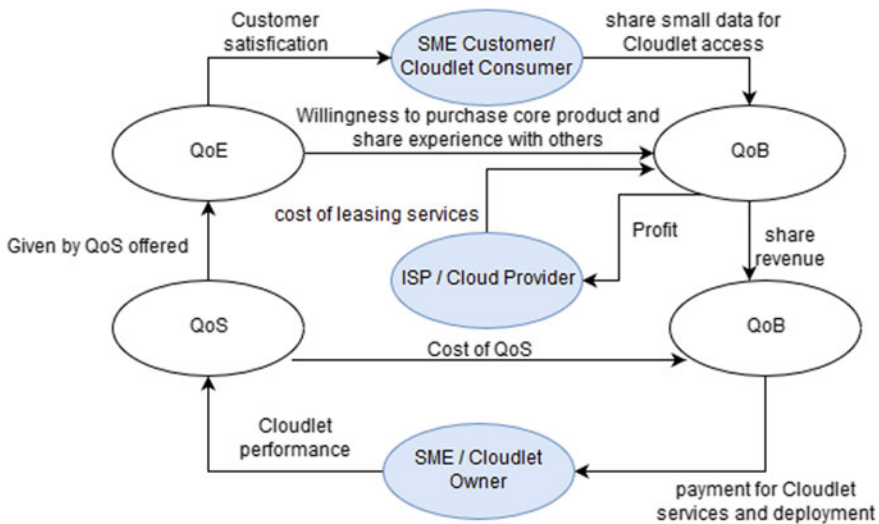


Fig. 1 Business ecosystem for role-players in a coffee shop value chain

service guarantees (QoS) that further results to QoE for Cloudlet Consumers. The QoS measures the performance of the Cloudlet instance and its metrics are found on the SLA between the agreed parties. The failure of a Cloudlet Owner to deliver the stated QoS results to service violations and this can lead to additional costs on the operational cost. To avoid the above, this study considers that in cases of violations a Cloudlet Owner will receive a notification with a detailed fault that caused the violation and possible reasons, how much throughput was achieved in the network, and how much session duration was extended to the failed tasks.

3 Simulation Description and Setup

A CloudSim Plus is an open-source tool used as a simulator for SLA evaluation in this study. It uses Java 8 and CloudSim 3 package to enable the implementation of simulating Cloud-based scenarios [8]. Below are some of the classes included in the tool to ensure an effective use-case simulation:

1. Allocation policies: Provides a mechanism to enable the provisioning of resources in a network such as a VmAllocationPolicySimple policy that is used when the Host with less processor cores is replaced by a VM to improve network stability.
2. Schedulers: Distributes the initiated tasks across multiple VMs in the network. The time-shared and space-shared are types of scheduling algorithms used to scheme tasks to VMs. This simulation uses a time-shared scheduling algorithm, meaning the tasks are processed at the same time.

The tool was pulled from a GitHub repository and imported as a maven project to Netbeans 8.2 IDE, running on Windows 10 64-bit operating system. The CloudSim Plus tool was used due to the lack of simulation tools that enable implementation of management policies such as SLA on the edge. Although a tool such as Edge-CloudSim enables simulation of edge scenarios but currently lacks support of SLA management policies [9]. A Cloudlet in CloudSim Plus tool is considered as mimics of Cloud consumer tasks initiated for processing by the VMs in the network. Therefore, the aim of this simulation is to provide an understanding to the effect of using the SLA to manage and improve resource provisioning in the network. By adopting an SLA, it is envisaged that a deployed Cloudlet in a coffee shop will help the owners to minimize Cloudlet operational cost.

The simulation was configured as per Table 1. Throughout the simulation, the number of Cloudlets, Hosts, and broker entities were kept constant. The only parameter that was continuously changed during the simulation was the number of VMs for processing initiated tasks.

Table 1 Important simulation parameters

Parameter	Content
VMs	10, 15, 20, 25, 30, 35, 40, 45, 50
Number of cloudlets or initiated tasks	150
Number of hosts	20
Number of brokers	1
Number of SLA contract	1
Cloudlet scheduling algorithm	Time-shared

3.1 Simulation Metrics

The simulation took into consideration QoS metrics such as availability and task completion time to ensure efficient network performance. The metrics are parameters of SLA referred to as SLOs and the SLA is written in a JSON format.

1. Task Completion time: Amount of time it takes VMs to process initiated tasks measured in Milliseconds (ms). Threshold for this metric is 100 ms with minimum of 10 ms.
2. CPU utilization: The percentage of power for use by the VMs to process initiated tasks, the threshold is 90%.
3. Availability: is a percentage of available resources in a network to process initiated tasks. The minimum and threshold of availability were set to 100%.
4. Wait time: The amount of time each task is suspended before processing. This was disregarded due to the use of the time-shared scheduling algorithm.
5. Throughput: The total number of tasks processed by the VMs. The minimum and threshold of throughput is 100% to avoid SLA violations.

The aim of the simulation is to address and develop a mechanism to avoid SLA violations in order to ensure service guarantees in the network. A mechanism used to detect SLA violation occurrence in the network was developed. An assumption that if SLA violations occur, a notification with detailed information is sent to the Cloudlet Owner and for unsuccessful tasks, the session duration is extended as shown in Fig. 2 results display.

4 Simulation Results and Discussion

This section provides illustrations (shown by Figs. 3, 4 and 5) and description of the results obtained from the simulation.

The results show that an increase in the number of VMs in the network has a positive effect on the percentage of SLA. The more resources are added, the higher

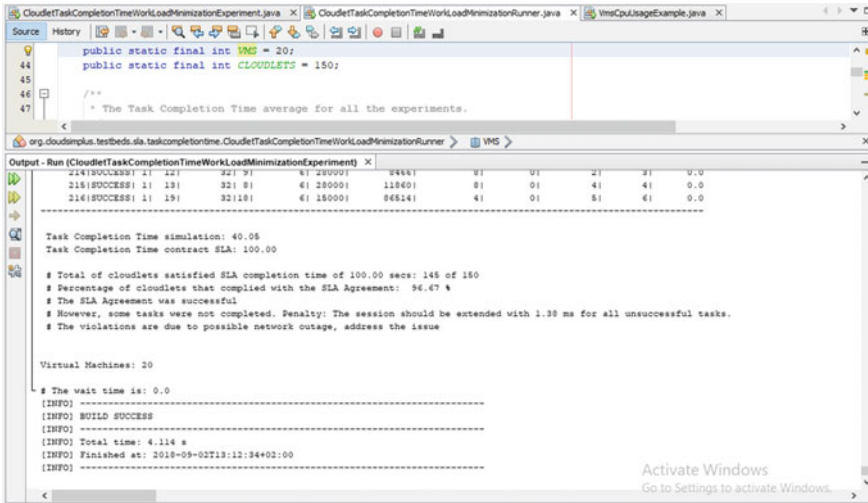


Fig. 2 Overview of SLA simulation results

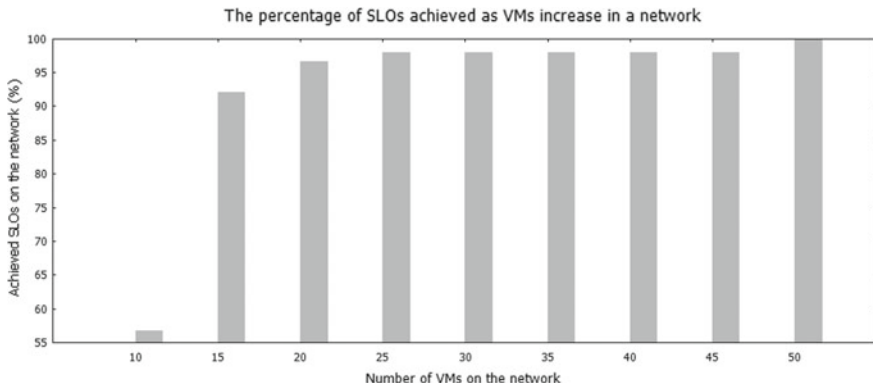


Fig. 3 Effect of increasing the number of VMs to SLOs achieved on the network

the number of SLOs and throughput achieved on the network. The SLA percentage was obtained through comparing the QoS metric values obtained during simulation with that stated on the SLA contract. The comparison process occurs automatically using the attached SLA JSON format file with declared QoS metric thresholds. The thresholds in the file are compared to the final results of the network performance. In a case where one or more of the metrics from the SLA are violated, the SLA percentage decreased as shown by Fig. 3. Figure 3 can give the same pattern of behavior as network throughput, a gradual increase due to the increment of the number of VMs placed in the network. This means that for a coffee shop to ensure connectivity service guarantees to Cloudlet Consumers, requires efficient resources

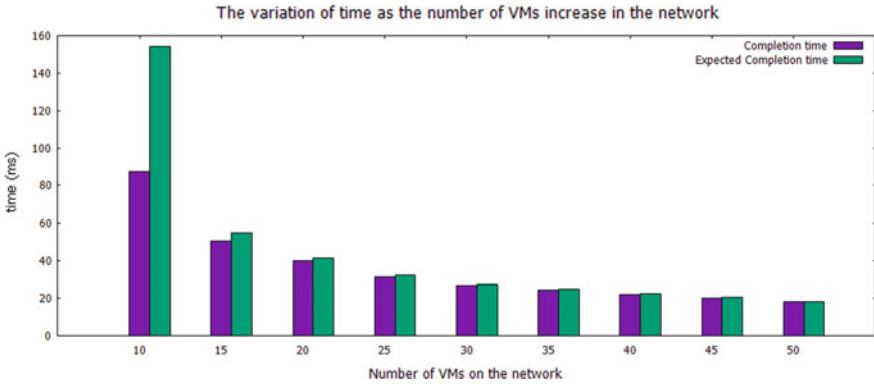


Fig. 4 Comparison between task completion and expected completion time per VM increase

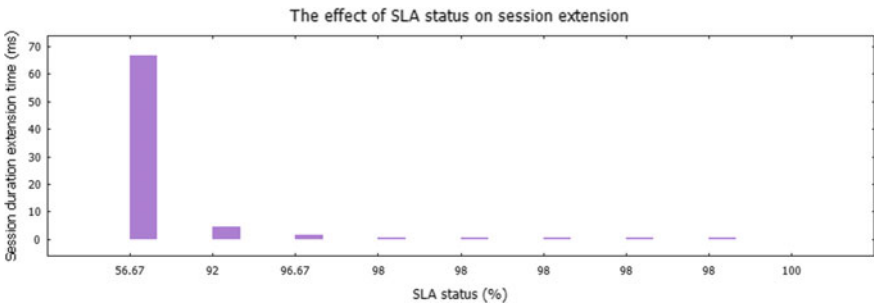


Fig. 5 Decrease in session extension time as SLA percentage increases in the network

on the edge. This can further improve the experience of usage to Cloudlet Consumers making their stay more comforting. On the other hand, the above will increase the purchase patterns of the offered core services in the coffee shop. Despite the increase of VMs on the network, it was observed from 25 to 45 VMs the SLA percentage value remained constant as shown in Fig. 3. The simulation was repeated several times to make reason of the pattern but in all instances the results remained the same. However, the completion time of the tasks (shown in Fig. 4) differed, however, it might be caused by the simulator limitations.

In Fig. 5, a comparison between task completion and expected completion times was illustrated to indicate the effect of increasing VMs. Both task completion and expected completion times decreased with increase in number of VMs on the network, meaning the task processing duration decreased. The decrease in task processing time also improves consumers QoE, in relation to other graphs, it is evident that it does guarantee high throughput or SLA percentage. The decrease in both task completion and expected completion time influences the session extension duration shown in Fig. 5. Since the connection to consume Cloudlet resources is based on a session duration, the function of the session extension duration is to extend time to all the

tasks that were not completely processed by the Cloudlet. Which may have been either caused by network outage or lack of resources in the network. The session extension presented as SET in the computation (2) is the difference of the expected completion time (ECT) and task completion time (CT). Furthermore, ECT [shown in the computation (1)] is the ratio of 150 tasks issued and their task completion time produced by the total network throughput. The session extension is a compensation of SLA violations. Meaning an increase in session extension can result in an increase in a Cloudlet operational cost.

$$ECT = (n * CT)/Cn \quad (1)$$

$$SET = ECT - CT \quad (2)$$

The increase in operational cost may have a bad effect on a coffee shop due to lack of financial support. Therefore, the elimination of session extension lies on the effective management and allocation of resources in the network which will ensure service guarantees and this is achieved through using and monitoring SLA on the network.

5 Related Work

Few publications in respect to the validation of SLA in Cloud scenarios exist in literature and non on Cloudlet-or edge-server-based scenarios. In Cloud scenarios: existing literature focus on establishing mechanisms to detect SLA violations and some include the calculation of SLA violation costs to compensate consumers. The below discussed literature addresses the importance and functionality of SLA in Cloud-based scenarios conducted in a CloudSim tool and they serve as a background to guide the evaluation of SLA for a deployed Cloudlet in a coffee shop.

SLA-based performance framework that extends the Web-SLA was proposed in [10]. The framework transforms low-level to high-level metrics, to measure and analyse performance of Cloud system against the agreed SLA by the parties. The simulation setup considered different application of SLA on different VM policies such as maximize throughput provision policy (MTPP), minimize response time provision policy (MRPP), and maximize utilization policy (MUPP). Based on the comparison, the results proved that MUPP performs better than the other policies due to deadline satisfaction as the tasks increased in a network.

To improve the resource provisioning, SLA violation detection mechanism was proposed to help allocate efficient processors for task processing on the network [11]. One of the assumptions taken into consideration by the authors is that the VMs allocated by the broker agent are based on the tasks initiated. The simulation setup considered two scenarios: Scenario 1: The number of VMs (16 VMs) greater than the number of resources (15) requested by the submitted jobs. In Scenario 2:

The number of VMs (14 VMs) less than the number of resources (27) requested by the submitted jobs. Based on the setup, no SLA violations were encountered on scenario 1 due to efficient resource allocation where else, scenario 2 resulted in high SLA violations. The work lack information with regards to the compensation of the detected SLA violations. Another work similar to the literature above proposed a novel approach implemented on a clients end to monitor SLA compliance [7]. Part of the contribution of this work mentions that the monitored SLA compliance is sent as feedback to the provider. The monitoring of SLA compliance as clients initiates requests for processing on the Cloud follows the steps listed below:

1. A client generates fetch task information using a Gen (SLA) function. The function accepts SLA as an input. The Gen (SLA) function generates a task which consists of instructions to collect relevant information with regards to SLA compliance status on the Cloud.
2. The information fetches task is forwarded along with a set of tasks for processing to the Cloud instance.
3. When the tasks arrive on the Cloud, a log is created. The log documents the arrival time and hash of the tasks. Then the Cloud executes the tasks and the SLA status to the provider.

Work done in [9] lacks proof of concept regarding the proposed mechanism also, actionable policies in a case of a detected SLA violation. The gap of SLA violation compensation which is not covered by the above studies was done in [12]. The scholars developed a Cloud SLA availability framework that compares SLAs of different Cloud providers, the framework calculates both the availability of resources and penalty cost to select a penalty degree that will yield more profit. The SLA penalty in this study is dependent on the percentage of resource availability. To improve profit for Cloud providers, the authors developed a business framework that helps providers to get a better penalty degree for their SLA. This also helps to determine the availability based SLA for cloud services.

6 Conclusion

The evaluation of SLA carried out in this paper considers the use of the mechanisms implemented on the existing literature as a guide to ensure that the Cloudlet meets consumers' service guarantees. Part of the contribution is the extension of user session which is considered as a mechanism by the service provider to compensate consumers for SLA violation. The SLA violations are dependent on the breach of Cloudlet QoS metrics stated on the agreed SLA. In cases where an SLA is violated, the following is done to ensure service guarantees and improve QoE for consumers:

1. The Cloudlet Owner receives a notification with regards to SLA violation and possible issue that might have caused it, in order to mitigate the issue.
2. The Cloudlet consumer session is extended based on the network status.

Therefore, the use of SLA to guide the provision and management of Cloudlet resources will ensure service guarantees and QoE to consumers. This can be achieved through explicitly defining the level of service a Cloudlet must meet in the SLA, SLA monitoring, and efficient resource provisioning in the network. Furthermore, the monitoring of the SLA by Cloudlet Owners such as a coffee shop will help such businesses minimize SLA violation plus operational cost of the Cloudlet.

Acknowledgements The authors would like to extend their acknowledgement to the Council of Scientific and Industrial Research (CSIR) for funding and the Department of Computer Science at the University of Zululand for support.

References

1. K. Bilal, O. Khalid, A. Erbad, S.U. Khan, Potentials, trends, and prospects in edge technologies: fog, cloudlet, mobile edge, and micro data centers. *Comput. Netw.* **130**, 94–120 (2018)
2. Z. Pang, L. Sun, Z. Wang, E. Tian, S. Yang, A survey of cloudlet based mobile computing, in *2015 International Conference on Cloud Computing and Big Data (CCBD)* (IEEE, 2015), pp. 268–275
3. Y. Gao, W. Hu, K. Ha, B. Amos, P. Pillai, M. Satyanarayanan, Are cloudlets necessary? School of Computing Science, Carnegie Mellon University, Pittsburgh, PA, USA, Technical report. CMU-CS-15-139 (2015)
4. S. Durst, I. Edvardsson, Knowledge management in SMEs: a literature review. *J. Knowl. Manag.* **16**(6), 879–903 (2012)
5. M.N. Nxumalo, M.O. Adigun, I. Mba, An envisaged SLA-based cloudlet business model for ensuring service guarantees, in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)* (IEEE, 2018), pp. 1–4
6. S.E. Middleton, M. Surridge, B.I. Nasser, X. Yang, Bipartite electronic SLA as a business framework to support cross-organization load management of real-time online applications, in *European Conference on Parallel Processing* (Springer, 2009), pp. 245–254
7. K.S. Suneel, H.S. Guruprasad, A novel approach for SLA compliance monitoring in cloud computing. *Int. J. Innov. Res. Adv. Eng.* **2**(2), 154–159 (2015)
8. M.C. Silva Filho, R.L. Oliveira, C.C. Monteiro, P.R. Inácio, M.M. Freire, CloudSim plus: a cloud computing simulation framework pursuing software engineering principles for improved modularity, extensibility and correctness, in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)* (IEEE, 2017), pp. 400–406
9. C. Sonmez, A. Ozgovde, C. Ersoy, EdgeCloudSim: an environment for performance evaluation of edge computing systems, in *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)* (IEEE, 2017), pp. 39–44
10. A.B. Khattak, A. Jalal, Performance based service level agreement in cloud computing. *Int. J. Res. Publ.* **4**(4), 20–30 (2015)
11. S.M. Musa, A. Yousif, M.B. Bashi, Sla violation detection mechanism for cloud computing. *Int. J. Comput. Appl.* **133**(6), 8–11 (2016)
12. Q. Xia, W. Liang, W. Xu, Throughput maximization for online request admissions in mobile cloudlets, in *2013 IEEE 38th Conference on Local Computer Networks (LCN)* (IEEE, 2013), pp. 589–596

A New Use of Doppler Spectrum for Action Recognition with the Help of Optical Flow



Meropi Pavlidou and George Zioutas

Abstract In this work, we present two new procedures for activity recognition that are based on the Fourier frequencies that are generated when the optical flow values of successive frames of video are processed simultaneously. In the first algorithm, we correlate these 2D Doppler Fourier spectra with the mean spectra of each activity class. These correlation vectors, which include only 30 features in number, are categorized using a reduced robust SVM classification model. This first procedure is of low computational cost for action recognition tasks for numerable activity classes. For large numbers of activity classes, we propose a new method of aggregated weighted spectra of optical flow values across the whole video. The above-mentioned Fourier spectra are concatenated with a short vector representing the distributions of the moving edges. These methods are insensitive to the presence of background as well as to the positions of the subjects and their shapes and can encode the information of a part or of the whole of a video into relatively short vectors. The results of the two procedures seem to be competitive to state-of-the-art action recognition methods when tested on the KTH Royal Institute Database and on the UCF101 Database for action recognition tasks.

Keywords Action recognition · Doppler frequencies · Spectral density estimation · Kalman Filter · Optical flow · SVM classification · Principal component analysis

1 Introduction

Activity recognition plays an important role for many different applications such as health care, human–computer interaction or social sciences. Action recognition algorithms are based either on global features or on local features. Spatiotemporal feature points based on local movements aim at robustness to pose, image clutter, occlusion

M. Pavlidou (✉) · G. Zioutas
Aristotle University of Thessaloniki, Thessaloniki, Greece
e-mail: mepa@auth.gr

G. Zioutas
e-mail: zioutas@eng.auth.gr

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_35

and object variation in [1]. In [2], dynamic time warping and posterior probability model the gait of the subjects. The work of Efros et al. [3] proposes a spatiotemporal descriptor based on optical flow estimation. Probabilistic latent semantic analysis and latent dirichlet allocation in [4] estimates the probability distributions of the spatial–temporal words. Neural networks for computational intelligence applications [5–7] are applied to the frames directly or to trajectories extracted by multiple frames [8].

Many of the algorithms that are currently used for activity recognition include large vectors of data and are sensitive to the different frame sizes and the resolution of the input videos. In an effort to increase the amount of information coming from multiple frames, we decided to combine the Doppler frequencies that are generated from the optical flows of the moving edges of multiple frames.

In this paper, we take advantage of the Doppler frequencies that are generated when we concatenate the 2D optical flow values of the pixels of the moving edges. The vectors that are produced from our two methods, two-dimensional correlation coefficient Doppler spectroscopy descriptor (CCDS) and weighted Doppler spectroscopy descriptor (WDS), are relatively short, may include information from a part of the movement video or from the whole video and are insensitive to noise, background or brightness.

2 Two-Dimensional Correlation Coefficient Doppler Spectroscopy Descriptor (CCDS)

In the CCDS algorithm, the first step is to identify the moving edges of two consecutive frames with the help of the Kalman filter. In the second step, the Horn–Schunck [9] optical flow values of those edges are computed, as in Fig. 1a. Since those optical flow values regard two consecutive video frames and are present in the same matrix,

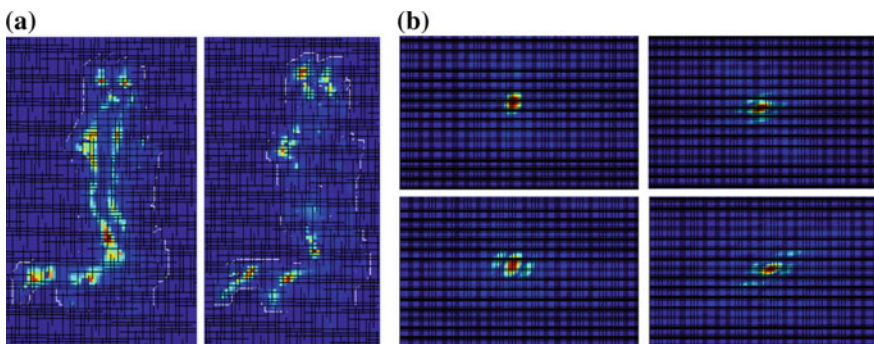


Fig. 1 **a** Horizontal and vertical optical. **b** Two-dimensional spectral density of the flow velocities present from two con-upper-half and the lower half of the jogging man seductive frames of a person jogging frame

Doppler frequencies are generated. Those Doppler frequencies are detected when we estimate the two-dimensional Fourier spectrum of this matrix as shown in Fig. 1b.

The discrete Fourier transform Y of an m -by- n matrix X :

$$Y_{p+1,q+1} = \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} w_m^{jp} w_n^{kq} X_{j+1,k+1} \tag{1}$$

For better classification results, and inspired from [10] where halves of the frames are used in computations, we divide the optical flow values in half into those belonging to the upper half and the lower half, and also we divide them into the vertical and the horizontal optical flow values. Consequently, we now can compute the Doppler spectra of four two-dimensional matrices as we can see in Fig. 2. In the next step, we calculate the values of the final vector prior the classification. These values are the 2D correlations between the four spectra of each frame, as in Fig. 2, and the average spectra of each class of data are, for example, shown in Fig. 3a, b.

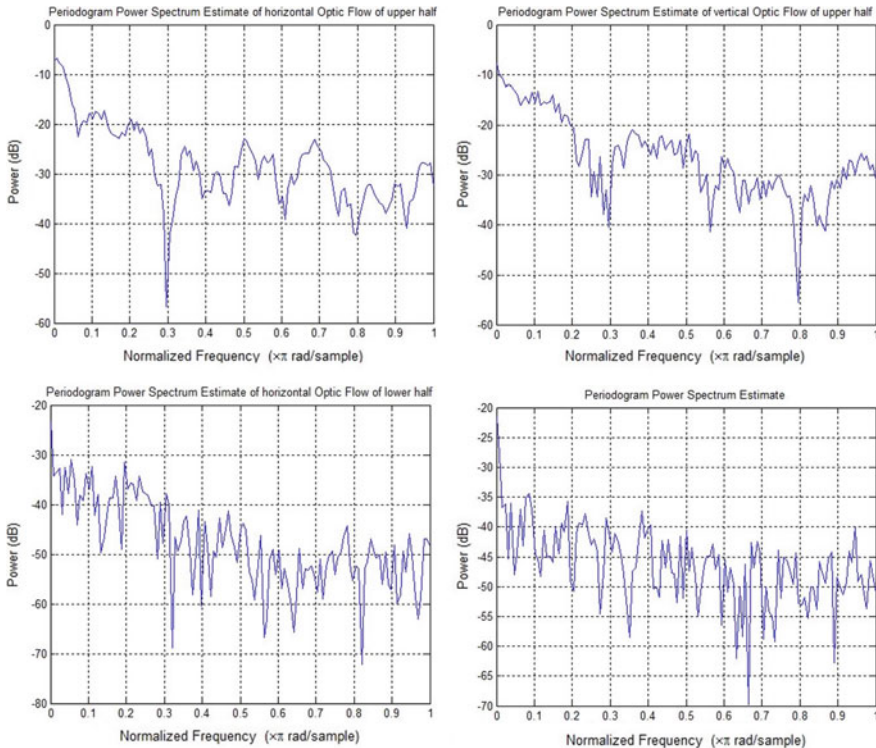


Fig. 2 Periodograms of horizontal and vertical values of optical flows for the upper and lower half of a boxing video

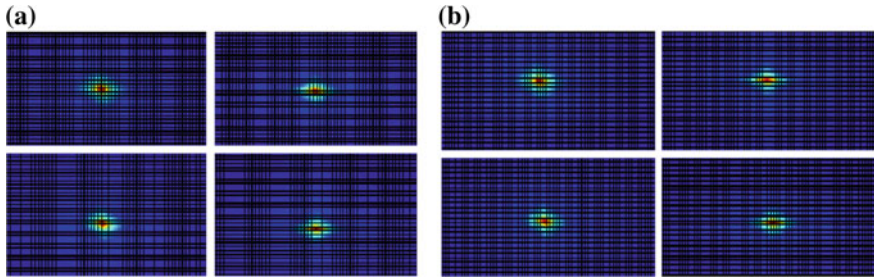


Fig. 3 **a** Averaged spectra of the Doppler frequencies of the horizontal (a), (c) and vertical (b), (d) optical flow values of the upper (a), (b) and lower (c), (d) halves of class Boxing of the KTH Activity Database [11]. **b** Averaged spectra of the Doppler frequencies of the horizontal (a), (c) and vertical (b), (d) optical flow values of the upper (a), (b) and lower (c), (d) halves of class walking of the KTH activity database [11]

Two-dimensional correlation analysis is a mathematical technique that is used to study changes in measured signals. Let us assume that we have in our disposition two spectral densities of two two-dimensional optical flow velocities, F^k and F^M . F^k is the spectral density of the optical values of frame k and F^M The two-dimensional correlation is the mean spectral density of the optical values of all the frames of class M , for example, of the activity of jogging. Then, the two-dimensional correlation, Corr_{2D} , between F^k and F^M is calculated as:

$$\text{Corr}_{2D} = \frac{\sum_m \sum_n \left(F_{mn}^k - \overline{F_{mn}^k} \right) \left(F_{mn}^M - \overline{F_{mn}^M} \right)}{\sum_m \sum_n \left(F_{mn}^k - \overline{F_{mn}^k} \right) \sum_m \sum_n \left(F_{mn}^M - \overline{F_{mn}^M} \right)} \quad (2)$$

Corr_{2D} is used as a metric of similarity of the spectral density of each frame with the mean spectral density of the optical flow velocities of all the frames in a specific class. If we correlate the spectral densities for all the frames and all the classes, for the horizontal and vertical velocities, we end up with a vector of correlation coefficients of size $2N$ for each frame. N is the number of the classes, and therefore, the number of the means of the spectral densities is 2 because we have the densities for the horizontal and vertical velocities separately. Concluding the CCDS method so far, the proposed algorithm for the estimation of the correlated spectral density activity descriptors follows the subsequent procedure:

- Use Kalman filter to detect motion in each video and choose only those frames that are positive for movement detection and the pixels in each frame that are positive for movement.
- Convert each frame from RGB to greyscale.
- Compute the optical flow values using Horn–Schunck algorithm for the pixels of interest.
- Estimate the absolute values of the optical flow as we are interested in the velocity of the edges and not the direction of the movement.

- Combine the optical flow values of the moving edges of two consecutive frames by adding them, creating a 2D matrix that holds the information of pixels moving relatively to each other with specific velocities and thus generating Doppler frequencies.
- Estimate the spectral densities of the vertical and the horizontal concatenated optical flow values in the upper half and the lower half of the image.
- Correlate those spectrum densities with the mean spectrum densities of each activity class.

At this point, we have in our disposition a vector for each frame of every video. This vector holds the concatenated spectral densities of the optical flow values, horizontal and vertical, of the upper half and lower half, which were detected by the Kalman filter. The next step is to apply the classification method for each vector so that we can recognize human activity.

3 Weighted Doppler Spectroscopy Descriptor (WDS)

We already described how, as objects move across the frames, the pixels that show nonzero velocities create moving frames. In Fig. 5, the horizontal and vertical velocities of two consecutive frames of a person jogging are present. However, in our experiments we noticed that when the number of classes is increased and along with them the number of frames and the range of the activities, then CCDS which draws information based on only two consecutive frames is easier to misclassify. The CCDS information of the two frames is the partial information of a movement that may regard seconds or even a fraction of a second and may not be able to represent the nature of the action in a video.

At this point, the need for an algorithm that efficiently combines the knowledge of all the frames taking into consideration the direction of the evolution of the action seems as the next step that we need to take. So how about we gather the information not from just two frames, because it is too scarce, but from the whole video, all of its frames? How about having the vertical and the horizontal velocities of all the frames present in two matrices? In this way, we can produce a single matrix that holds the Doppler spectrum information of the optical flow values of all the frames of the video and represents the whole action and not just a piece of the action like the information provided by CCDS that regards only two frames. Below we present an example of the spectra of all the video frames of a surfing video that belongs to the UCF101 Dataset [12]. The respective Doppler spectra of the upper and the lower halves of the horizontal and vertical velocities would be those of Fig. 4b.

In Fig. 4b, we can see the spectra of the velocities, horizontal and vertical, of moving pixels of all frames of class Surfing of the UCF101 Dataset, of the upper halves and lower halves. The velocities of all the frames are summed up. This unweighted summation provides no information on the evolution of the action depicted in the video. Therefore, a weight is multiplied with the velocities of the moving edges of

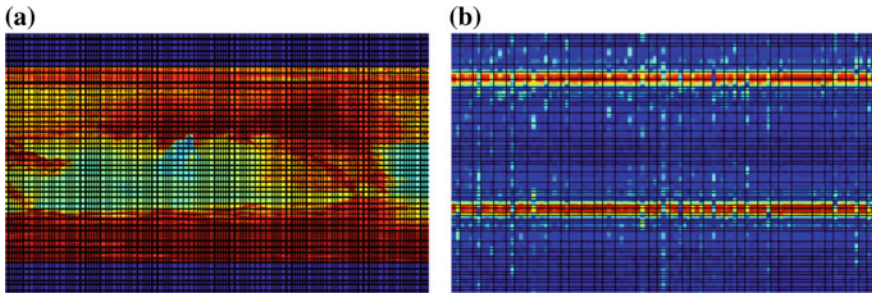


Fig. 4 **a** Moving edges of all frames of class Surfing of the UCF101 Dataset combined in one matrix. **b** Spectra of the velocities, horizontal and vertical, of moving pixels of all frames of class Surfing of the UCF101 Dataset, of the upper halves and lower halves. These are the spectra of the summed up velocities of every frame

every frame. For the first frame, the weight is the number of the frames of the video. For the second frame, this number is reduced by one and so on until the weight for the last video becomes equal to one. Finally, these weighted velocities are summed up and normalized, and their spectra are calculated for the upper half of the frames, the lower half, the horizontal and the vertical velocities. The matrix of the Spectra of every video is reshaped into a vector. We can see distinct shapes that differ from one another in terms of shape, inclination, dispersion and intensity. Moreover, for the sake of computational complexity, we have reduced the resolution of the spectra during the validation step down to a level that produces reliable results without the need of extremely large matrices for the spectra. Moreover, in order to clean each class of the Training subset, we apply the robust principal component analysis (Robust PCA) method [13–16] to clean the data of each class of the Training matrix, thus producing the final clean vector for each video without outliers. Perform one-against-all (OAO) SVM classification, and assign each data point to the class for which it gathers the most scores, above 50%. This vector will be concatenated with the reshaped spectra vector, and this final vector will comprise our new action recognition descriptor.

Concluding the WDS method, the proposed algorithm for the estimation of the weighted Doppler spectral density activity descriptors follows the subsequent procedure:

- Use Kalman filter to detect motion in each video, and choose only those frames that are positive for movement detection and the pixels in each frame that are positive for movement.
- Convert each frame from RGB to greyscale.
- Compute the optical flow values using Horn–Schunck algorithm for the pixels of interest.
- Estimate the absolute values of the optical flow as we are interested in the velocity of the edges and not the direction of the movement.
- Combine the optical flow values of the moving edges of all consecutive frames by adding them, with decreasing weights that decline from the number of frames to

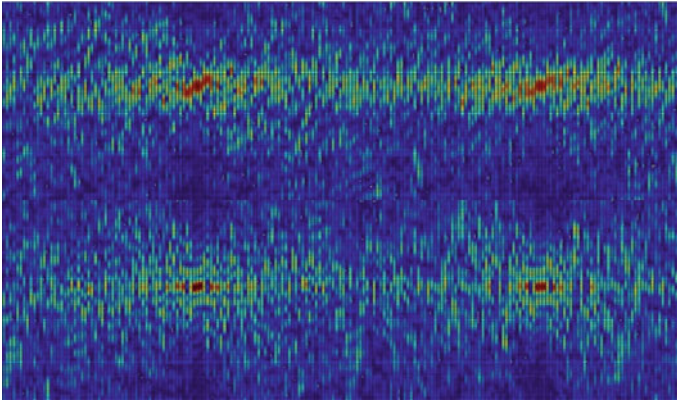


Fig. 5 Spectra of the velocities, horizontal and vertical, of moving pixels of all frames of class Surfing of the UCF101 [12] dataset, of the upper halves and lower halves. These are the spectra of the weighted summed up velocities of every frame

- 1, creating a 2D matrix that holds the information of edges moving relatively to each other with specific velocities and thus generating Doppler frequencies.
- Estimate the spectral densities of the vertical and the horizontal concatenated optical flow values in the upper half and the lower half of the image. Concatenate the matrices and reshape them into one vector for each video.
- Apply the robust PCA method to clean the data of each class of the Training matrix, thus producing the final clean vector for each video without outliers.
- Perform OAO linear SVM classification, and assign each data point to the class for which it gathers the most scores, above 50% .

4 Experimental Results

4.1 *KTH Royal Institute of Technology Human Activity Database*

We tested our model on numerous human activity datasets. The first results come from KTH Royal Institute of Technology [11]. This video database containing six types of human actions, walking, jogging, running, boxing, hand waving and hand clapping, performed several times by 25 subjects in four different scenarios: outdoors, outdoors with scale variation, outdoors with different clothes and indoors as illustrated below. Currently, the database contains 2391 sequences. All sequences were taken over homogeneous backgrounds with a static camera with 25 fps frame rate. The sequences were downsampled to the spatial resolution of 160×120 pixels and have a length of four seconds in average.

Table 1 Classical SVM classification on periodogram of quantized optical flow values of KTH Royal Institute Database

SVM	Yeo [17]	Sch. [11]	Dol. [1]	Nieb. [4]	Hist. [10]	Per.
handcl.	0.89	0.60	0.76	0.93	0.78	0.9015
handw.	0.95	0.74	0.88	0.77	0.75	0.8673
box.	0.82	0.97	0.93	1.00	0.86	0.9059
jog.	0.58	0.60	0.13	0.52	0.92	0.8976
run.	0.91	0.55	0.84	0.88	0.92	0.9344
walk.	1.00	0.84	0.89	0.79	0.89	0.9201

We notice that in the cases of handclapping, hand waving and boxing, Nieb. [4] outperforms the periodogram descriptors by 2%. However, this method is outperformed in the cases of jogging, running and walking and specifically has a recognition score of 13% in the case of jogging. Sch. [11] method is outperformed by 2–38% in every case except boxing where the recognition rate is 7% higher. Yeo [17] is also outperformed in handclapping, boxing, jogging and running by 2–31% while it outperforms the periodogram descriptors by 8% in walking and running. Finally, Hist. [10] method, which does not display recognition rates lower than 75% and shows high recognition efficiency, is outperformed from 1.44 to 12% in every case except the boxing case where the result is better by 2.3% compared to the periodogram algorithm. Concluding, we can say that the new method outperforms the up-to-date action recognition methods and is characterized by steady behaviour across the six different actions of handclapping, hand waving, boxing, jogging, running and walking (Table 1).

4.2 University of Central Florida, UCF101, Human Activity Database

UCF101 [12] is an action recognition dataset of realistic action videos, collected from YouTube, having 101 action categories. UCF101 includes 13,320 videos from 101 action categories, providing with a large diversity in terms of actions and with the presence of large variations in camera motion, object appearance and pose, object scale, viewpoint, cluttered background, as well as illumination conditions and is considered to be one of the most challenging dataset to date. The action categories in UCF101 can be divided into five types: human–object interaction, body-motion only, human–human interaction, playing musical instruments and sports. Specifically, the action categories for UCF101 dataset are: Apply Eye Makeup, Apply Lipstick, Archery, Baby Crawling, Balance Beam, Band Marching, Baseball Pitch, Basketball Dunk, Basketball, Bench Press, Biking, Billiards, Blow Dry Hair, Blowing Candles, Body Weight Squats, Bowling, Boxing Punching Bag, Boxing Speed Bag, Breaststroke, Brushing Teeth, Clean and Jerk, Cliff Diving, Cricket Bowling,

Cricket Shot, Cutting in Kitchen, Diving, Drumming, Fencing, Field Hockey Penalty, Floor Gymnastics, Frisbee Catch, Front Crawl, Golf Swing, Haircut, Hammer Throw, Hammering, Handstand Pushups, Handstand Walking, Head Massage, High Jump, Horse Race, Horse Riding, Hula Hoop, Ice Dancing, Javelin Throw, Juggling Balls, Jump Rope, Jumping Jack, Kayaking, Knitting, Long Jump, Lunges, Military Parade, Mixing Batter, Mopping Floor, Nun chucks, Parallel Bars, Pizza Tossing, Playing Cello, Playing Daf, Playing Dhol, Playing Flute, Playing Guitar, Playing Piano, Playing Sitar, Playing Tabla, Playing Violin, Pole Vault, Pommel Horse, Pull Ups, Punch, Push Ups, Rafting, Rock Climbing Indoor, Rope Climbing, Rowing, Salsa Spins, Shaving Beard, Shot-put, Skate Boarding, Skiing, Ski jet, Sky Diving, Soccer Juggling, Soccer Penalty, Still Rings, Sumo Wrestling, Surfing, Swing, Table Tennis Shot, Tai Chi, Tennis Swing, Throw Discus, Trampoline Jumping, Typing, Uneven Bars, Volleyball Spiking, Walking with a dog, Wall Pushups, Writing On Board and Yo Yo. The downloaded data for the UCF101 dataset did not include any videos for the classes. For the purpose of classifying UCF101 activity videos, up-to-date classification methods have been applied. In [18], a ConvNet model which forms a temporal recognition stream is formed. This model is formed by stacking optical flow displacement fields between several consecutive frames which facilitate the description of the motion between video frames and make the recognition easier, as the network does not need to estimate motion implicitly. The Temporal ConvNets achieved a remarkable 81.2% recognition rate on the UCF101 database. [19] applied a flexible and efficient Temporal Segment ConvNet and succeeded a 94.2% recognition rate on the UCF101. We also inserted the WDS vectors in Sect. 5 Accelerated SVM method performs one-against-one (OAO) classification for each of the 101 classes of the UCF101. Each data sample is assigned to the class that gets the higher score among the OAO classification tasks. If the sample is correctly assigned, then it is regarded as being correctly recognized. The recognition rates of the WDS algorithm are presented in Table 2. The training and testing vectors can be provided through an e-mail to mepa@auth.com.

5 Conclusions

In this research, two new action recognition methods were developed for action recognition that are based on a new use for the Doppler frequencies. We combined the optical flows of the moving pixels of two subsequent frames for the CCDS method and for all the frames with the help of weights for the WDS method. The presence of Optical flow values in 2D matrices that express the evolution of the movement and generate Doppler frequencies. These frequencies are subsequently used for Activity Recognition with the help of Support Vector Machines. In the case of the WDS, a robust accelerated SVM method was used since the number of videos for the first experiment was limited and the number of the training and testing vectors is also limited. When we proceeded to the second experiment with the UCF101 dataset, the number of videos and therefore frames increased significantly. This fact

Table 2 Recognition rates using the WDS method on the UCF101 Database

Apply Eye Makeup	92.68	Biking	86.4864	Clean and Jerk	(Missing data)
Apply Lipstick	78.1250	Billiards	100	Cliff Diving	70.1244
Archery	90	Blow Dry Hair	88.8888	Cricket Bowling	97.4358
Baby Crawling	91.8919	Blowing Candles	(Missing data)	Cricket Shot	91.4893
Balance Beam	86.6667	Body Weight Squats	74.19354	Cutting in Kitchen	76.6666
Band Marching	100	Bowling	100	Diving	88.8888
Baseball Pitch	100	Boxing Punching Bag	73.20354	Drumming	93.3333
Basketball Dunk	78.378	Boxing Speed Bag	100	Fencing	35.4838
Basketball	100	Breast Stroke	85.7142	Field Hockey Penalty	80
Bench Press	46.666	Brushing Teeth	94.444	Floor Gymnastics	94.2857
Frisbee Catch	97.1428	Horse Race	91.1764	Long Jump	100
Front Crawl	100	Horse Riding	82.6086	Lunges	71.4285
Golf Swing	65	Hula Hoop	62.8571	Military Parade	82.857
Haircut	72.2222	Ice Dancing	97.7272	Mixing Batter	92.1052
Hammer Throw	82.0512	Javelin Throw	84.3750	Mopping Floor	96.6666
Hammering	90.4761	Juggling Balls	70.5882	Nun chucks	86.4864
Handstand Pushups	83.3333	Jump Rope	76.4705	Parallel Bars	96.8750
Handstand Walking	93.5483	Jumping Jack	82.5000	Pizza Tossing	87.0967
Head Massage	80.4878	Kayaking	97.4358	Playing Cello	93.4782
High Jump	91.1764	Knitting	88.2352	Playing Daf	90.4761
Playing Dhol	65.2173	Punch	98.9010	Skiing	89.4736
Playing Flute	67.4418	Push Ups	50	Ski jet	92.8571
Playing Guitar	84.4444	Rafting	80.6451	Sky Diving	80

(continued)

Table 2 (continued)

Playing Piano	89.6551	Rock Climbing Indoor	77.5000	Soccer Juggling	78.0487
Playing Sitar	88.6363	Rope Climbing	81.8181	Soccer Penalty	94.7368
Playing Tabla	80.6451	Rowing	81.5789	Still Rings	93.5483
Playing Violin	89.2857	Salsa Spin	91.8918	Sumo Wrestling	84.3750
Pole Vault	83.3333	Shaving Beard	88.8888	Surfing	91.4285
Pommel Horse	85.2941	Shot-put	95	Swing	89.4736
Pull Ups	85.71428	Skate Boarding	90.9090	Table Tennis Shot	92.3076
Tai Chi	64.2857	Tennis swing	95.6521	Throw Discus	88.8888
Trampoline Jumping	87.8787	Typing	100	Uneven Bars	86.2068
Volleyball Spiking	71.8750	Walking with dog	85.2941	Wall Pushups	77.7777
Writing on Board	85.7142	Yo Yo	72.9729		

along with the significant increase in the number of the classes, from 6 classes of the first experiment to 101 classes, created the need for a more robust method that would save computational time and cost. This is how we multiplied each optical flow frame values with weights, and we added them in order to get one single vector for each video. The vector is the reshaping of a 2D spectrum of size 102×102 which results in a vector of 10,404 elements. If we want even better resolution, we will probably increase the resolution of the 2D Fourier transform which will generate larger vectors that will include more frequencies. The choice of our parameters for the experiments was performed by using one-third of the videos from every one of the 101 classes for validation purposes. The data after the final processing and checks are available by sending an e-mail to the authors.

Acknowledgements The KTH Royal Institute of Technology video database [11] that was used is publicly available for non-commercial use.

The UCF101, Human Activity Database [12] is freely available here: <https://www.crcv.ucf.edu/data/UCF101.php>.

References

1. P. Dollar, V. Rabaud, G. Cottrel, S. Belongie, Behavior “recognition via spatiotemporal features”, in *Proceedings 2nd Joint IEEE International Workshop on VS-PETS, Beijing* (IEEE Computer Society Press, Los Alamitos, 2005), pp. 65–72
2. L. Wang et al., Fusion of static and dynamic body biometrics for gait recognition. *IEEE Trans. Circ. Syst. Video Technol.* **14**(2), 149–158 (2004)
3. A.A. Efros, et al., Recognizing action at a distance, in *ICCV*, vol. 3 (2003)
4. J.C. Niebles, H. Wang, L. Fei-fei: Unsupervised learning of human action categories using spatial-temporal words, in *BMVC* (2006)
5. L.P. Wang, X.J. Fu, *Data Mining with Computational Intelligence* (Springer, Berlin, 2005)
6. X.J. Fu, L.P. Wang, Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. *IEEE Trans. Syst. Man Cybern Part B Cybern* **33**(3), 399–409 (2003)
7. L.P. Wang, On competitive learning. *IEEE Trans. Neural Networks* **8**(5), 1214–1217 (1997)
8. V. Luong, L.P. Wang, G. Xiao, Deep networks with trajectory for action recognition in videos, in *The 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES 2014)*, Singapore, 10–12th Nov 2014
9. B.K.P. Horn, B.G. Schunck, Determining optical flow. *Artif. Intell.* **17**(1–3), 185–203 (1981)
10. S. Danafar, N. Gheissari, Action recognition for surveillance applications using optic flow and SVM, in *Asian Conference on Computer Vision* (Springer, Berlin, 2007)
11. C. Schuldt, I. Laptev, B. Caputo, Recognizing human actions: a local SVM approach, in *ICPR* (2004), pp. 32–36
12. K. Soomro, A.R. Zamir and M. Shah, UCF101: a dataset of 101 human action classes from videos in the wild. *CRCV-TR-12-01*, November, 2012.
13. M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal components analysis. *Technometrics* **47**, 64–79 (2005)
14. M. Hubert, S. Engelen, Fast cross-validation of high-breakdown resampling algorithms for PCA. *Comput. Stat. Data Anal.* **51**, 5013–5024 (2007)
15. M. Hubert, P.J. Rousseeuw, T. Verdonck, Robust PCA for skewed data and its outlier map. *Comput. Stat. Data Anal.* **53**, 2264–2274 (2009)
16. S. Serneels, T. Verdonck, Principal component analysis for data containing outliers and missing elements. *Comput. Stat. Data Anal.* **52**, 1712–1727 (2008)
17. P. Ahammad, C. Yeo, S.S. Sastry, K. Ramchandran, Compressed domain real-time action recognition, MMSP, in *Proceedings of 8th IEEE Workshop on Multimedia Signal Processing* (IEEE Computer Society Press, Los Alamitos, 2006) pp. 33–36
18. K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos, in *Advances in Neural Information Processing Systems* (2014)
19. L. Wang, et al., Temporal segment networks: towards good practices for deep action recognition, in *European Conference on Computer Vision* (Springer International Publishing, 2016)

Meeting Challenges in IoT: Sensing, Energy Efficiency, and the Implementation



Toni Perković, Slaven Damjanović, Petar Šolić, Luigi Patrono
and Joel J. P. C. Rodrigues

Abstract IoT became the design language for the future smart infrastructures. Therein, IoT devices are sensing the changes and delivering the data through some efficient radio, while the whole system is meant to be power efficient. In this paper, we gather state-of-the-art technology for building IoT device and analyze the power consumption of the whole system, therefore, providing the useful estimate on possibilities to use it in the future infrastructures. Further on, specific use case is given that justifies the adoption of the device in the real environment.

Keywords Internet of Things · Long range · Low power · Smart environment

1 Introduction

Internet of things (IoT) became the hottest topic in modern technological applications, where the research directions enable the new possible applications that make life easier. The fundamental idea in IoT is to put all data of all things on Web, where

T. Perković (✉) · S. Damjanović · P. Šolić
University of Split, Split, Croatia
e-mail: toperkov@unist.hr

S. Damjanović
e-mail: Slaven.Damjanovic.00@fesb.hr

P. Šolić
e-mail: psolic@fesb.hr

L. Patrono
University of Salento, Lecce, Italy
e-mail: luigi.patrono@unisalento.it

J. J. P. C. Rodrigues
National Institute of Telecommunications (Inatel), Santa Rita do Sapucaí, Brazil
e-mail: joeljr@ieee.org

Instituto de Telecomunicações, Lisboa, Portugal

University of Fortaleza (UNIFOR), Fortaleza, CE, Brazil

today's technologies are making it feasible. These are advancing in every aspect and becoming less robust, less expensive, and less power hungry. IoT is enabling huge investments which will further increase as shown by prognosis of reputable vendors such as CISCO and ERICSSON [1, 2]. As relatively young research area with devices with still limited capabilities, there is a vast space for further advances in device performances and capabilities in all layers of its ISO/OSI architecture. This mainly relates to further extension of wireless interrogation range and smarter control of power constraints, while also providing new applications. Once these requirements are improved, new applications in different user domains will emerge and further increase IoT market size.

The application space of IoT devices is vast. For example, smart cities can be characterized with several application points of view: smart grid with integrating home appliances, smart lightning systems, smart cars, tags, health, energy, parking, and homes. These applications can be achieved by using specified software which gathers data from technologies that need to ensure sensing capabilities (accelerometer, magnetometer, pressure, temperature, humidity, etc), which data gets processed by some low-power embedded computing architecture, while at the same time providing connectivity [3].

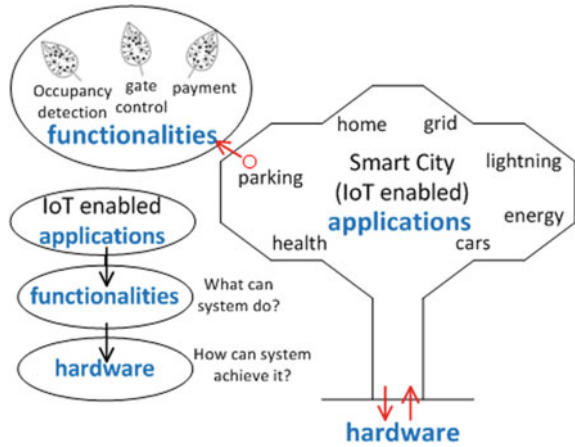
To ensure the feasibility of the proposed approach, the enabling devices are necessary to be less robust and power efficient, which due to the application scenario might hard to be meet. In this paper, we give an overview of sensing technologies that enable IoT applications and challenges to be met during the sensing procedure. To meet these needs it is necessary to investigate which parameters might be crucial in deploying the system, and which constraints should be taken care when considering such architectures. Further on, the example of implementation on agricultural IoT device is provided.

2 Challenges for IoT Devices

IoT could be implemented in many different environments. However, to better understand the scope of its applications, we provide the overall system architecture depicted in Fig. 1. Therein, it is necessary to firstly define the applications (given in the smart city context) that can be achieved with less or more functionality depending on how the hardware is used. However, to adopt IoT in real environment, we discuss several key IoT challenges mainly related to the way on how to gather data [4].

Battery Lifetime and Power Consumption [5]: The consumption of device largely depends if the device is continuous in sensing, processing, and transmitting/receiving potential data. More samples from the sensor mean better data interpretation, but on the other side increases power consumption. It is necessary to determine: (a) how long the device will be sensing prior deciding to further actions and (b) the strategy on how to manipulate and communicate with respect to the received data.

Fig. 1 Generalized smart city architecture; smart parking application example



Bandwidth/Data Rate/Throughput/Latency: Practical applications implicate use cases that need some level of required data rate. Related limitations are tightly bind to the network that support IoT architecture. It is well known that improving data rates generally means increase in power consumption or the increase in the frequency bandwidth. Generally, both are limited, as power constrains are battery issue, while frequency band is limited with the technology. For each purposes, it is necessary to consider which applications need which data link and determine which technology would better fit to the given needs, especially taking care of possible latency. To build up the IoT network, one must rethink about employing limited resources to serve different kinds of data traffic flows where different environments can appear. In the meantime, it is necessary to assure that the related algorithms that are involved in the whole procedure and protocol stack are simple enough.

Range: Each application in IoT deployment have specific needs related to the range where it is expected to deliver sensed data. This again largely depends on geographic constraints for given application. To convey data over larger distances, it is necessary to insert more the power into radio communication, leading to larger battery consumption. Therefore, application scenario needs to be explicitly determined following green approach to reduce power consumption.

3 Related Low-Power Networks Solutions

To cope with the power consumption and data transmission issues, it is necessary to understand which technologies can be applied to deliver data in timely manner. The communication technologies for deploying low-power wireless solutions for IoT can be classified in the following categories.

3.1 *Low Range*

Low-power local networks are generally used for short-range solutions (less than 1000m), but can be considered for transmissions over larger areas when organized in, for example, mesh topology. Popular examples are listed below.

Radio Frequency Identification (RFID): To communicate with low-power devices, (i.e., tags) over distances of up to 100 meters RFID can be exploited. Although primarily designed for pure identification, recent RFID devices can be used for sensing and sending sensed data through same protocols [6].

Bluetooth Low Energy (BLE): Bluetooth technology is used to enable connection with devices that use low data rate (at max. 1 Mbps) and consume information within a short distance range (in theory, up to 100 m). After improvements, Bluetooth 4.0 was introduced (i.e., BLE) with simpler pairing functions, higher data rate (24 Mbps max), and low-power consumption, aimed at connecting IoT devices [7].

Zigbee: Is an alternative technology to Bluetooth reaching similar distances (up to 100 m). Last version (3.0) is mainly designed for industrial environments and can achieve data rate of 250 kbps. Its design is robust, low power, and built for purposes of occasional data transmission.

3.2 *Large Range—LPWAN*

LPWAN (low power wide area networks) can establish communication between devices over ranges exceeding 1000 m. This technology applies to the low-power radio communication networks, where a single base station can supply thousands of end devices. Below, we give examples of LPWAN radio technologies.

DASH7 given as multi-layered architecture aims to provide communication over a range about up to 2 km while working in the 433 MHz, 868 MHz, and 915 MHz [8] bands. Low latency, highly secure (128-bit AES encryption), with the mobility support, and data rate up to 167 Kbps gives an interesting advantage for IoT applications.

Sigfox is a cellular system, where areas are covered with Sigfox operator base stations. End devices are connected to base stations using the BPSK (binary phase-shift keying) modulation [9]. The system works in band of 868 MHz, using the available spectrum of 400 channels with 100 Hz bandwidth. Achieved range can be 30–50 km in rural areas and about 3–10 km in urban areas. Every base station can cover around one million of end devices, while every device is limited to about 140 messages sent per day using 100 bps data rate.

LoRaWAN the network architecture is considered as a “star of stars” topology, while the LoRa enables the long-range radio communication link. The protocol determines the network capacity, QoS, and security. It uses chirp spread spectrum modulation that can reach data rates 290 bps–50 kbps; while at the same time is

considered very high power efficient. The range that can be achieved, depending on the power of radio, reach from 2–5 km in urban areas up to 45 km in rural areas.

NB-IoT (Narrowband-IoT) is mainly focused on indoor applications. It is based on LTE-M to provide connectivity for a wide range of low-power sensor devices within IoT networks [10].

4 Overview of Low-Cost Sensor Technologies

In this section, we give a brief overview of existing and low-cost sensor technologies that can be used for the realization of IoT ecosystem. As the focus of this work is on the design and implementation of the LPWAN systems based on the fast and efficient prototypes, we give an overview of Arduino-based sensors. Moreover, since LPWANs are battery-powered devices [11], a special focus is placed on sensors and their consumption, aimed at minimizing the overall power consumption of the system.

4.1 *Strategies for Reducing the Consumption of Battery-Powered LPWAN Devices*

As we have stated, our realization is done on simple LPWAN prototypes based on Arduino microcontrollers. The Arduino platform is already being used for data collection systems that include wireless sensor devices for collecting data on temperature and humidity in solid structures [12], monitoring user activity with WiFi network integration [13], monitoring boats on moorings [14], etc. On the other hand, LPWAN devices in most cases do not perform complex operations, and they are basically reduced to periodical readings of sensor status/values, and sending these values to the gateway devices. This makes Arduino an ideal candidate for choosing our prototype for building a LPWAN device. Arduino is based on the Atmel ATmega328P microprocessor and large number of Arduino platforms exists that enable simple and fast creation of sensor prototypes. As the goal is to reduce the overall consumption of the LPWAN device, the Arduino Pro Mini is the ideal candidate for creating such a device. Arduino Pro Mini comes in two variants, with 5 V running at 16 MHz clock speed, and 3.3 V running at 8 MHz clock speed. As the majority of the Arduino compatible sensors and radio modules are operating on 3.3 V (such as LPWAN technology LoRa), we will work with the Arduino Pro Mini microcontroller on 3.3 V. Moreover, Arduino running on at 8 MHz clock speed microcontrollers additionally reduce the overall power consumption. In order to further reduce the consumption of the Arduino Pro Mini, it is recommended to make hardware modifications by removing the LEDs and a voltage regulator that consumes a lot of energy from a battery-powered device [15]. Furthermore, as the LPWAN device is not active most

Table 1 Characterization of sensor components for LPWAN Arduino-based systems

Sensor type	Description	Voltage (V)	Cons. active	Cons. power-down
DHT11	Temp. humidity	3.3, 5.6	1.5 mA	50 μ A
DHT22	Temp. humidity	3.5, 5.5	2.5 mA	150 μ A
BME280	Temp. humidity pressure	1.71–3.6	3.6 μ A	0.1 μ A
BME680	Temp. humidity pressure gas	1.71–3.6	0.09–12 mA	0.15 μ A
BH1750	Light	2.4–2.6	120 μ A	0.01 μ A
UV SH1145	Proximity	1.71–3.6	4.3 mA	–
TSL2561	Infrared	2.7–3.6	0.24 mA	3.2 μ A
ADXL345	Accelerom.	3.3	40 μ A	0.1 μ A
MAX9812	Microphone	2.7–3.6	230 μ A	100 nA
GY-MAX4466	Microphone	2.4–5.5	24 μ A	5 nA
DS18B20	Thermometer	3.0–5.0	1 mA	750 nA
LMx93-N	Flame detect.	2.0–3.6	0.4 mA	–
VL53L0X	Laser range	2.6–3.5	5 μ A	16 μ A
MAG3110	Magnetometer	1.95–3.6	8.6–900 μ A	2 μ A
HMC5883L	Compass	2.16–3.6	100 μ A	2 μ A

of the time, we can reduce the Arduino consumption using software modifications that include setting the device into “power-down” mode, which reduces Arduino consumption in the inactive period to less than 1 μ A. Using some exiting Arduino libraries (such as low-power library from rocketscream) consumption can be reduced by disabling modules such as analog-to-digital converter (ADC), two-wire interface (TWI), serial peripheral interface (SPI), watch-dog timer (WDT), and brown-out detection (BOD) [16]. These modules can be normally used once the device wakes up from power-down (inactive) mode.

In Table 1, we outline sensor components that have low-battery consumption and may provide the foundation for creating fast and reliable LPWAN prototypes. The majority of these sensor devices are already low-power devices in power-down mode, but to further reduce energy consumption, it is either suggested to completely shut-down the access to the power for these devices. This can be implemented either by connecting the sensors’ power pin to Arduino digital pins (operating on 3.3 V), in a way that Arduino simply cuts-off access to the power via this pins once it enters power-down mode. Otherwise, we could use an external transistors or low-power components such as TPL5110 Power Timer to cut-off power to the complete LPWAN device, while keeping consumption below 100 nA [17]. Furthermore, TPL5110 uses a built-in timer that can be regulated with resistors to vary from once-every 100 ms up to once every two hours which is extremely helpful if devices have large duty cycle and require periodical sensing and transmissions. However, if sensors require

a specific time period inside which they need to transmit information over the radio, waking up Arduino from deep sleep can be regulated by external low-power RTC clock. DS3231 [18] is low-power/low-cost RTC clock that can be used to wake-up Arduino from deep sleep using some predefined duty cycle. With a clock precision of ± 2 ppm (5 s drift a month), DS3231 makes a good choice for the realization of low-cost LPWAN systems. Next, we give a case study example of a LPWAN prototype network created using Arduino-based sensors.

5 eAgrar: Case Study of Arduino-Based LPWAN Network for Agricultural Monitoring

Our goal is to create cost-effective agricultural devices that cover a larger geographical area, such that users could have a better knowledge about the state of the whole crop across a large field, and not just a small portion within the field that could potentially have its own microclimate which differs from the rest of the field. It is known that moisture remains in the bays where the temperature is lower for several degrees, which favors to the development of diseases. Keeping the device at low cost can help field owners to install a larger number of devices over a larger area and get better and more detailed information about the crops. LPWAN sensor networks present an ideal candidate for the development of device that periodically informs users about the status of the field (temperature, humidity, moisture, air pressure, etc).

The main feature of the developed LPWAN network is large radio communication range (up to 10km), flexibility of use, and battery-powered end devices. LPWAN systems are characteristic for delay-tolerant networks that require periodical and non-frequent message transmissions. This way, LPWAN devices spend most of their time in power-down mode, which reduces energy consumption and allows battery-operated devices to be available for a larger duration of time.

As we already noted, the initial premise is to create a battery-powered end sensor device with a capability to autonomously operate at least one year without replacing the existing battery. The assumption is closely related to the fact that users want to plant a potentially large number of devices over a larger agricultural area that will periodically send data to the centralized system without any significant hardware interventions on the operation of end devices (e.g., changing battery) during the time period of at least one year. Figure 2 shows the architecture of the developed system.

5.1 Implementation

Our wireless sensor network consists of sensor units, i.e., end devices, scattered over a large geographical area, and a central unit, a gateway that collects data from sensor units acting as a concentrator. The data collected from end devices is further sent

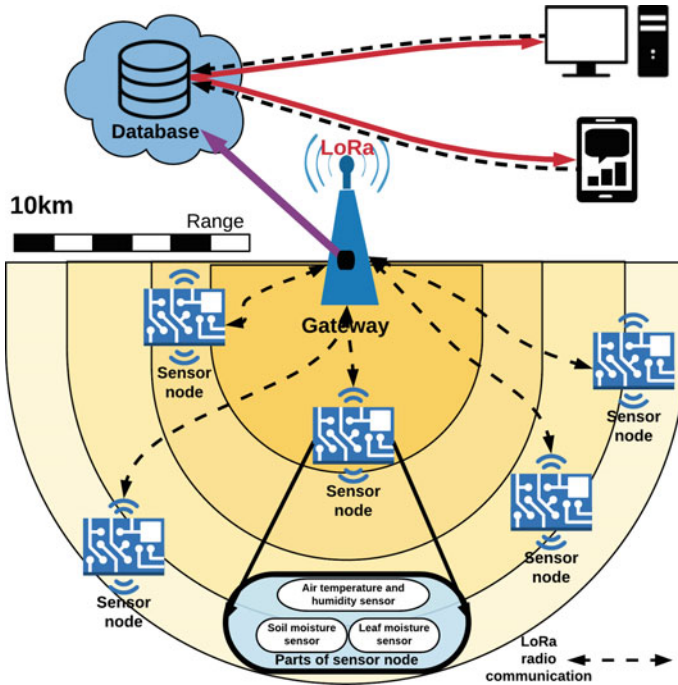


Fig. 2 The proposed eAgrar architecture based on LoRa LPWAN network

to the database or to the end user (application or Web page). Radio communication between the end devices and centralized system is based on the LoRa (long range) radio module technology. According to the LoRa specification, the data can cover the distance range up to 10 km, which makes this technology suitable for our implementation. Assuming that the range is only half of the predicted capability, i.e., 5 km, it would still give us a coverage of almost 80 km².

Implementation of eAgrar system. In this section, we give implementation details of our system that will ensure a lifetime of minimum one year without replacing the battery. For the implementation of end sensor device of our eAgrar system, we used Arduino Pro Mini with ATmega328p microcontroller that operates on 3.3 V with 8 MHz clock speed. We made slight modifications on Arduino Pro Mini board [15] to minimize consumption where we physically removed all LEDs and the voltage regulator. To minimize the consumption during the sleep period of Arduino (inactive state), we used low-power library from roocketscream [15]. As we show later, with low-power library we can reduce Arduino consumption below 1 μ A. To wake-up Arduino from deep sleep we used a low-power/low-cost RTC clock—DS3231 [18] according to the given duty cycle.

We used HopeRF RFM95 radio module for communication between the end sensor devices and the gateway. This module allows us to communicate at large distances (up to 10 km) using LoRa modulation. The transmission power goes up to 20 dBm

with 125 mA consumption during this time period. Module also has capabilities to use FSK, GFSK, MSK, and GMSK modulation. Due to its small price (less than 5 USD), HopeRF presents a suitable and preferred solution for message transmission over large distances.

We used BME280 sensor for temperature, humidity, and pressure measurement. This sensor is particularly characterized by low-energy consumption (only 3.6 μ A in active mode, and 0.1 μ A in sleep mode), with an extremely fast response time that supports new application requirements and high precision in a wider range of temperatures. Next, we used FC-37 sensor module capable of detecting the amount of moisture on the surface of the leaf. To measure the humidity of the ground we used FC-28-D sensor device. This is a high-quality sensor that signals whether the humidity percentage in the ground is high or low. At the end, we used a lithium (Li-SOCI2) 3.6 V battery with a 9000 mAh capacity. Such a battery allows us to create an autonomous electronic device that operates a large time period. We used this battery to directly supply our microcontroller without any prerequisite for voltage regulator. Figure 3 (left) shows the implementation of our end device.

For the implementation of eAgrar gateway device, we used Raspberry Pi 3 (RPi3) and Arduino Pro Mini operating on 5 V. HopeRF RFM95 that supports LoRa radio communication was connected to the Arduino microcontroller for sending and receiving messages from end sensor devices. This way, Arduino simply forwards all received data over a serial port to the RPi. The main logic on RPi is written in Python that sends data to the central mySQL database placed in a cloud. To establish a communication with the cloud system in areas without wired Internet connection, we used HUAWEI E3131 3G/UMTS Surfstick. We also used 50 W solar panel along

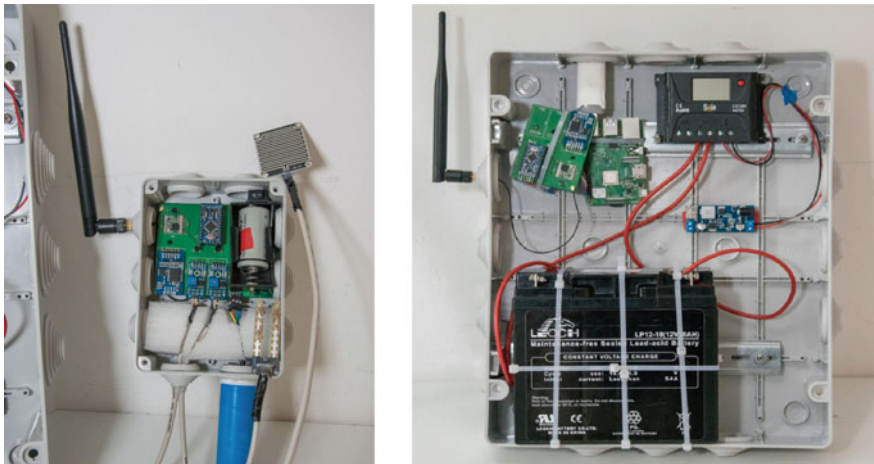


Fig. 3 (left) LPWAN end device used for the implementation of eAgrar network, (right) LPWAN prototype of a gateway system

with a voltage regulator. Figure 3 (right) shows the implementation of a gateway device (without external solar panel).

5.2 Preliminary Results

Consumption—Fig. 4 shows current consumption within active state of sensor device. We measured the consumption with Handyscope HS6 DIFF oscilloscope. As shown in Fig. 4, the external RTC wakes up Arduino microcontroller from deep sleep (period 1 in Fig. 4), after which Arduino powers up all sensors (period 2). After reading sensor values, Arduino sends a message to the gateway device (period 3), and waits back for the response (period 4). If end sensor device does not receive a response within a predefined time period, it retransmits a message. We limit the number of retransmissions to three messages leading to a max. four messages sent over a radio prior going to deep sleep state. Summarizing the total active state period, it bounds to approximately 1.5 s. The average consumption within active state equals:

$$I_{\text{active}} = \frac{0.55 \text{ s} \cdot 10 \text{ mA} + 0.7 \text{ s} \cdot 50 \text{ mA} + 0.05 \text{ s} \cdot 125 \text{ mA} + 0.3 \text{ s} \cdot 55 \text{ mA}}{1.6 \text{ s}} \quad (1)$$

$$= 39.5 \text{ mA}$$

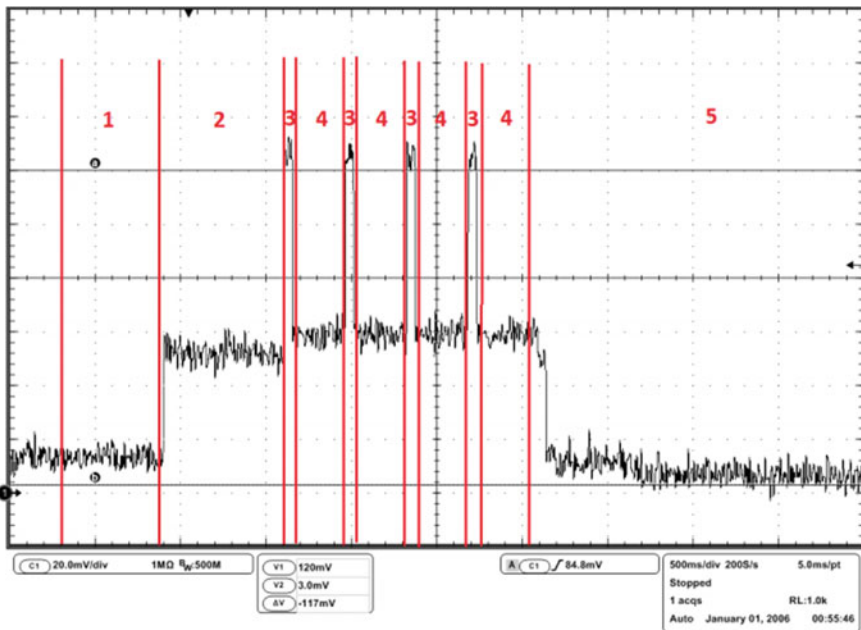


Fig. 4 Consumption of end sensor device of the proposed eAgrar system

where 10 mA denotes consumption after waking up from the sleep state, 50 mA denotes consumption during sensor reading period, while 125 mA and 55 mA denote consumption during message transmission and reception, respectively. Within sleep state (point 5), the average consumption equals $I_{\text{sleep}} = 0.3$ mA. If we consider that the readings are sent to the gateway every 20 min, the average consumption of end device will be $I_{\text{avg}} = 0.332$ mA. Taking into account that battery capacity falls due to self-discharge by 10–15%; we can assume that total battery capacity is around 7650 mAh. From the above analysis, the average lifetime of end sensor device equals 23,042 h = 960 days = 2.63 years. Taking into account battery self-discharge, possible retransmission, it is a realistic assumption that our prototype-based low-cost LPWAN device can successfully operate within a period of one year.

Measurements—Recall, we used LoRa radio module because of its ability to send data over large distances (over 10 km). For the purpose of our test, we placed gateway device 1.5 m high from the ground at fixed position and tested its coverage moving away end device from the gateway. We used the ping-pong method where one device sends the message and the other responds. With the lowest output power that we were able to send data from the distance up to 800 m. After boosting the output power to the strongest available 20 dBm, we were able to send and receive messages from 7.3 km. Due to the uneven terrain were unable to send messages over larger distances.

6 Conclusion

In this paper, an analysis about the state-of-the-art technologies both in sensing and data transmission is described, while stating and discussing the most important challenges that IoT devices should meet. Given the technical arguments, to implement the specific use-case scenario, it is necessary to investigate which hardware options would satisfy one needs. Being optimal in that sense would increase the battery lifetime and improve the quality of service. Finally, an example of IoT device implementation, based on Arduino platform and LoRa-based radio module is provided where the most significant results were analyzed.

Acknowledgements This work was partially supported by the Croatian Science Foundation under the project “Internet of Things: Research and Applications”, UIP-2017-05-4206; by National Funding from the FCT—Fundação para a Ciência e a Tecnologia through the UID/EEA/50008/2013 Project; by Finep, with resources from Funttel, Grant No. 01.14.0231.00, under the Centro de Referência em Radiocomunicações—CRR project of the Instituto Nacional de Telecomunicações (Inatel), Brazil; by Finatel through the Inatel Smart Campus project; by Brazilian National Council for Research and Development (CNPq) via Grant No. 309335/2017-5.

References

1. Ericsson, *Cellular Networks for Massive IoT*. Ericsson white paper, Jan 2016
2. D. Evans, *The Internet of Things: How the Next Evolution of the Internet is Changing Everything*. CISCO white paper, Apr 2011
3. O. Vermesan, P. Friess, *Internet of Things—from research and innovation to market deployment*, vol. 29 (River Publishers Aalborg, 2014)
4. V. Gazis, M. Goertz, M. Huber, A. Leonardi, K. Mathioudakis, A. Wiesmaier, F. Zeigerh, Short paper: IoT: challenges, projects, architectures, in *18th International Conference on Intelligence in Next Generation Networks (ICIN)*, 17–19 Feb 2015
5. W. Shyr, C. Lin, H. Feng, Development of energy management system based on Internet of Things technique. *Int. J. Electr. Comput. Energ. Electr. Commun. Eng.* **11**(3) (2017)
6. P. Šolić, Z. Blažević, M. Škiljo, T. Perković, Enabling IoT through Gen2 RFID: PHY/MAC research opportunities, in *6th International EURASIP Workshop on RFID Technology (EURFID)* (2018)
7. A. Kurawar, A. Koul, P. Viki, T. Patil, Survey of bluetooth and applications. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* **3**(8), 2832–2837 (2014)
8. M. Weyn, G. Ergeerts, L. Wante, C. Vercauteren, P. Hellinckx, Survey of the DASH7 alliance protocol for 433 MHz wireless sensor communication. *Int. J. Distrib. Sens. Netw.* **9** (2013)
9. U. Raza, P. Kulkarni, M. Sooriyabandara, Low power wide area networks: an overview. *IEEE Commun. Surv. Tutor.* (2017) (Online)
10. Nokia, LTE evolution for IoT connectivity (2015). Online: <http://resources.alcatel-lucent.com/asset/200178>. Accessed 16 Apr 2017
11. J. de Carvalho Silva, J. JPC Rodrigues, A.M Alberti, P. Solic, A.L.L. Aquino, LoRaWAN—a low power WAN protocol for Internet of Things: a review and opportunities, in *2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)* (2017)
12. N. Barroca, L.M. Borges, F.J. Velez, F. Monteiro, M. Gorski, J. Castro-Gomes, Wireless sensor networks for temperature and humidity monitoring within concrete structures. *Constr. Build. Mater.* **40**, 1156–1166 (2013)
13. K.-Y. Lian, S.-J. Hsiao, W.-T. Sung, Intelligent multi-sensor control system based on innovative technology integration via ZigBee and Wi-Fi networks. *J. Netw. Comput. Appl.* **36**(2), 756–767 (2013)
14. M. Bešlić, T. Perković, I. Stančić, G. Pavlov, M. Čagalj, eMooring: distributed low power wide area system to control moorings, in *2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)* (2017)
15. How to run Atmega328p for a year on coin cell battery. Online: <http://www.home-automation-community.com/arduino-low-power-how-to-run-atmega328p-for-a-year-on-coin-cell-battery/>. Accessed June 2018
16. Open Source Building Science Sensors (OSBSS), A low-cost Arduino-based platform for long-term indoor environmental data collection. *Build. Environ.* **100**(2), 114–126, (2016)
17. Adafruit TPL5110 Power Timer Breakout. Online: <https://learn.adafruit.com/adafruit-tpl5110-power-timer-breakout/>. Accessed Nov 2018
18. Extremely Accurate I2C-Integrated RTC/TCXO/Crystal. Online: <https://datasheets.maximintegrated.com/en/ds/DS3231.pdf>. Accessed June 2018

Small Cells Handover Performance in Centralized Heterogeneous Network



Raed Saadoon, Raid Sakat, Maysam Abbod and Hasanein Hasan

Abstract Before the full release of 5G, 4G will continue to be developed to support many new use cases and applications. By making some modifications and enhancements on its structure, 4G can handle many new features that could be seen as 5G specifics. Such enhancement includes the use of small cells in heterogeneous networks, mobile edge computing, and cloud computing, virtualizations and software-defined networks (SDN), and network slicing. Heterogeneous network supported by computing functionality is one of the most recent added flavor topologies to LTE technology and the upcoming next-generation mobile network. Mobility is yet one of the main challenges in such a system. Allowing the user to surf the world of the data without interruption while in the move between the different cells of the heterogeneous network in the future 5G network is required to keep the high level of the user's quality of experience. In this paper, new mobility-enhanced schemes are presented for improving the handover performance of a mobile UE with dual connectivity in heterogeneous network.

Keywords 5G · LTE · HO · HetNet · Mobility

R. Saadoon (✉) · R. Sakat · M. Abbod · H. Hasan
Collage of Engineering, Design and Physical Sciences, Brunel University London, London, UK
e-mail: raed.saadoon@brunel.ac.uk

R. Sakat
e-mail: raid.sakat@brunel.ac.uk

M. Abbod
e-mail: maysam.abbod@brunel.ac.uk

H. Hasan
e-mail: 0732799@alumni.brunel.ac.uk

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_37

1 Introduction

Passing through its different generations from 1G to 4G, evolution of mobile network examine different fundamental changes and challenges. Systems evolved from analog network providing voice-only service to an all IP packet core network providing multimedia services. During this evolution, the structure of the mobile network itself passed through different changes, in the two main parts of the structure; the core network and the radio network. The way the mobile user accesses the network was also affected by this evolution. In addition, the experience of the mobile users accessing the network also played a role in this evolution. The mobile user is the end user of the network and it is the customer of the mobile network operators (MNO); therefore, keeping a good level of experience is a fundamental requirement in order to keep a good level of revenue. Customer satisfaction is a need for the mobile network operators (MNO), and as the mobile networks evolved, the customers become more demanding and keeping the level of satisfaction became more challenging. Mobility is one of the most attractive attributes of the mobile network; the ability of the mobile user to move through the network and benefiting from the services is one of the charming characteristics that keep the user attracted to the mobile networks, at the same time this attribute represents one of the most challenging areas for the mobile network operators. In order to keep the level of satisfaction, the mobile network operators must provide the mobile users with seamless connectivity and all-time services, especially as the mobile devices are no longer a luxury and become an essential part of people's everyday life and business. People are using their mobile devices while on the move for different purpose and needs, they are surfing the Internet, checking their e-mails, connecting with each other through social media, and streaming audio and video feeds. The integration between mobile networks and computer networks with the advanced capabilities of the devices led to big amount of data being generated, and the vast amount of this data is due to mobile networks and the attached devices, such as mobile smartphones, tablets, wearables, and many other devices. According to Cisco Visual Networking Index (VNI) 2017 [1], by 2021, that is, one year after the 3GPP to submit the final specifications of the 5G at the ITU-R WP5D meeting in February 2020, 58% of the world population will be using the Internet, on average of 3.5 networked devices and connections per person, and global IP traffic will reach 3.3 Zettabyte. It also notes that traffic from wireless and mobile devices will represent 63% of total IP traffic, smartphones will exceed 86% that is four-fifths of mobile data traffic, and over 78% that is three-fourths of the world's mobile data traffic will be video. Based on the latter and taking the requirement for the 5G mobile network into consideration, which includes higher connection speed of up to 10 Gbps, latency of about 1 ms, increasing the bandwidth per unit area, 100% coverage and almost the same for availability, it can be seen that the current structure of the 4G represented by the structure of LTE-A will not be able to cope with such requirements, but could be the base for the future mobile network or 5G; therefore, and as a contribution toward the design of 5G, we aim to design a reliable HO technique for mobile UEs in heterogeneous network where the small cells are used for data only delivery. The

system is based on the current LTE-A architecture and adding computation functionality to the design to handle the SCs HO in order to reduce the delay and to increase the throughput.

1.1 Heterogeneous Network

Unlike the homogeneous network that consists of macrocells only, a heterogeneous network is composed of multiple site types, i.e., macrocells, microcells, picocells, femtocells, and either Wi-Fi hotspots. Such network architecture produces challenges in terms of co-channel interference and RF planning, as the various types of BTS typically sharing the same channel bandwidth [2]. One of the obstacles in the heterogeneous network is the intercell interference mainly in the boundaries of the cells that can cause handover problems. Solutions such as intercell interference coordination (ICIC) and enhanced intercell interference coordination (eICIC) techniques as specified by 3GPP specifications in release 8 and 10, respectively, can be used to mitigate such issues [3]. In ICIC techniques the signal to noise ratio (SNR) is improved at the cost of reducing the total number of resource blocks, i.e., the bandwidth available for transmission, by avoiding the use of the resource blocks used by neighboring eNodeBs. eICIC also improves the SNR by benefiting from almost blank subframe (ABS), which allow neighboring eNodeBs to reduce their transmission during certain time intervals. In addition small cells can be used to provide dedicated services in terms of network slicing [4].

1.2 Small Cells

A promising solution that can cope with mobile traffic explosion is the use of the small cells (SC). In 3GPP TR 36.932 [5], a node that transmits power that is lower than the BS classes of macro node can be considered as small cell. In such situation, pico and femto cells that have a transmit power of 0.25 W and 0.1 W, respectively, are both considered as small cells, which is not the case for microcells. 3GPP does not specify a separate power class for microcell. A wide area BTS, i.e., macrocell with reduced transmit power could be designed and considered as microcell. In some other scenarios, Wi-Fi hotspots [6] and remote radio heads (RRHs) within a centralized radio access network (C-RAN) are also considered for small cells deployment options.

As in [7], the solution scenarios target the deployment of small cells with and without the coverage of macrocells, indoor and outdoor hotspots, with both ideal and non-ideal backhaul, with the possibility of small cells to use the same or different frequency band when under the macrocell layer. The distribution technique is also an important factor; therefore, the spars and dense distribution are also considered within the 3GPP TR. Study on the impact of each architecture option will lead to

better decision on how to deploy small cells and the service that could be delivered, taking in consideration the users' locations and dual connectivity in heterogeneous network.

1.3 Dual Connectivity

Dual connectivity is an interesting and important extension to the carrier aggregation (CA) principle that was adopted in release 12 of the 3GPP. CA is the most successful feature of the LTE-advanced system that was introduced in 3GPP release 10. CA increase the channel bandwidth by combining multiple RF carriers coming from the same co-located eNB, by giving the UE the ability to receive and transmit multiple signals in both directions, i.e., downlink and uplink [2, 6]. With dual connectivity the UE will have the ability to be connected with both macro and small cell simultaneously. This will allow the maximum data rate to be aggregated from the macro and small cells layers. With such feature small cells could be used in the most efficient way as the macrocell could be used to maintain the coverage keeping the UE connected all the time to the system even in the absence of the small cell layer coverage. 3GPP TR 36.842 [8] provides different architecture options for the deployment of such heterogeneous network that uses small cells layer and UE dual connectivity feature. Benefit of such architecture, based on different scenarios include: continuous connectivity, reduce signaling overhead with core network, increase the peak data rate, and the possibility to use the small cells layer for dedicated services such as content delivery. Although such architecture brings many benefits, yet it brings many challenges as it increases the system complexity. Challenges such as: efficient distribution of the small cells and the discovery scenario of such small cells by the UE that may exhaust the battery power, the miss measurement, and exchange of the signals between the macro and small cells layers may lead to a ping-pong situation and handover problems during the mobility. Implementing such architecture of heterogeneous network in which the UE with dual connectivity can be connected to different sites simultaneously (i.e., with the MeNodeB for continuous connection to the core network and with the small cells for dedicated purpose) may enhance the system performance to higher levels by increasing the capacity and reducing the latency and system messaging overhead as the small cells will be in direct connection and fully controlled by the MeNodeB without huge intervention from the core network as shown in Fig. 1.

2 Related Work

Wireless and mobile networks have an increasing impact in the world. Smart devices connected to such networks become more powerful and are evolving in a way that surpasses the evolution of the mobile network systems, benefiting from the combination of the computing technology, and the powerful capabilities of the software

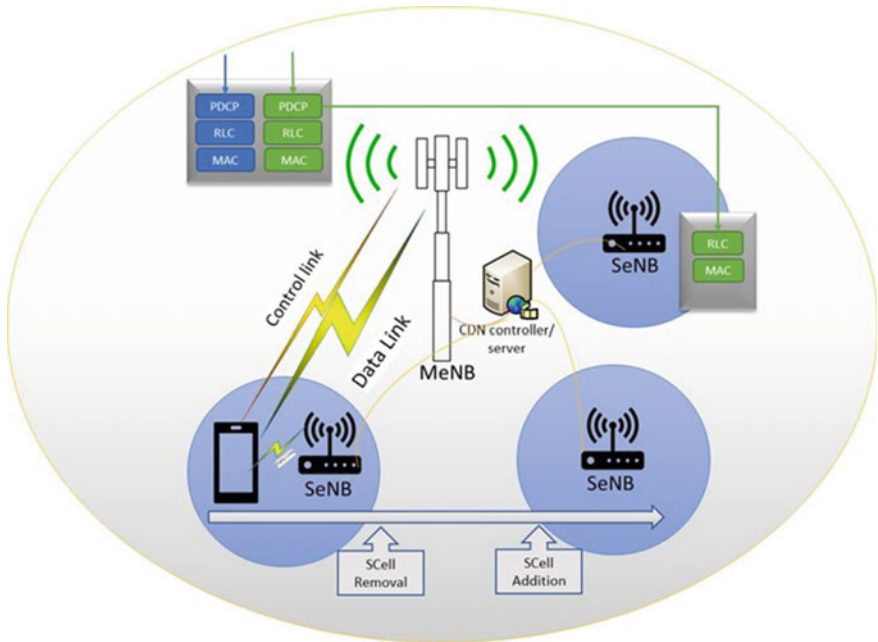


Fig. 1 System architecture

programs. These powerful mobile devices become the primary point of access to the Internet system. Wireless mobile devices can be found everywhere, it can be found in our homes, cars, and offices and it is wearable, portable, and attached to mobile equipment and devices like drones. The attractive features of such devices that make them popular and widely adopted by users are their lightweight, wireless connection, and mobility. While in the move, devices with such exiting features are exchanging information and generating huge amount of data. Taking into consideration the new applications and services provided to the mobile users by different service providers, in addition to the use of data in different fields by different parties such as marketing, results in creating a more demanding users that require all-time connection with more bandwidth. The latter becomes a burden on the current wireless and mobile system, and therefore the design of next-generation mobile network must take these issues into consideration. In order to support such demands, the next-generation mobile network will be multipurpose systems with network slicing that provides dedicate services specific to the need of each customer. In order to implement such network, mobile systems need to merge and benefit from the capabilities of the computer system in addition to other access network techniques. As the next-generation mobile network will be a key enabler to other systems and the base for the Internet of things (IoT) it should be designed to integrate networking, computing, and storage resources into a programmable unified infrastructure that will allow for an optimized usage of

resources and be able to process the huge amount of generated data by mobile users [9].

This demand encouraged many companies and institutes to group together in an alliance or working groups, in an effort to provide their vision for the next-generation mobile network, in addition to many academic works presented for the same purpose. Most of the work is based on the system architecture of the current LTE-A system and trying to solve the problem of the data delivery, especially when the user is moving. A paper presented by Ericsson to the 3GPP RAN suggests a technique for CP and UP separation between MeNodeB and SeNodeB to increase mobility robustness and minimize UE context transfer [10]. At the same event FiberHome Technologies Group, presented a study for small cell discovery in HetNet based on existing uplink signal, based on cell listening to reduce signaling overhead [11]. In [12], the authors presented a work for small cells dual connectivity with bearer split in the uplink direction in order to increase the user throughput by means of coordination between the cells. In the same scenario, authors in [13] proposed a flow control algorithm to forward the data from the MeNodeB to the SeNodeB, results showed that there was a trade-off between user throughput and latency. The authors in [14] presented a scenario for dual connectivity in which the processing of the data and the control of the SC is within a controller on the edge of the network, i.e., mobile edge computing (MEC), the results showed that the proposed scheme can reduce the packet loss and latency when routing full load in the network. In terms of computing capabilities to support next-generation mobile system, separation of CP and UP techniques supported by software-defined network (SDN) to enable content caching was presented in [15] where the separation of CP from UP will eliminate the need for mobile dedicated solutions as it removes the mobile tunneling to allow dynamic relocation of the cache. In the road to 5G, a cache-enabled wireless heterogeneous network (HetNet) with (SDN) and split of the CP and UP is proposed in [16], where the macrocell and SCs with different cache abilities are overlaid and cooperated together in the backhaul. Based on the results, it was found that the proposed network has much higher throughput and energy efficiency than current LTE networks. Fundamental trade-offs exist between the throughput and the density of SCs.

3 Architecture

In this section, the architecture of the HetNet system for UE's with DC where the mobility is controlled by the centralized MeNodeB will be explained with the fundamental principles of parameters and measurement involved in the design of the system.

3.1 System Model

The architecture is based on the 3GPP LTE-A Evolved Packet System (EPS), with the same main components for radio and core networks (release 12) [8]. Small cells will be distributed as hot spots covering specific areas under the coverage of the macrocell layer, The small cell layer with frequency (F2) will be located at the center of the hot spot, where the macro with frequency (F1) will be like an umbrella covering the small cells layer. In our hypothesis, the macro and small cells layers are connected via an ideal backhaul.

As in Fig. 2 that shows the division of control between MeNodeB and SeNodeB in DC, a UE with dual connectivity will maintain an RRC-connection with the MeNodeB at all times, while receiving user plane data from the MeNodeB and SeNodeBs. Hence, there will be only one S1-MME connection per UE. The main services required by the UE will be the responsibility of the MeNodeB as it keeps all-time connection with the UE, while SeNodeBs could be used for specific services such as content delivery. In such way, the mobility of the UE will be controlled through the MeNodeB, as the MeNodeB will be responsible of the RRC signaling with the MME where there is no need to move the RRC context of the UE between the SeNodeB's. In this manner, handover between SeNodeB's will be like adding and removing new cells as all the information and resource management (RRM) are in the MeNodeB

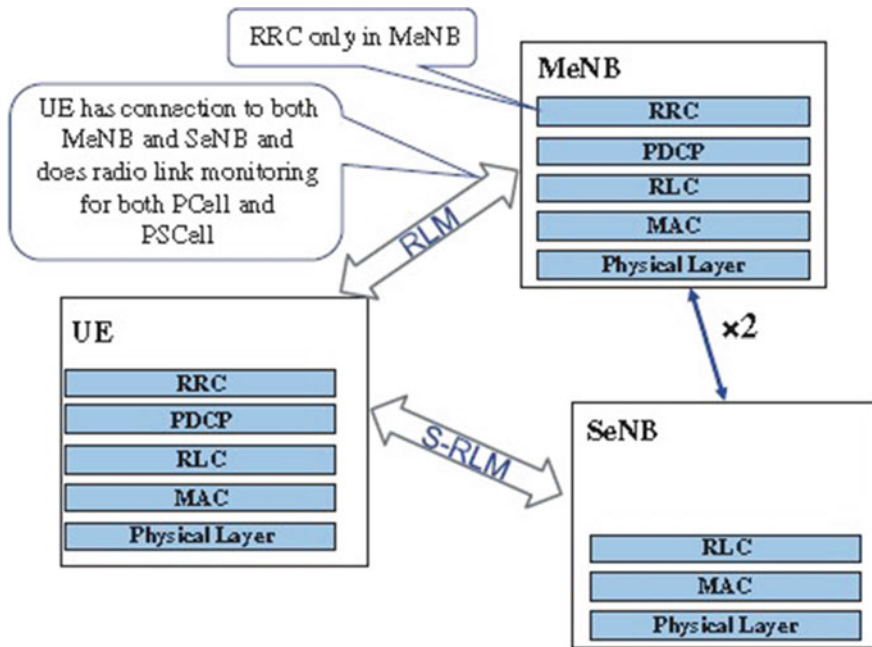


Fig. 2 Division of control between MeNodeB and SeNodeB [6]

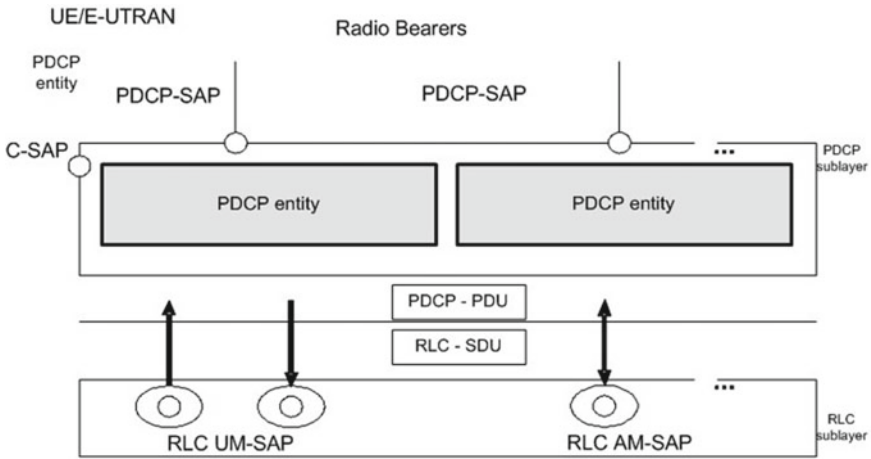


Fig. 3 PDCP layer, structure view [17]

server. The system will be a centralized radio access network similar to the C-RAN with RRH; but, instead of the RRH that just have a RF layer, the SeNodeBs will have a protocol stack that eliminates the need for the RRC and Packet Data Convergence Protocol (PDCP) layers and have the RLC, MAC, and PHY layers. The service data units (SDUs) coming from the PDCP layer are then encapsulated and directed to the RLC layers in each eNB that will convert them to protocol data units (PDUs). Passing through the rest of the protocol stack, UE will receive the data from one or more eNB, assuming the scheduler in the MeNodeB server has shared the total set of the resource blocks, separate IFFT functionality could be used to separate between the signals. The (PDCP) layer (Fig. 3) within its responsibility is to transfer the user plane and control plane data by the mean of PDCP entities, several PDCP entities may be defined for each UE, and each PDCP entity is carrying data for one data bearer.

In terms of mobility, a UE in RRC connected mode relies on network controlled UE assisted mobility, in which the network informs the UE via a dedicated RRC signals when it has to perform handover process to the required cell. When a UE follows a certain trajectory, while in dual connectivity, it can have its PCell on the macro layer, i.e., MeNodeB, while SCell can be configured within the small cell layer, i.e., SeNodeB. Two types of handover could occur within such architecture, between different MeNodeBs and between different SeNodeBs. The assumption is that MeNodeB handover is based on the RSRP A3 event as the neighboring cells in A3 reporting event can be either intra-frequency or inter-frequency which can be well serve the PCell in the MeNodeB, while addition and removal of SCells that is SeNodeBs intra-frequency handover for under the MeNodeB coverage is based on the A6 event trigger. A network will keep control of the mobility of the UE as the MeNodeB is controlling the C-plane. Consider adding the computation and storage capability from other working groups, e.g., ETSI MEC, a CDN network controller could be

located within the macrocell layer (MeNodeB). Information about the network and its users will be gathered at the controller. Such information includes mobility, number of users per cell, list of candidate cells for under the MeNodeB coverage. This can provide significant impact in the UE data rate. When those SeNodeBs are well known by the control server of the MeNodeB and sharing the same C-plane, all the information necessary to provide the required service can be started directly after the UE is under the SeNodeB coverage. The triggering criteria for SCell addition is configured by the MeNodeB controller by mean of certain event (i.e., A3 and A6), in this case, RRM measurement is not reported to the core network as it is already the MeNodeB's server responsibility. Similar architecture for controlled RRM and handover between small cells could be a promising method to reduce signaling overhead toward the core network, as well for increasing the data rate per UE.

3.2 Reporting Measurements

Reporting of the measurement required to perform the HO and event triggering are based on the RSRP and RSRQ made by the UE and then reported to the eNodeB as defined in the 3GPP specification in [TS 36.133 and TS 36.214]. The RSRP is the average power received by the UE from a single cell and can be measured as:

$$\text{RSRP}_{i,\text{UE}} = P_i - L_{\text{UE}} - L_f \quad (1)$$

where

$\text{RSRP}_{i,\text{UE}}$ is the reference signal received power for a UE received from cell i
 P_i is the transmit power for eNodeB i
 L_{UE} path loss gain from the UE to the source eNodeB
 L_f is the fast fading channel gain.

RSRQ is the reference signal received quality and can be measured as:

$$\text{RSRQ} = N \times \frac{\text{RSRP}}{\text{RSSI}} \quad (2)$$

where N is the number of resource blocks (RB).

RSSI is the carrier received signal strength indicator of the total received power from all sources around the UE, serving and non-serving cells, including interference and noise. RSSI is measured as:

$$\text{RSSI} = \text{RSRP}_{s,\text{UE}} + \text{RSRP}_{\text{int,noise}} \quad (3)$$

Hence, assuming that the UE measures the reference signals from the neighboring wireless channels on periodic basis; the collected measurement passes through a processing mechanism of averaging and filtering before any event to be triggered.

After applying the layer 3 filtering the processed measurements are to be reported to the MeNodeB where the handover decision is to be made. The processing mechanism of layer 3 filtering eliminates the fading effect and provides more stable measurement values by estimation inaccuracies from the reported measurements to ensure the accuracy of handover decision and can be calculated using the below equation:

$$R_{L3}^n = (1 - \alpha)R_{L3}^{n-1} + \alpha \cdot M_n \quad (4)$$

where

R_{L3}^n is the updated filtered measurement result
 R_{L3}^{n-1} is the older filtered measurement result
 $\alpha = 1/2^{(K/4)}$ where K is the appropriate filter coefficient that can configured independently for RSRP and RSRQ
 M_n latest received measurement result from physical layer.

Layer 3 filtering is applied only in RRC connected and is applied during handover to ensure no handover is triggered due to bad measurements as no event is to be triggered before evaluated by layer 3 filtering.

4 Performance Evaluation

After introducing some basic terminologies and concepts, a study of the UE behavior moving under the coverage of MeNodeB and receiving data from a content server attached to that MeNodeB while experiencing handover procedure between the small cells will be presented. The study is based on the system architecture explained in Sect. 3 of this work and will be implemented using Riverbed modeler formally well known as OPNET modeler.

4.1 Simulation Setup

The Riverbed LTE Modeler 18.7 is based on the 3GPP Rel.8; therefore, a modification to the system has to be made to support dual connectivity using the device creator operation to create custom LTE nodes for our network [OPNET documentation create Custom Device Model Operation]. The handover process state for the UE is reside in the (lte_as) node mode which also contains the RRC-connection states and the measurement states as shown in Fig. 4. The modification to the (lte_as) states will require adding new states as secondary link states in similar way to the existing states in order to serve the requirement of dual connectivity by the UE. In this case, the UE will have only one RRC-connection state with the MeNodeB that keeps it connected to the system and have the secondary states required for DC, as explained in Fig. 2.

Δ is the A6 offset.

Data forwarding, while handover is supported and the SeNodeB modification event is used by the MeNodeB to modify, establish or release bearer contexts. The report interval function and num reports attributes within the Riverbed modeler govern the number of reports sent to the MeNodeB. UE will perform cell search and generate reports, when the reported measurements violate the handover triggers, the MeNodeB decides to handover the UE to a different cell from a list of the SeNodeB in its attached server. The modification of small cells handover procedure is shown in Fig. 5.

The system performance of a UE under the coverage of MeNodeB and receiving data from both MeNodeB and one of the SeNodeBs while moving in a specified defined trajectory that insures the UE will be under the coverage of one SeNodeB within the MeNodeB coverage area will be examined based on the parameters in Table 1.

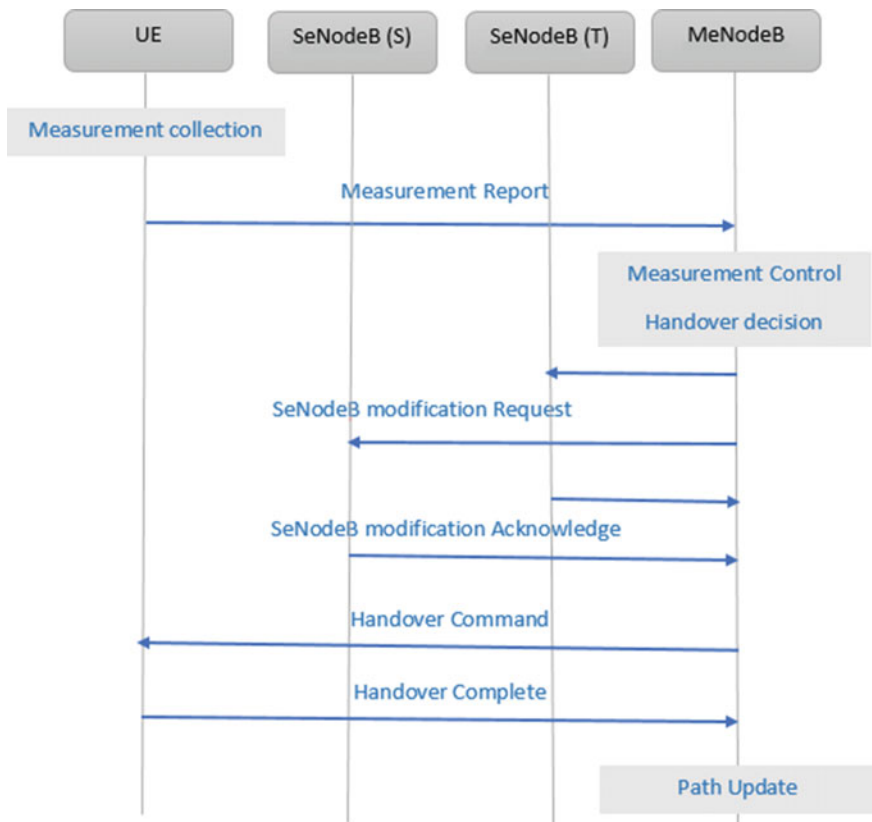


Fig. 5 SeNodeB modification procedure

Table 1 Simulation parameters

Parameters	Values
Scenario	Scenario#2 in the 3GPP TR 36.842 V12.0.0 (2013-12)
Deployment	1 three-sector MeNodeB site 4 SeNodeB sites per macrocell area
MeNodeB carrier frequency (F1)	2 GHz
SeNodeB carrier frequency (F2)	3.5 GHz
LTE B.W./duplexing	20 MHz/FDD
MeNodeB Tx power	46 dBm
SeNodeB Tx power	30 dBm
UE Tx power	23 dBm
Traffic model	HTTP browsing with heavy load video
File inter-arrival time	Exponential
Service type	As requested by UE

4.2 Result Discussion and Analysis

In this work, we consider a HetNet scenario with small cells deployed under the coverage of MeNodeB. The scheme proposed is analyzed based on the discussed settings and scenarios explained in the proceeded sections of this paper.

In the first scenario Mobile_UE initially is associated with eNodeB_0, serves as MeNodeB, and to eNodeB_3, serves as SeNodeB. eNodeB_0 is configured to receive periodic measurements from the Mobile_UE. Trajectory is set for the UE for mobility within the MeNodeB coverage area and UE configured to perform cell search and select the best suitable SeNodeB, the UE model supports the measurement of RSRP for each cell to aid in the cell selection process. RSRP is measured for the MIB packets. Cell selection threshold that is configured on the SeNodeBs is compared with the scanned SeNodeB RSRP.

Figure 6 shows RSRP distribution during mobility. When UE moves and approaching another small cell coverage area, the macrocell will trigger measurement reporting to the UE and the UE starts to generate reports to perform handover. This appears in Fig. 6 at times 4, 5, 6, and 7 min, at these times UE will perform HO trigger between small cells, while keeping attached to the MenodeB, as shown in Fig. 7, as the red line represents the MeNodeB to which the UE is attached for the simulation period, while the blue lines represent the SCs. As seen in the results, handover is performed when the serving cell measurements reported by Mobile_UE violates the handover triggers. As explained earlier in the simulation setup based on Eq. (5).

The key performance factors chosen for the next scenarios are the throughput and accumulative network burden.

It can be observed that the throughput in bits/sec and the response of the network is acceptable when the UE is connected to the MeNodeB provided that no association

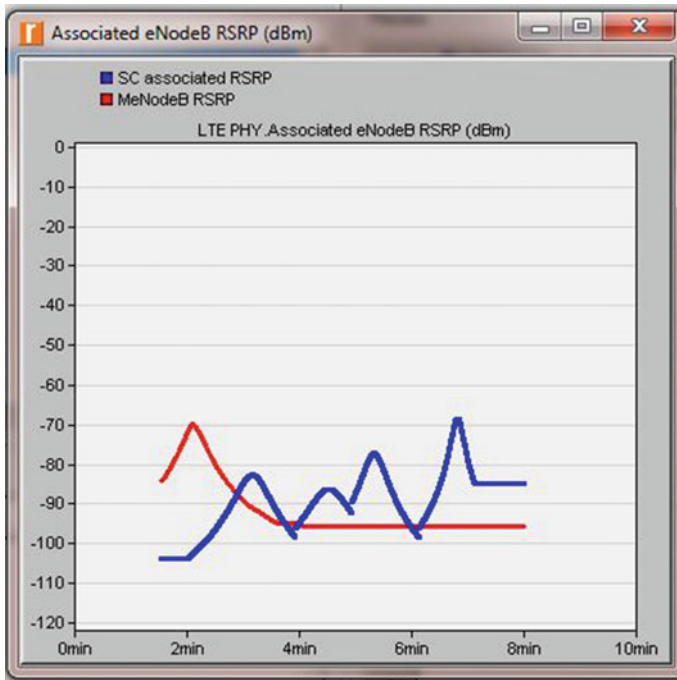


Fig. 6 HO region based on the RSRP

with any SeNodeB in the network is set, it jumps higher when the UEs start receiving data from SeNodeBs. This is the expected response as the burden decrease on the MeNodeB (i.e., more traffic is carried out by SeNodeB). The third scenarios are for UEs under the coverage of the MeNodeB without small cells represented by green, and with small cell with normal load represented by the blue color, and when the same network runs and routes non-live video content in a full load scheme in its data plane represented by the red color.

In theory, the throughput (bits/sec) increases in two cases:

- When the data traffic increases.
- When the elapsed time decreases.

Hence, in our model, as shown in Fig. 8, when the dual connectivity is triggered in the MeNodeB the network delivers and performs relatively slow at the beginning of the simulation. However, it starts to perform even better in the second scenario when the DL is connected to the SeNodeB, as the SC has its own resources that could be dedicated for the UE's under its coverage.

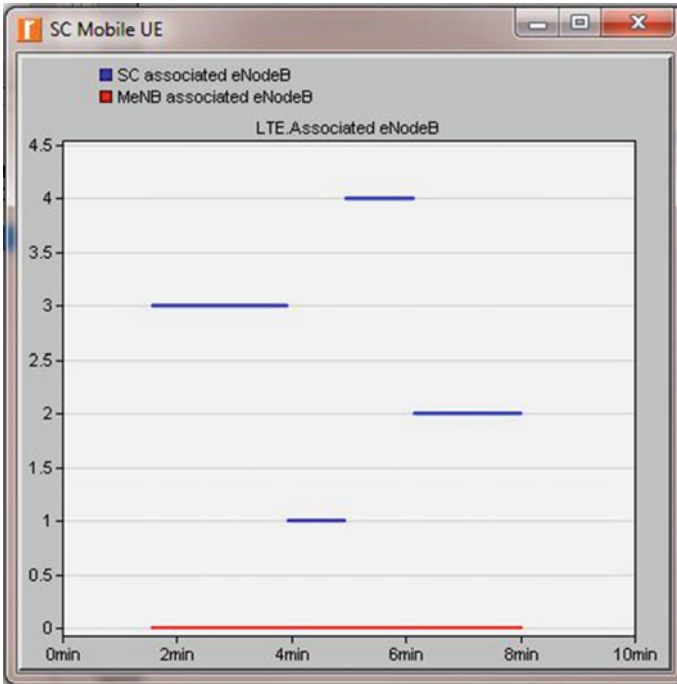


Fig. 7 LTE associated eNodeB

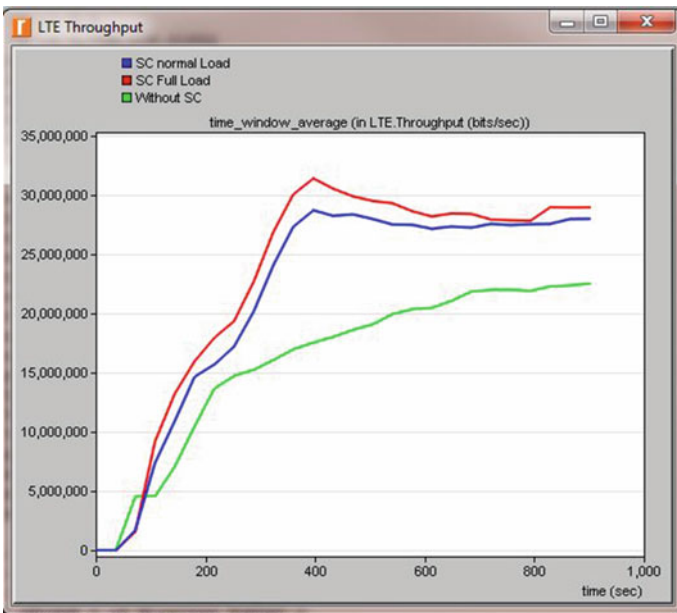


Fig. 8 Throughput

5 Conclusion

The increased demand of wireless data services driven by smart devices capabilities and mobile Internet led to an exponential growth on the traffic generated by wireless systems. In order to cope with this explosive growth of generated data, traditional macrocell only networks have evolved to heterogeneous networks (HetNet) in order to improve the performance of the system. In such system, the MeNodeB will act as the main point of connection for the UEs' while small cells are deployed to provide specific services for the users throughout DC. This architecture of user-centric network is considered as one of the most promising techniques for future 5G system. One of the main issues with such architecture is the frequent handover due to the large number of the small cells deployed under a relatively small area. In this paper, a centralized mobility management system was proposed to overcome such challenges. Under the proposed architecture, the macrocell (MeNodeB) is the anchor for UEs movement, while the other small cell will act as the SeNodeBs (SCs), which only provide service for the users under its coverage.

References

1. Cisco Visual Networking Index, *Global Mobile Data Traffic Forecast Update, 2016–2021* (Cisco, 2017). <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862>
2. C. Johnson, *Long Term Evolution in Bullets*, 2nd edn, ver. 1 (Northampton, England, 2012)
3. S. Sesia, I. Toufik, M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*, 2nd edn. (Wiley, 2011)
4. SCF197, *URLLC and Network Slicing in 5G Enterprise Small Cell Network* (Small Cell Forum, 2018)
5. 3GPP TR 36.932 V13.0.0 (2015-12)
6. H. Holma, et al., *LTE small cell optimization: 3GPP evolution to release 13* (Wiley, 2015)
7. 3GPP TR 36.932 V14.0.0 (2017-03)
8. 3GPP TR 36.842 V12.0.0 (2013-12)
9. R.S. Sadoon, Explosion of data (BIGDATA), in *Internet of Things and Big Data Analysis: Recent Trends and Challenges*. Nov 2016. ISBN-10:0692809929 (Chapter 3)
10. Ericsson, Tdoc R2-123706, 3GPP TSG-RAN WG2 # 83
11. FiberHome Technologies Group, Small cell discovery in HetNet based on existed uplink signal, 3GPP TSG-RAN WG2 #83 R2-132295, Barcelona, Spain, 19th–23th Aug 2013
12. S.C. Jha, K. Sivanesan, R. Vannithamby, A.T. Koc, Dual connectivity in LTE small cell networks, in *2014 IEEE Globecom Workshops (GC Wkshps)* (Austin, TX, 2014), pp. 1205–1210. <https://doi.org/10.1109/glocomw.2014.7063597>
13. H. Wang, C. Rosa, K.I. Pedersen, Dual connectivity for LTE-advanced heterogeneous networks. *Wireless Netw.* **22**(4), 1315–1328 (2015)
14. R. Saadoon, R. Sakat, M. Abbod, Small cell deployment for data only transmission assisted by mobile edge computing functionality, in *2017 Sixth International Conference on Future Generation Communication Technologies (FGCT)* (Dublin, 2017), pp. 1–6
15. J. Costa-Requena, M. Kimmerlin, J. Manner, R. Kantola, SDN optimized caching in LTE mobile networks, in *2014 International Conference on Information and Communication Technology Convergence (ICTC)* (Busan, 2014), pp. 128–132

16. J. Zhang, X. Zhang, W. Wang, Cache-enabled software defined heterogeneous networks for green and flexible 5G networks. *IEEE Access* **4**, 3591–3604 (2016). <https://doi.org/10.1109/access.2016.2588883>
17. 3GPP TS 36.323 V8.6.0 Rel. 8

CRISP-DM/SMEs: A Data Analytics Methodology for Non-profit SMEs



Jhon Montalvo-Garcia , Juan Bernardo Quintero 
and Bell Manrique-Losada 

Abstract The exponential increase in information due to technological advances and the development of communications has created the need to make decisions based on the data analysis. This trend has opened the doors to new approaches to data understanding and decision-making. On the one hand, companies need to follow data analytic methodologies to manage large volumes of information with big data tools. On the other hand, there are non-profit small and medium-sized enterprises (SMEs) that make efforts to address data analytics according to their different sources and types. They find challenges such as lack of knowledge in methodological and software tools, which allow timely deployment for decision-making. In this paper, we propose a data analytics methodology for non-profit SMEs. The design of this methodology is based on CRISP-DM as a reference framework, is represented by Software Process Engineering Metamodel (SPEM) and is characterized by being simple, flexible, and low implementation costs.

Keywords Data analytics · CRISP-DM · Non-profit SMEs

1 Introduction

The information processing according to Palmer and Hartley [1] has gradually become the basis for achieving a competitive advantage and, therefore, organizations have to believe that they have the right information at the right time and for the right people. The company's managers should be provided with the appropriate

J. Montalvo-Garcia (✉)

Seventh-day Adventist Church of Colombia, Cra 84 #33aa-169, Medellin, Colombia
e-mail: jhonmontalvo@gmail.com

J. B. Quintero · B. Manrique-Losada

Universidad de Medellin, Cra 87 #30-65, Medellin, Colombia
e-mail: juquintero@udem.edu.co

B. Manrique-Losada

e-mail: bmanrique@udem.edu.co

© Springer Nature Singapore Pte Ltd. 2020

X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_38

tools for the exploitation and data analysis that will allow them to obtain the necessary knowledge in the strategic decision-making process [2]. This is how in the last decade, data warehouses (DW) have become an essential component to achieve competitiveness with modern decision support systems in most companies that handle large volumes of data.

However, these companies only represent a minimal part of the business world. Small and medium-sized enterprises (SMEs) are the dominant form of business organization in all countries of the world, representing more than 95% and up to 99% of the business population [3]. SMEs are considered important at the local, national, and global levels, playing an important role in the national economy [4], and in the social sector; as non-profit companies, who pursue social and community purposes. Although they do not generate profit distribution or enrichment to the partners, strategic and financial decisions involving data analysis are made.

Therefore, non-profit SMEs collect information from different sources and are interested in business intelligence systems [5] and in the trend toward data analytics (DA) that is increasing by technological advance. Despite this interest, the development of data analytics projects is frustrated, since its implementation is usually a complex task due to the generated costs. These costs are associated with technological infrastructure, administrative costs, personnel training, and software tools [6]. Thus, the implementation of analytical projects in non-profit SMEs has few alternatives, since the DA focuses on large companies that have the greater financial capacity [7] and high-volume data management.

To reduce the gap between non-profit SMEs and DA projects, an analytical methodology is proposed according to the needs of these organizations. As a consequence, this paper is organized into four sections, starting with section two, which describes the background of non-profit organizations and the most recognized data analytics methodologies in the industry. Section three presents a simplified CRISP-DM proposal for non-profit SMEs. Finally, in section four, conclusions and future work are presented.

2 Background

2.1 Non-profit SMEs

Non-profit companies are legal and social entities created to produce goods and services, whose legal status does not allow them to be a source of income for the units that establish, control or finance them [8]. Therefore, they can benefit to associates, third parties or the general public from social projects. Non-profit has taken a great importance in the world, not only as organizations that provide social services but as generators of employment and promoters of economic activity. Because of this, they are grouped at an international level according to their social purpose [8] and

are inspected, monitored or controlled by different state entities such as mayoralities, governorships, and government ministries [9].

Non-profit companies have their main source of income from funds received by natural persons, legal entities or public entities representing countries, through donations or subventions. Regarding the organizational structure, non-profit companies have a high degree of complexity, due to the wide range of organizational possibilities (central, divisional, functional or geographical). Therefore, non-profit companies are recognized within SMEs as foundations, universities, corporations, associations, cooperatives, churches, and among others.

2.2 Data Analytics in Non-profit SMEs

The ability of SMEs to succeed in the face of larger competitors is centered on personal intuition and the ability to provide superior service. Since big data is changing the business landscape, some big competitors are using big data to improve product quality, marketing operations, and customer relationships. This new efficiency of the largest competitors can be a real threat to the sustainability of the SME business [10], especially for non-profit companies. However, there are other concerns in the SMEs that stand in the way of the growth of the DA and they are the new regulations imposed by the government every year. A clear example of this is the implementation of the International Financial Reporting Standards (IFRS), data protection policies, electronic invoicing, and strict social security regulations, which have increased in a large percentage from 2016 to 2017, as is the case of Colombia [11]. This has put aside the advance in digital transformation, a key aspect in the career of data analytics.

Despite these concerns, an EMC study reveals that 58% of Colombian organizations have current plans to implement big data technologies. The other 42% say that the lack of interest corresponds to the fact that the business culture is not ready yet (46%); but also, because it is very expensive to implement it with respect to the current economic situation (28%), and there is a lack of understanding regarding this trend (25%) [12]. Thus, to minimize the complexity, costs and lack of staff training, non-profit SMEs require that a DA project can be deployed quickly and that it can be easy to model and replicate to other areas of the organization. It also demands that the developed models can be easy to improve in front of any external change that affects the organization and being flexible for the integration of different data sources. Figure 1 shows the practices that non-profit SMEs need when implementing a DA project.

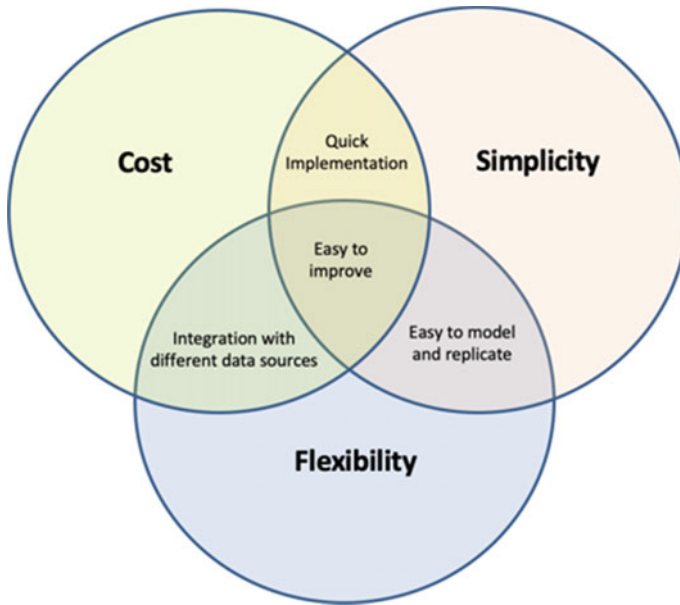


Fig. 1 Data analytics practices in non-profit SMEs

2.3 Data Analytics Methodologies

The most referenced models found in the scientific community and proposed for the development of DA projects are as follows: knowledge discovery in databases (KDD), sample, explore, modify, model, assess (SEMMA), and cross-industry standard process for data mining (CRISP-DM). CRISP-DM is the most used in recent years [13]. At the beginning of 1996, the KDD model became the first accepted model in the scientific community that established the main stages of an information exploitation project. Then, from the year 2000, with the great growth that emerged in the area of data mining, two new models were developed that propose a systematic approach to carry out the process: SEMMA and CRISP-DM.

2.4 Reference Model for Data Analytical Project in Non-profit SMEs

CRISP-DM is considered the standard and most referenced methodology to develop data mining and knowledge discovery projects [14]; it is flexible and can be easily adapted to each analytical task in terms of DM processes [15]. CRISP-DM goes into greater detail about the tasks and activities to be carried out in each phase of the data

mining process, while KDD and SEMMA provide only a general guide of the work by each phase.

For this proposal of data analytics for non-profit SMEs, CRISP-DM was taken as reference model for the following reasons: (i) it is the model most referenced by its wide acceptance; (ii) all its phases and activities are properly organized, structured and defined; and finally, (iii) it facilitates the understanding and revision of a project. An adaptation of CRISP-DM was made, excluding some tasks to be simple to implement, improve and replicate; as also flexible to adapt and pay for the non-profit organization, in order to reduce the effort of personnel, time, and costs in the development of DA projects.







3 Methodological Proposal

CRISP-DM/SME’s methodology is proposed as a result of the analysis of data analytics methodologies and the needs of non-profit SMEs. The methodology includes roles of data science, descriptions by phase, activities required in each phase, work products, guidance, and tools. To represent the methodology, a diagram built in SPEM is used since it is a standard language for the modeling of software development processes oriented to work products. The methodological proposal uses the notation represented in Fig. 2 [16].

3.1 Graphic Representation

The methodological proposal is represented in the SPEM diagram, in Fig. 3. The diagram shows that the sequence of the phases is not strict and can interact between

Fig. 2 SPEM elements used

Stereotype	SPEM 2.0
Role Use	
Phase	
Activity	
Work Product Use	
Guide	
Tool Definition	

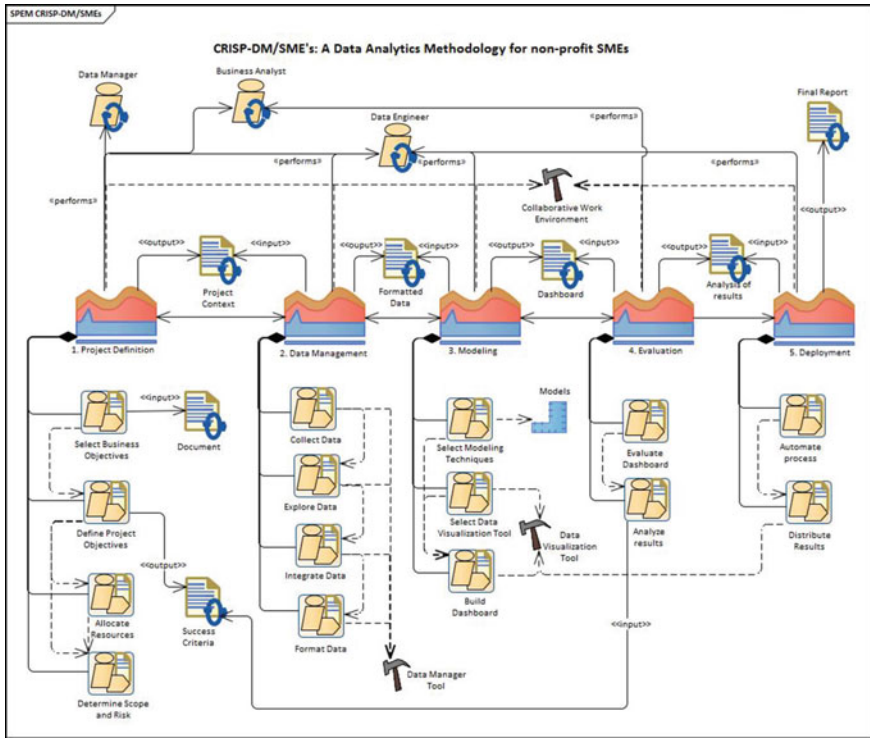


Fig. 3 CRISP-DM/SMEs

each phase. Therefore, project execution can move forward or backward if necessary. The phases described in the diagram contain a set of tasks necessary to guarantee the quality of the project and produce the work products. In addition, technological tools, guides, and models can be used to develop activities. Activities are executed through one or several roles that can be internal or external to the organization.

3.2 Roles

CRISP-DM involves three roles in the data science industry [17], whose profiles are necessary for the development of tasks. Data engineer is the main role within DA projects in the non-profit SMEs. Data manager is responsible of leading the DA team and providing the financial, human, and software resources for the project. Business analyst is responsible for improving business processes and is the intermediary between the data engineer and data manager.

3.3 Phases

Phase 1: Project Definition. Select business goals that will impact the data analytics project. This information is obtained from the strategic plan of the non-profit SME or another similar organizational document. The objectives of the data analytics project are defined and aligned with the business. As a result of this task, the success criteria of the DA project are obtained. Human and financial resources are estimated and allocated for the realization of the project. And finally, the scope and risks are determined.

Phase 2: Data Management. The data is collected from the different sources to be explored. Then, it is integrated with the appropriate format according to the project. The data engineer uses a tool to manage the data and the phase will obtain the formatted data.

Phase 3: Modeling. The selection of the model is closely related to the visualization tool. The tool performs the analysis of the data applying the models chosen for the DA project. The dashboards are built according to the requirements defined in the first phase and contemplated in the success criteria. Samples are taken with a group of small data before proceeding with the evaluation phase, in order to validate the expected results.

Phase 4: Evaluation. The dashboard must be submitted to the stakeholders of the project for their respective assessment and acceptance. The results obtained with the model and the dashboard are evaluated according to the success criteria and the selected business objectives. If the dashboard does not meet the success criteria, then the business analyst will decide if the DA project should go back to an earlier phase to improve the model or the dashboard.

Phase 5: Deployment. This phase consists of automating the data source for the dashboard. A solution is built to integrate, update, and format the data for use in the model and the dashboard. Then, the dashboard is shared with the stakeholders of the project, through protected applications or links.

3.4 Work Product

A work product is the result of each phase of the DA project, which can be used in another phase. The project context is the main document for the other phases of the project, which must contain the business objectives related to the project, the specific objectives of the DA project, the resources allocated, and the scope and possible risks. The criteria for success are also elaborated as a work product. At the end of all tasks in the data management phase, a work product called formatted data is obtained and will be the source of origin for the modeling phase. This work product will have identified the size, fields, attributes, errors, and data types. You will also have the list of the different sources of data for the project. The dashboard will have graphics of descriptive or predictive analysis and it will be the work product that is shared

with all the stakeholders of the DA project in the deployment phase. The analysis of results is the work product that defines if it is necessary to redefine the project, the success criteria or the objectives and if it is necessary to return to an earlier phase or continue with the deployment. Finally, a final report is generated, documenting the effort made in the DA project with the implementation of the CRISP-DM/SME's methodology in order to contribute to the scientific and business community through success cases.

3.5 Tools and Guidance

A collaborative work environment tool is used by the team involved in the development of the project so that each one can control the assigned tasks. On the other hand, the data engineer uses the necessary tools to manage the data and build the dashboard. Each tool is defined in the project context. Additionally, guidance is used to select the models to follow for the data representation.

4 Conclusions and Future Work

The proposed methodology seeks to reduce the effort of implementing data analytics projects in non-profit SMEs, to improve the collection, storage, processing, and analysis of data. It induces the creation of awareness in decision-making based on the exploration of information. Having an analytical methodology allows to minimize complexity, costs, and helps to have trained personnel for the implementation of DA projects. It also facilitates rapid deployment, improvement, and integration into other projects.

The roles integration in the methodology helps assign responsibilities and tasks to each of the stakeholders who participate in the DA project, involving the skills of professionals in the data science industry. Additionally, the use of a dashboard is a decisive work product for data analysis since a dashboard graphically represents the result of the project. Finally, the usage of SPEM language for the representation of the methodological proposal allows an abstract description of the fundamental elements of the process of data analytics and also the description of how they are related to each other.

As future work, we propose the realization of a Web platform that allows the repository of each of the data analytics projects of non-profit SMEs, with the respective phases, tasks assigned to roles, work product, and guides on how to perform each task. This platform should be within the reach of the non-profit SMEs and should contribute to the reduction of effort in the process of development of the DA project.

References

1. A. Palmer, B. Hartley, *The Business and Marketing Environment*, 3rd edn. (McGraw-Hill, London, 2000)
2. E. Soto, La información como recurso estratégico generador de conocimientos, *Soportes audiovisuales e Informaticos* (Universidad de la Laguna, Spain, 2004), pp. 58–60
3. P. Pytel, A. Hossian, P. Britos, R. Garcia-Martinez, Feasibility and effort estimation models for medium and small size information mining projects. *Inf. Syst.* **47**, 1–14 (2015). <https://doi.org/10.1016/j.is.2014.06.004>. (Elsevier, Argentina)
4. R. Mullins, Y. Duan, D. Hamblin, P. Burrell, H. Jin, Z. Ewa, B. Aleksander, A web based intelligent training system for SMEs. *Electron. J. e-Learn.* **5**(1), 39–48 (2007). (ERIC, Germany, Poland, Portugal, Slovakia, United Kingdom)
5. G. Lawton, Users take a close look at visual analytics. *Computer* **42**(2), 19–22 (2009). <https://doi.org/10.1109/mc.2009.61>. (California)
6. H. Florez, Inteligencia de negocios como apoyo a la toma de decisiones en la gerencia. *Rev. Vincul.* **9**(2), 11–23 (2012). (Bogota)
7. T. Guarda, M. Santos, F. Pinto, M. Augusto, C. Silva, Business intelligence as a competitive advantage for SMEs. *Int. J. Trade Econ. Financ.* **4**(4), 187–190 (2013). <https://doi.org/10.7763/ijtef.2013.v4.283>. (Portugal)
8. W. Franco, D. Samiento, G. Serrano, G. Suarez, *Entidades sin animo de lucro* (Consejo Tecnico de la Contaduria Publica, 2015). <http://www.ctcp.gov.co>
9. Conferencia Colombiana de ONG, *Quiénes conforman el sector de las Entidades Sin Ánimo de Lucro ESAL en Colombia* (2016). <http://ccong.org.co/>
10. B. Ogbuokiri, C. Udanor, M. Agu, Implementing bigdata analytics for small and medium enterprise (SME) regional growth. *IOSR J. Comput. Eng. Ver. IV* **17**(6), 2278–2661 (2015). <https://doi.org/10.9790/0661-17643543>. (Nigeria)
11. Sinnetic, *PYMES se desaceleran en transformación digital e innovación por responder a múltiples requerimientos estatales*, vol. 18 (Sinnetic, Colombia, 2017), pp. 1–3
12. A. Gonzalez, *Big data y analítica en Colombia: A un paso de despegar* (2014). <https://searchdatacenter.techtarget.com>
13. KDnuggets, *What Main Methodology are you Using for Your Analytics, Data Mining, or Data Science Projects?* (KDnuggets, 2014) <https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>
14. H. Achmad, V. Sabur, A. Pritasari, H. Reinaldo, Data mining and sharing to create usable knowledge, implementation in small business in Indonesia. *Sains Humanika* **2**, 69–75 (2016). (Indonesia)
15. M. Dittert, R. Härtling, C. Reichstein, C. Bayer, *A Data Analytics Framework for Business in Small and Medium-Sized Organizations*, vol. 73 (Springer, Germany, 2018), pp. 1–13. <https://doi.org/10.1007/978-3-319-59424-8>
16. V. Menéndez, M. Castellanos, Software process engineering metamodel (SPEM). *Rev. Latinoam. Ing. Softw.* **3**(2), 92–100 (2008). <https://doi.org/10.18294/relais.2015.92-100>
17. DataCamp, *The Data Science Industry: Who Does What* (2018). <https://www.datacamp.com/community/tutorials/data-science-industry-infographic>

Bridges Strengthening by Conversion to Tied-Arch Using Monarch Butterfly Optimization



Orlando Gardella, Broderick Crawford, Ricardo Soto, José Lemus-Romani, Gino Astorga and Agustín Salas-Fernández

Abstract The problem existing in bridges that collapse due to undermining of the piers that sustain it, as well as collapses due to hydraulic action, generate high costs that can be reduced, making the repair of its structures. It is possible to reinforce them by modification, incorporating cable-stayed arches, and tensioning of hangers to support them. In this paper, a new approach is presented to solve the problem effectively, using a modern metaheuristic based on nature called monarch butterfly optimization it works imitating the way of migration, who uses the monarch butterflies. The results obtained are compared with those provided by black hole algorithm through the use of a known statistical test.

Keywords Reinforcement of bridges · Metaheuristics · Optimization · Monarch butterfly optimization · Combinatorial optimization

O. Gardella · B. Crawford · R. Soto · J. Lemus-Romani (✉)
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
e-mail: jose.lemus.r@mail.pucv.cl

O. Gardella
e-mail: orlando.gardella.r@mail.pucv.cl

B. Crawford
e-mail: broderick.crawford@pucv.cl

R. Soto
e-mail: ricardo.soto@pucv.cl

G. Astorga
Universidad de Valparaíso, Valparaíso, Chile
e-mail: gino.astorga@uv.cl

A. Salas-Fernández
Universidad Tecnológica de Chile INACAP, Vitacura, Chile
e-mail: jsalASF@inacap.cl

1 Introduction

Currently, the repair of bridges [12] is being studied, constructions that serve to connect two spaces that could not otherwise be accessed. The construction of new bridges may take several years, and a lot of money must be invested. One of the main problems that occur in bridges is the scouring of piers that sustain them as a result of hydraulic action [1, 4]. For this reason and despite the fact that there are other repair techniques, one solution is to modify the structure of the bridge, incorporating a solid steel arch that joins from end to end, and holds the board by means of hangers. The tensioning of hangers must be done sequentially and accurately, since any error can cause damage to the structure or simply its collapse [7]. To solve a problem like this one of great complexity, algorithms inspired by nature have been developed. In this research, monarch butterfly optimization (MBO) will be used to calculate the order and magnitude of the tessellation, which is a metaheuristic population-based algorithm, which is inspired by the migratory behavior of monarch butterflies. For the modeling of the bridge and the problem itself, we will use SAP2000 [11], a software for the analysis and design of structures, which allows through an API to pass information from a metaheuristic technique to the bridge, as well as request from it properties and relevant information of the structure. The API contains predefined functions that can be invoked from a library in different programming languages, allowing a realistic modeling of the bridge. The bridges, meanwhile, are predesigned in the software to separate their design and structuring of the optimization algorithm.

2 Problem

To build a new bridge requires investing high costs of money, also while it is being built, habitual traffic between two ends separated by water is disabled. For this reason, it is appropriate to reinforce the bridges to prolong the useful life of these and to lower the costs resulting from a collapse. The main causes of a bridge collapse are. 1. Scouring, removal of support material caused by water. 2. Erosion, wear produced on the surface produced by the rubbing of water. 3. Corrosion, deterioration generated by oxidation. To reinforce the bridges, a reinforcement was proposed by means of an arch that goes over the bridge deck from end to end. The construction process includes installing the arch to hold the hangers and the net that will support the board, tensioning them sequentially, and eliminating strains. The hangers can not be tightened simultaneously, also the combination between the tensioning order and the applied force must maintain the original bridge forces, that is, the difference in stresses of the modified board must be minimal with respect to the original board. On the other hand, excessive forces of tension in the hangers can cause damage to the board and even the collapse of the bridge. To obtain the efforts of the original board of a bridge without arch or hangers, the SAP2000 software is used. Once the hangers and the arch are installed, the hangers are sequentially tensioned with an adequate

tension (to be found using an optimization algorithm). To evaluate the effectiveness of the tensioning, the efforts of the board with the new configuration are calculated through the same SAP2000 software.

Data to obtain from the original bridge (Eq. 1):

$$\sigma_{o_{i,k}} = \frac{M_{o_{i,k}}}{I_{o_i}}; \quad k \in \{1, 2, \dots, k\}, i \in \{1, 2\} \tag{1}$$

$M_{o_{i,k}}$ = Vector that has the moment of the beam i of the original bridge for each cut k

v_i = Distance from the neutral axis to the fiber furthest from the beam i

I_{o_i} = Inertia of the original bridge, on the beam i

i = Beam

k = Cross-section.

Data to be obtained from the modified bridge (Eq. 2):

$$\sigma_{m_{i,k}} = \frac{P}{A} + \frac{P \cdot e \cdot V_i}{I_{m_{TOTAL}}} + \frac{M_{m_{i,k}} \cdot v_i}{I_{m_i}}; \quad k \in \{1, 2, \dots, k\}, i \in \{1, 2\} \tag{2}$$

$M_{m_{i,k}}$ = Vector that has the moment of the beam i of the modified bridge for each cut k

I_{m_i} = Inertia of the modified bridge, on the beam i

$I_{m_{TOTAL}}$ = Total inertia of the modified bridge beams

e = Eccentricity of the external prestressed strut

A = Total area of the board

P = Total axial effort, including losses.

2.1 Objective Function

Considering that the main objective is to maintain the properties of the beams of the board (that is to say, the efforts in each cut), the objective function is defined as the sum of the difference both higher and lower stresses, for each of the K cuts of each of the two beams, represented in Eq. 3

$$\min \sum_{i=1}^2 \sum_{k=1}^K |\sigma_{o_{i,k}} - \sigma_{m_{i,k}}| \tag{3}$$

where $\sigma_{o_{i,k}}$ is the tension of the original bridge on beam i and on the cut k . In as much, $\sigma_{m_{i,k}}$ is the tension of the modified bridge in the beam i and in the cut k . This function is evaluated for both the lower and higher tensions of the board, and the objective is to minimize the differences.

2.2 Constraints

The constraints for the problem are the following:

- The hangers can not be tightened simultaneously (Eq. 4 and 5).

$$ord_1, ord_2, \dots, ord_n \in \{1, 2, \dots, n\} \quad (4)$$

$$ord_w \neq ord_j ; \forall j, w \text{ con } j \neq w; j, w \in \{1, 2, \dots, n\} \quad (5)$$

- The effort of the modified bridge deck must not exceed the limits of the modified admissible band (Eqs. 6–9)

$$\sigma m \geq \sigma o \quad (6)$$

$$\sigma m \geq f_{ct} \quad (7)$$

$$\sigma m \leq f_{cmax2} \text{ (In intermediate stages)} \quad (8)$$

$$\sigma m \leq f_{cmax} \text{ (In final stage)} \quad (9)$$

σm = Tension (upper or lower) in the modified bridge beams

σo = Tension (upper or lower) in the original bridge beams

f_{ct} = Maximum tensile stress admissible by concrete

f_{cmax2} = Maximum compression stress due to concrete, enlarged

f_{cmax} = Maximum compressive stress admissible by the concrete.

3 Monarch Butterfly Algorithm

In nature, the population of monarch butterflies (MB) of the North American east is known by its migration toward the south during the end of summer/autumn from the north of EE.UU. and southern Canada to Mexico. In MB optimization [10], all individuals of the MB are found in two different locations. Northern USA and the Southern Canada–USA (location one) in addition to Mexico (location two). As a consequence, the locations MB are it behaves the two modes. First, through the migration operator and the migration rate, the position is updated generating the descendants. It is followed by tuning the positions for other butterflies by means of butterfly adjustment. In other words, the migration operator and the butterfly adjustment operator, determined the search direction of the MB in the MBO algorithm. The sum of new butterflies in the two modes it stays constant, being the initial population the one that defines the amount of butterflies, in order to keep the unchanged population and minimize evaluations.

The following rules summarize the migration behavior of MB:

Rule 1. All MB are only found on earth one or earth two. That is, MB on earth one and earth two make up the entire MB population.

Rule 2. Each individual MB child is generated by the MB migration operator on land one or on land two.

Rule 3. For the population to stay constant, and MB will disappear when generating a new child. In the MBO algorithm, the new MB by presenting a better fitness who an old MB, the replacement is done. Next to this, it may be that a new child presents worse performance fitness; in this case, we proceed to discard this solution. When this case is presented, the father remains, without being replaced.

Rule 4. Individuals of the MB that present better fitness are maintained for the next generation, where no operator can modify them. To ensure the quality and effectiveness of the MB population, avoiding deterioration with the increase of iterations.

3.1 Migration Operator

Simplifying the migration process, it can be summarized that MB stay on earth from April 1 to August (5 months) and land September 2 to March (7 months). Therefore, the number of MB on earth one and earth two is $\text{ceil}(\text{NP} \cdot p)$ (NP_1) and $\text{NP} - \text{NP}_1$ (NP_2), respectively. Here, $\text{ceil}(x)$ rounds x to the nearest integer greater than or equal to x ; NP is the total number of the population; p is the proportion of MB on land one. MB on earth one and earth two are called subpopulation one and subpopulation two, respectively. This migration process can be expressed as follows (Eq. 10).

$$x_{i,k}^{t+1} = x_{r_1,k}^t \tag{10}$$

where $x_{i,k}^{t+1}$ indicates the k -th element of x_i at generation $t + 1$ that presents the position of the MB i .

$x_{r_1,k}^t$ indicates the k -th element of x_{r_1} that is the newly generated position of the MB r_1 . t is the current generation number.

MB r_1 is selected randomly from Subpopulation one.

When $r \leq p$, the element k in the newly generated MB is generated by Eq. 10. Here, r can be calculated as (Eq. 11).

$$r = \text{peri} \cdot \text{rand} \tag{11}$$

peri indicates period migration and is set to 1.2 (12 months a year)

rand is a random number drawn from uniform distribution.

When $r > p$, the element in the newly generated MB is generated by Eq. 12.

$$x_{i,k}^{t+1} = x_{r_2,k}^t \tag{12}$$

Algorithm 1: Migration operator

```

1 begin;
2 for  $i = 1$  to  $NP_1$  (for all monarch butterflies in Subpopulation 1) do
3   for  $k = 1$  to  $D$  (all the elements in  $i$ th monarch butterfly) do
4     Randomly generate a number rand by uniform distribution;
5      $r = \text{rand} * \text{peri}$ ;
6     if  $r \leq p$  then
7       Randomly select a monarch butterfly in Subpopulation 1 (say  $r_1$ );
8       Generate the  $k$ th element of the  $x_i^{t+1}$  as Eq. 10.
9     end
10    else
11      Randomly select a monarch butterfly in Subpopulation 2 (say  $r_2$ );
12      Generate the  $k$ th element of the  $x_i^{t+1}$  as Eq. 12.
13    end
14  end
15 end

```

Fig. 1 Algoritmo 1, Migration operator

where $x_{r_2,k}^t$ indicates the k -th element of x_{r_2} that is the newly generated position of the MB r_2 .

MB r_2 is selected randomly from subpopulation two (Fig. 1).

3.2 Adjusting Operator Butterfly

The position of the MB can also be updated by the adjusting operator butterfly. For all the elements in MB j , if a randomly generated number rand is smaller than or equal to p , it can be updated as Eq. 13.

$$x_{j,k}^{t+1} = x_{best,k}^t \quad (13)$$

where $x_{j,k}^{t+1}$ indicates the k -th element of x_j at generation $t + 1$ that presents the position of the MB j .

$x_{best,k}^t$ indicates the k -th element of x_{best} that is the best MB in Land one and Land two.

t is current generation.

if rand is bigger than p , it can be updated as Eq. 14

$$x_{j,k}^{t+1} = x_{r_3,k}^t \quad (14)$$

where $x_{r_3,k}^t$ indicates the k -th element of x_{r_3} that is selected randomly in land two. Here, $r_3 \in 1, 2, \dots, NP_2$. Under this condition, if $\text{rand} > \text{BAR}$, it can be further updated as follows Eq. 15.

$$x_{j,k}^{t+1} = x_{j,k}^t + \alpha x(dx_k - 0.5) \tag{15}$$

Where BAR indicates butterfly adjusting rate. dx is the walk step of the MB j than can be calculated by performing Lévy flight Eq. 16.

$$dx = \text{Levy}(x_j^t) \tag{16}$$

In Eq. 15, α is the weighting factor that is given as Eq. 17.

$$\alpha = S_{\max}/t^2 \tag{17}$$

where S_{\max} is max walk step that a MB individual can move in one step and t is the current generation (Fig. 2).

Algorithm 2: Butterfly adjusting operator

```

1 begin;
2 for  $i = 1$  to  $NP_2$  (for all monarch butterflies in Subpopulation 2) do
3   Calculate the walk step  $dx$  by Eq. 16;
4   Calculate the weighting factor by Eq. 17;
5   for  $k = 1$  to  $D$  (all the elements in  $j$ th monarch butterfly) do
6     Randomly generate a number  $\text{rand}$  by uniform distribution;
7     if  $\text{rand} \leq p$  then
8       | Generate the  $k$ th element of the  $x_j^{t+1}$  as Eq. 13.
9     end
10    else
11      | Randomly select a monarch butterfly in Subpopulation 2 (say  $r_3$ );
12      | Generate the  $k$ th element of the  $x_j^{t+1}$  as Eq. 14.
13      | if  $\text{rand} > \text{BAR}$  then
14        |  $x_{j,k}^{t+1} = x_{j,k}^t + \alpha x(dx_k - 0.5)$ ;
15      | end
16    end
17  end
18 end
19 end

```

Fig. 2 Algoritmo 2, butterfly adjusting operator

4 Implementation

This problem was presented by Valenzuela [9] using genetic algorithm and was compared to black hole algorithm [2, 3] (BH) by Matus [8]. BH obtained better results; in this work, we compare MBO with BH (Table 1).

Order contains natural numbers and represents the order in which the magnitudes of force for each N suspender cable will be applied. Magnitude contains real numbers that represent the percentage of tensile force that will be applied for each N suspender cable (Fig. 3). The constraints will be reviewed by the SAP2000 software.

Table 1 Representation of solution for $N = 3$

3	1	2	0.99	0.87	0.28
Order P1	Order P2	Order P3	Magnitude P1	Magnitude P2	Magnitude P3

Algorithm 3: Monarch Butterfly Optimization algorithm

```

1 begin;
2 Step 1: Initialization. Set the generation counter  $t = 1$ ; initialize the population
  P of NP monarch butterfly individuals randomly; set the maximum generation
  MaxGen, monarch butterfly number  $NP_1$  in Land 1 and monarch butterfly
  number  $NP_2$  in Land 2, max step  $S_{MAX}$ , butterfly adjusting rate BAR,
  migration period  $peri$ , and the migration ratio  $p$ .
3 Step 2: Evaluate each monarch butterfly according to its position.
4 Step 3: while the best solution is not found or  $t < MaxGen$  do
5   | Sort all the monarch butterfly individuals according to their fitness. Divide
   | monarch butterfly individuals into two subpopulations (Land 1 and Land
   | 2);
6   | for  $i = 1$  to  $NP_1$  (for all monarch butterflies in Subpopulation 1) do
7   |   | Generate new Subpopulation 1 according to Algorithm 1.
8   | end
9   | for  $j = 1$  to  $NP_2$  (for all monarch butterflies in Subpopulation 2) do
10  |   | Generate new Subpopulation 2 according to Algorithm 2.
11  | end
12  | Combine the two newly-generated subpopulations into one whole
   | population;
13  | Evaluate the population according to the newly updated positions;
14  |  $t = t+1$ .
15 end
16 Output the best solution.
17 End.
```

Fig. 3 Algoritmo 3, monarch butterfly optimization algorithm

4.1 Parameter Settings Used in Experiments

Parameter configuration was performed using the parametric sweep technique [5]. The instances in MBO were executed with the configurations shown in Table 2.

4.2 Experimental Results

MBO was implemented in Python 3.7 and executed on four servers. A physical server with Intel Xeon processor with 4 cores of 2.33 Ghz, 4 GB of RAM and three VMWare virtual servers with Intel Xeon processor with 2 1.9 Ghz virtual cores, 4 GB of RAM. All of them with Windows 7 operating system.

The problem contains 11 instances. Each of them was executed in 15 experiments. The following tables show the comparisons of results obtained in MBO and BH by Matus [8]. Where BH corresponds to a black hole algorithm, MBO a monarch butterfly optimization.

Table 3 shows the minimum fitness values of MBO and BH and the difference between each of them.

Table 2 MBO settings

Population	p	BAR	Smax	peri	D	Iterations
6	5/12	5/12	1	1.2	3	1000

Table 3 Fitness differences

Instance	Min. MBO	Min. BH	Difference
AB-TC	527,402,607	517,068,779	10,333,827
CC-TC	525,414,720	515,608,056	9,806,664
CR-AA1	518,490,642	511,024,809	7,465,833
HW-TC	526,378,552	516,990,684	9,387,868
PT-TC	529,781,444	517,173,901	12,607,543
PV-TC	530,093,584	517,407,410	12,686,174
RC-AA1	518,146,374	510,072,744	8,073,630
RD-AA1	521,011,555	508,556,024	12,455,531
TC-TC	535,136,905	519,571,134	15,565,771
VC-TC	526,331,613	515,963,267	10,368,345
WR-TC	526,673,501	518,685,865	7,987,636

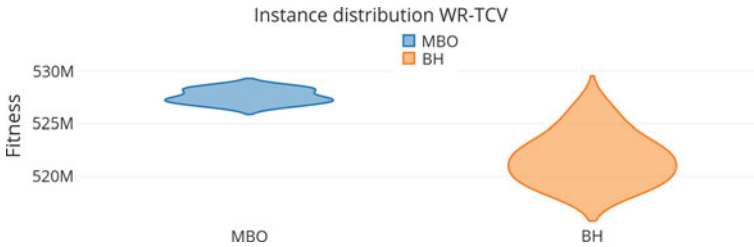


Fig. 4 Instance distribution

4.3 Instance Distribution

The distribution of the data obtained by the two algorithms is compared through violin plot (Fig. 4). All the plots show that the MBO is stuck in the local optimum unlike BH.

5 Comparison Between

To show the differences between MBO and BH, it is necessary to perform statistical tests. Tests were carried out with Mann–Whitney–Wilcoxon [6] because the distribution of the data can not be defined a priori and also it is independent samples. Below are the results for each of the instances (Table 4).

For all instances, the same results are presented, MBO is better than BH with an error rate of 99%, which indicates that MBO has worse results than BH.

Table 4 p-value Mann–Whitney–Wilcoxon test

Instance	2 better than 1	1 better than 2
AB-TCV	1.5296211e-006	0.99999847
CC-TCV	1.53348888e-006	0.999998467
CR-AA10	1.39924268e-006	0.999998601
HW-TCV	1.53348888e-006	0.999998467
PT-TCV	1.5296211e-006	0.99999847
PV-TCV	1.48760546e-006	0.999998512
RC-AA10	1.40283555e-006	0.999998597
RD-AA10	1.53348888e-006	0.999998467
TC-TCV	1.53348888e-006	0.999998467
VC-TCV	1.87011077e-006	0.99999813
WR-TCV	1.53348888e-006	0.999998467

6 Conclusion

The new algorithm was used to solve the problem of bridge reinforcement through cable-stayed arch. As can be seen, the results obtained indicate that the MBO algorithm is not better than BH to solve this problem, according to what is corroborated by the statistical analysis carried out. In the future, we can work on some modifications that allow us to get out of stagnation, which could be the use of automatic configuration of parameters through tools external to the metaheuristics, or implementing autonomous search in the algorithm, so that one or more parameters can be regulated according to the problem.

Acknowledgements Broderick Crawford is supported by Grant CONICYT / FONDECYT / REGULAR / 1171243, Ricardo Soto is supported by Grant CONICYT / FONDECYT / REGULAR / 1190129. This work was funded by the CONICYT PFCHA / DOCTORADO NATIONAL SCHOLARSHIPS / 2019 - 21191692.

References

1. G.F. de Medeiros, M. Kripka, Optimization of reinforced concrete columns according to different environmental impact assessment parameters. *Eng. Struct.* **59**, 185–194 (2014)
2. J. García, B. Crawford, R. Soto, P. García, A multi dynamic binary black hole algorithm applied to set covering problem, in *International Conference on Harmony Search Algorithm* (Springer, Heidelberg, 2017), p. 42–51
3. A. Hatamlou, Black hole: a new heuristic optimization approach for data clustering. *Inf. Sci.* **222**, 175–184 (2013)
4. J.H. Kim, S.H. Kim, Y.K. Kwak, Development and optimization of 3-d bridge-type hinge mechanisms. *Sens. Actuators A: Phys.* **116**(3), 530–538 (2004)
5. F. Lobo, C. F. Lima, Z. Michalewicz, *Parameter Setting in Evolutionary Algorithms*, vol. 54 (Springer Science & Business Media, 2007)
6. P.E. McKnight, J. Najab, Mann-Whitney u test, in *The Corsini Encyclopedia of Psychology* (2010), p. 1–1
7. E. Muñoz, E. Valbuena, Los problemas de la socavación en los puentes de colombia, in *Revista de Infraestructura Vial, numeral* (2006), p. 15
8. B.C. Sebastián Matus de la Parra, Ricardo Soto. Optimización del refuerzo de puentes mediante arco atirantado con black hole algorithm (2018)
9. M. Valenzuela, Refuerzo de puentes de luces medias por conversión en arco atirantado tipo network (2015)
10. G.-G. Wang, S. Deb, X. Zhao, Z. Cui, A new monarch butterfly optimization with an improved crossover operator. *Oper. Res.* **18**(3), 731–755 (2018)
11. E.L. Wilson, A. Habibullah, *Sap 2000, Structural Analysis Program* (Computers and Structures Inc., University Avenue, Berkeley, California, USA, 1995)
12. F. Yong-li, On reinforcement methods of bridge. *Shanxi Archit.* **2**, 203 (2008)

Deep Learning Approach for IDS



Using DNN for Network Anomaly Detection

Zhiqiang Liu, Mohi-Ud-Din Ghulam, Ye Zhu, Xuanlin Yan, Lifang Wang,
Zejun Jiang and Jianchao Luo

Abstract With the astonishing development of the Internet and its applications in the last decade, cyberattacks are changing quickly, and the necessity of protection for communication network has improved tremendously. As the primary defense, the intrusion detection system plays a crucial role in making sure the network security. Key to intrusion detection system is actually to determine a variety of attacks effectively as well as to adjust to a constantly changing threat scenario. DNN or Deep Neural Network on NSL-KDD dataset for effective detection of an attack. Firstly, the dataset was preprocessed and normalized and then fed to the DNN algorithm to create a model. For testing purpose, entire dataset of NSL-KDD was used. Finally, to analyze the accuracy and precision of the DNN model, we use accuracy and precision matrices. The proposed DNN-based strategy enhances network anomaly detection and opens new analysis gateway for intrusion detection systems.

Keywords Deep learning · DNN · Intrusion detection system · Network security

1 Introduction

Together with the progressively in-depth amalgamation of social life and the Internet, the Web is changing how people learn and work, though additionally, it exposes us to progressively powerful security threats. Cybersecurity is a pair of processes and technologies created to safeguard computers, networks, data, and programs from unauthorized access and attacks, modification, and obliteration. A substantial research milestone in the information security area is the intrusion detection system. It can quickly determine an intrusion, which may be a continuing intrusion or may be an intrusion which has currently transpired. Involving the critical difficulties in cybersecurity will be the provision of an effective and robust intrusion detection system. The way you can identify many network attacks, mainly not earlier seen attack types, is a crucial issue being resolved urgently.

Z. Liu · M.-U.-D. Ghulam (✉) · Y. Zhu · X. Yan · L. Wang · Z. Jiang · J. Luo
Northwestern Polytechnical University, Xi'an, People's Republic of China
e-mail: gmohiudin@nwpu.edu.cn

Normal system behavior and network traffic, to analyze the network behavior, are studied by anomaly-based methods. They discover the anomaly or attack whenever the system or network behavior deviates from the standard or general behavior. Anomaly-based methods are more used because of their adaptation capabilities to zero-day or new attacks. Another advantage of using anomaly-based methods is that profile for normal behavior of a system or network is different for each application, protocol, thus making it hard for the attacker. Furthermore, the information on which attack alert is triggered is usually to identify the misuse. The main downside of using anomaly-based technique is the higher rate of false identification of attack or known as a false alert. The false alert is when it is normal traffic but identified as an attack. The reason behind this is before unseen activity which categorized as attack or anomalies. Traditional and less effective signature-based anomaly detection methods are used regardless of advancement in the IDS technology. There is much reason behind this, associated cost as well as high false error rate, lack of reliable training and testing network traffic data and sustainability of the behavior system, are one of the many factors for this averseness to transition. Reliance on these kinds of techniques, in the current situation, leads to ineffective and inaccurate detection. The particulars of this particular challenge are making a commonly recognized anomaly detection method capable of overcoming limitations brought on by the continuing changes happening in contemporary networks.

Machine learning methodologies [1] happen to be popular in identifying different kinds of attacks along with a machine learning strategy can assist the system administrator to go up the equivalent procedures for stopping intrusions. Nevertheless, because of feature engineering and brief learning, traditional machine learning cannot handle the anomaly or attack problem which occurs in a real network environment. Combined with the vibrant advancement of information sets, many classification jobs can result in reduced accuracy. Besides, for high-dimensional learning along with considerable learning, brief learning is not suitable for efficient evaluation and prediction. On the other hand, deep learning overcomes this problem, and it can get far better information from the given data and thus can create a far better model.

ML is a subsequent part of AI and is related to computational statistics that is also focused on predicting using machines/computers. Since mathematical optimization provides essential techniques and principals to the domain industry, it has a reliable connection with machine learning (ML). ML is often time conflated with the data mining which is more about data analysis and is known as unsupervised learning. Behavioral profiles for different entities can be created using ML, and by utilizing those behavioral profiles, anomalies can be identified. ML “is a field of study which provides computer the capability to find out without being programmed,” as defined by Arthur Samuel. Although the second subfield concentrates, ML is at times conflated with data mining more on exploratory data analysis, and it is recognized as unsupervised learning.

DL is a brand-new area in machine learning studies. The inspiration of it is based on the establishment of neural networking which simulates the human brain for analytical learning. The human mind mechanism to understand information like pictures sounds, as well as texts, is mimicked by it.

The idea of deep learning, which offers the desire for solving the optimization issue of the deep structure, was proposed by Hinton in 2006. Adhering to this particular, several related theoretical and functioning investigation research papers published with outstanding achievements [2] that centered on speech recognition, action recognition, and image recognition. The simple fact that deep learning theory, as well as technological innovation, has had a swift advancement in the past few years means that a brand-new era of artificial intelligence has opened and offered an entirely new method to build smart intrusion detection technology.

An artificial intelligence-based anomaly detection system, developed by using a DNN or deep neural networks, is proposed in this paper. The advanced version of the KDDCup99 dataset, NSL-KDD dataset, is used in this study. Performance metrics like accuracy, recall, f-measure, and the false alarm rate are used to calculate the efficiency and effectiveness of the purposed system.

2 Related Works

2.1 Intrusion Detection Researches

In the continually expanding world of network, network and information security become more and more dependent on intrusion detection systems. To maximize the anomaly detection mechanism and increase its accuracy and precision, supervised/semi-supervised and unsupervised machine learning approaches are in use. KNN [3], naïve Bayes, decision trees, support vector machines [4], and artificial neural network [2] are few of the examples.

Liskov et al. [5] provided a comparative assessment of supervised as well as unsupervised learning approaches about the detection accuracy of theirs and ability to identify unidentified attacks. Martinez-Balleste and Solanas [6] offered clustering algorithms for anomaly detection. Bhattacharyya et al. [7] compared the effectiveness of the NSL-KDD dataset on distinct classification algorithms such as naïve Bayes, support vector machines, as well as decision trees [8].

An SVM-based intrusion detection model on NSL-KDD dataset was presented by Wang et al. [9]. They claim to achieve the 99.92% accuracy. Though, critical details about the NSL-KDD dataset, e.g., training/testing samples and dataset statistics. Also, in the case of large data size, the performance of SVM [4] declines. Since during the network analysis, network traffic data is enormous, so SVM [4] is not an appropriate choice.

2.2 DNN

ANN [2] is considered to be the subsequent part of the ML or machine learning. Professors McCulloch and Pitts [10, 11] presented the first research paper regarding the neural network with the title “A Logical Calculus of the Ideas Immanent in Nervous Activity.” That paper explained the behavior of the human neural network and presented a concept of ANN, the first time in history.

With the availability of ReLU, BP or back-propagation, and dropout, trend for DNN is developing very fast. A deep neural network consists of input, output, and hidden layers. Each layer has multiple nodes, and all the computation is performed on these nodes, basically simulating the working of neurons in the human neural network. By multiplying the weight with the input value, we get the magnitude of the reaction whenever a node shows stimulus of a particular degree or above. Since a node has many inputs, each input carries different weights so that these weights can be adjusted among different inputs. Sum of all multiplied values is fed to activation function, and output is used for regression or classification analysis. Use of DNN [12] is extensive in many fields including image recognition, prediction, and learning systems.

2.3 Dataset

Analysis showed that KDD Cup 99 dataset carries the statistical issues which lead to poor approximation and estimation. Those issues were addressed in NSL-KDD dataset. Some files are available to download, for further studies and research. The table shows the detail about each file. NSL-KDD dataset enhances some of the shortcomings of the KDD Cup 99 dataset. The dataset consists of train and test dataset named as KDDTrain+ and KDDTest+, respectively. The test dataset has a distinct number of typical network traffic records, and it covers four significant attack classes as shown in Table 1.

1. KDD Cup 99 dataset have the problem of containing redundant records. All such records are removed in NSL-KDD dataset to make its classifier able to produce unbiased results.

Table 1 Attack classification

Attack class	Attack type
DOS	Back, Land, Neptune, Pod, Smurf, Teardrop, Apache2, Udpstorm, Processtable, Worm
PROBE	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint
R2L	Guess_Password, Ftp_write, Imap, Phf, Multihop, Waremaster, Warezclient, Spy, Xlock, Xsnoop, Snmguess, Snmgetattack, Httpunnel, Sendmail, Named
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps

2. Enough number of records are available to train and test the model.

The specific types of attacks classified into four major categories. Table 1 shows the detail.

3 DNN-Based Network-IDS

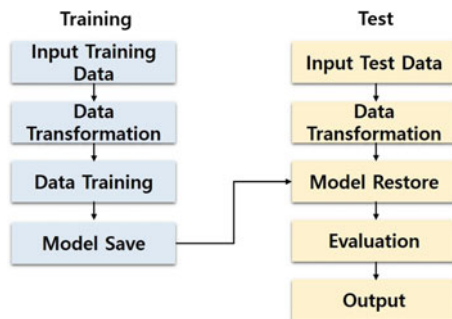
The proposed intrusion detection system (IDS) is illustrated in Fig. 1.

3.1 Preprocessing and Normalization of Data

NSL-KDD dataset enhances some of the shortcomings of the KDD Cup 99 dataset. The dataset consists of train and test datasets named as KDDTrain+ and KDDTest+, respectively. The test dataset has a distinct number of typical network traffic records—the neural network based on numeric values. We cannot use NSL-KDD dataset directly because of the presence of non-numeric features in the dataset. To overcome this, non-numeric features are converted using 1 – *n* numeric encoding.

1. One of the three non-numeric features of NSL-KDD dataset is “protocol-type” feature. Protocol-type feature has three distinct attributes, and these three distinct attributes encoded in binary vectors as (1, 0, 0), (0, 1, 0), (0, 0, 1).
2. Rest of the two non-numeric features in NSL-KDD dataset are “service” and “flag.”. Flag feature consists of 11 unique attributes, and service feature consists of 70 unique attributes. Just like the “protocol” feature, “service” and “flag” will also be encoded into numeric values.

Fig. 1 General DNN-IDS workflow



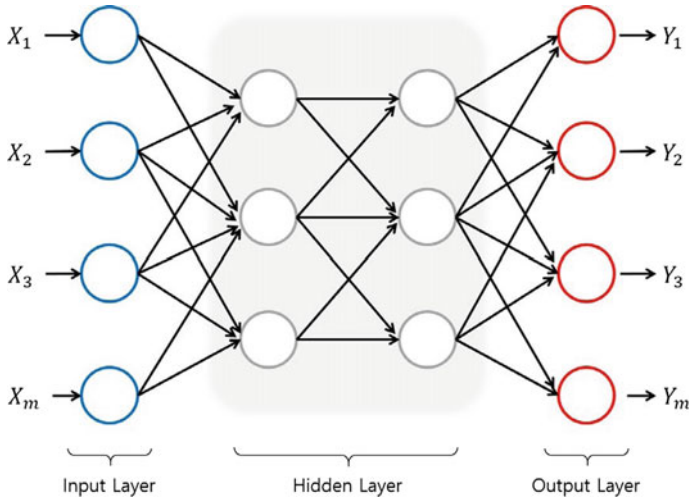


Fig. 2 Classic DNN (deep neural network)

3.2 *DNN Model*

A classic DNN model is showed in Fig. 2. For this particular study, we use a deep neural network with 200 hidden layers.

3.3 *Activate Function*

We use ReLU as the activation function. ReLU can improve the performance since it has the complicated classification better which makes it unique and better performing then linear activation function.

3.4 *Back-Propagation*

The stochastic optimization [13] method is used for back-propagation [14] often known as an optimization.

4 Experimental Results

4.1 Experimental Environments

Due to the complex calculation of DNN, we use GPU to spend less time building the model. From some GPUs available for deep learning, Intel i7 with 32 GB RAM and NVIDIA GTX used for training and testing of the model.

4.2 Evaluation Measure

Almost all performance metrics are used to evaluate the overall performance of the proposed approach of ours. The attribute values that resulted from the training in addition to being testing processes of the NSL-KDD dataset are utilized to calculate these general performance metrics. That classification (or prediction) result divided into four classes:

- True positive (TP): correct positive classification, i.e., identified anomaly occurrence correctly, as an anomaly.
- False positive (FP): Incorrect positive classification, i.e., identified regular occurrence wrongly, as an anomaly.
- True negative (TN): Correct negative classification, i.e., identified normal occurrence correctly as normal.
- False negative (FN): Incorrect negative classification, i.e., identified anomaly occurrence wrongly as normal.

Performance metrics are calculated from the following:

- Accuracy (AC): AC indicates the total percentage of correct predictions (true positive or true negative), which is obtained by Eq. 1

$$\text{Accuracy (AC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

Precision (P): p signifies the percentage of accurate predictions; It is obtained by dividing the total correct predictions with the total number of true and false predictions demonstrated in Eq. 2

$$p = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

Recall (R): Recall (R) signifies the correct percentage of accurate predictions of attack which we get by dividing it by the total number of attacks or intrusions, as shown in Eq. 3

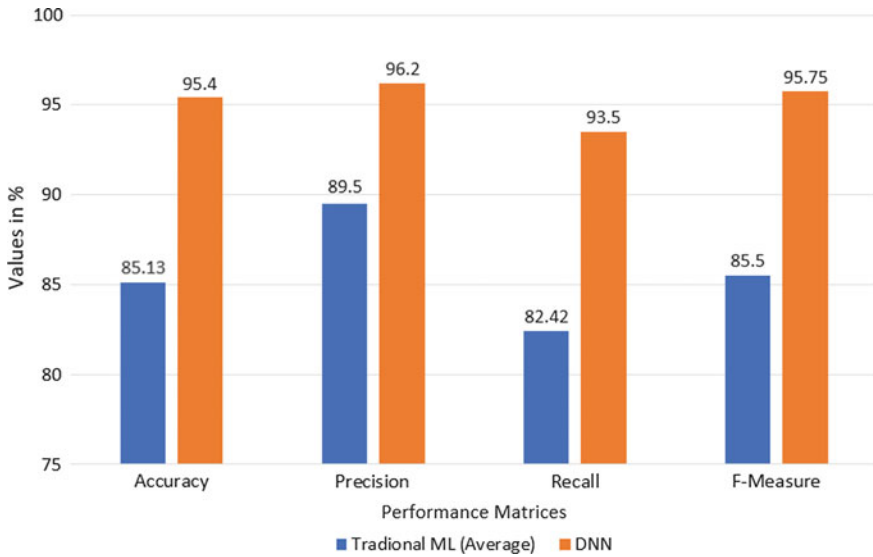


Fig. 3 Performance comparison of traditional machine learning models and DNN model

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

F-measure (F): It is thought to be the most crucial statistic of network intrusion detection which presents each precision (P) and also recall (R), as revealed in Eq. 4

$$\text{F-measure} = \frac{2 * P * R}{P + R} \quad (4)$$

Figure 3 shows the performance of DNN in all are four metrics. Average accuracy of DNN is approximately 93% which is pretty good as compared to the traditional machine learning algorithms.

5 Conclusion and Future Work

The proposed approach in this paper is another mean of detecting network anomalies in network traffic. Utilizing the Deep Neural Network (DNN) model improve not only the efficiency and accuracy of the detection but also the DNN can handle large dataset size. An improved version of KDDCUP99 dataset, NSL-KDD dataset, is used for both training and testing purposes. Not only drastically high precision and detection rate, averaging 97% but also false alarm rate decreased. Studies significantly to date have analyzed and classified only traffic data. Future work will be focused on developing an RNN-based model to overcome the DDoS attack type.

References

1. S. Chung, K. Kim, A heuristic approach to enhance the performance of intrusion detection system using machine learning algorithms, in *Proceedings of the Korea Institutes of Information Security and Cryptology Conference (CISC-W'15)* (2015)
2. N. Gao, L. Gao, Q. Gao, H. Wang, An intrusion detection model based on deep belief networks, in *2014 Second International Conference on Advanced Cloud and Big Data (CBD)* (2014), pp. 247–252
3. D. Shin, K. Choi, S. Chune, H. Choi, Malicious traffic detection using K-means. *J. Korean Inst. Commun. Inf. Sci.* **41**(2), 277–284 (2016)
4. S. Jo, H. Sung, B. Ahn, A comparative study on the performance of SVM and an artificial neural network in intrusion detection. *J. Korea Acad.-Ind. Cooperation Soc.* **17**(2), 703–711 (2016)
5. P. Laskov, P. Dssel, C. Schfer, K. Rieck, Learning intrusion detection: supervised or unsupervised? in *Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP), Cagliari, Italy*, ed. by F. Roli, S. Vitulano (Springer, Berlin, 2005), pp. 50–57
6. A. Solanas, A. Martinez-Balleste, *Advances in Artificial Intelligence for Privacy Protection and Security (Intelligent Information Systems)* (World Scientific, Hackensack, NJ, 2010) (Online)
7. D.K. Bhattacharyya, J.K. Kalita, *Network Anomaly Detection: A Machine Learning Perspective* (CRC Press, Boca Raton, FL, 2013)
8. M. Tahir, W. Hassan, A. Md Said, N. Zakaria, N. Katuk, N. Kabir, M. Omar, O. Ghazali, N. Yahya, Hybrid machine learning technique for intrusion detection system, in *5th International Conference on Computing and Informatics (ICOI)* (2015)
9. W. Wang et al., HAST-IDS: learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. *IEEE Access* **6**, 1792–1806 (2018)
10. W. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**(4), 115–133 (1943)
11. W. McCulloch, W. Pitts, Results of the KDD'99 classifier learning. *ACM SIGKDD Explor. Newsl.* **1**(2), 63–64 (2000)
12. O. Al-Jarrah, A. Arafat, Network intrusion detection system using neural network classification of attack behavior. *J. Adv. Inf. Technol.* **6**(1) (2015)
13. G. Dahl, T. Sainath, G. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 8609–8613
14. D. Kingma, J. Ba Adam, *A Method for Stochastic Optimization*, arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)

Outlier Detection Method-Based KPCA for Water Pipeline in Wireless Sensor Networks



Mohammed Aseeri, Oussama Ghorbel, Hamoud Alshammari,
Ahmed Alabdullah and Mohamed Abid

Abstract Water is considered as the most important resource in our life. Pipelines are considered as very important solution to transport water. So, due to the existence of the harsh environmental condition, different detection ways are not great to monitor pipelines. Therefore, all used systems need to be improved to become more efficient. For this reason, wireless sensor networks (WSNs) are used in water pipeline field. This latter are employed to solve different problems. In this paper, the low-cost damage detection technique for outlier is provided to discuss the task amounts. Our proposed solution uses kernel principal component analysis (KPCA). We aim at analyzing the nature of information. Determine if it is normal or abnormal to help to identify specific events in WSN field for water pipeline. Using real data collected from different stations in WSNs, this solution shows a higher performance in finding abnormal data.

Keywords Outlier detection method · Wireless sensor networks · Data classification · Kernel principal component analysis · Feature extraction

M. Aseeri · A. Alabdullah
King Abdulaziz City for Science and Technology (KACST), Riyadh, Kingdom of Saudi Arabia
e-mail: masseri@kacst.edu.sa

A. Alabdullah
e-mail: aalabdullah@kacst.edu.sa

O. Ghorbel (✉) · H. Alshammari
Jouf University, Al-Jouf, Kingdom of Saudi Arabia
e-mail: oaghorbel@ju.edu.sa

H. Alshammari
e-mail: hhalshammari@ju.edu.sa

O. Ghorbel · M. Abid
National Engineers School of Sfax, CES Research Unit, Sfax University, Sfax, Tunisia
e-mail: mohamed.abid@enis.rnu.tn

Digital Research Center of Sfax, Sakiet Ezzit, B.P. 275, 3021 Sfax, Tunisia

1 Introduction

Outlier detection has gained recently a very important role in water network to detect damage. Based on global sources for water loss around the world, it tells that various countries have a loss range of 20–30%. An economic loss is produced by undetected leak. It increases the charge of water supply networks. Then, systems of monitoring and diagnostic water are developed to be used for damage detection as soon as possible [1]. Damaged pipeline causes huge economic and raw material losses as shown in Fig. 1. These techniques increase the on-line system of water distribution. It gives the possibility of detecting events that deviate from the expected values. Also, it can be associated with leakage, bursts, or water quality contamination [2].

In the literature, numerous multivariate statistical techniques for outlier detection systems are developed like Fisher discriminant analysis (FDA), principal component analysis (PCA), etc. [3]. In comparison, the PCA approach generally represents high-dimensional process data in a reduced dimension via reconstruction. To detect outlier, we use wireless sensor networks (WSNs). This latter contain a big number of low-cost sensors forming an ad hoc network [4]. Most applications in WSNs need reliable data to give to the end user pure information [5].

Chosen as a best solution, outlier detection methods provide good gainful information and allow elaboration of obtained data [6]. The use of kernel principal component analysis (KPCA) is considered as a new field in wireless sensor networks (WSNs). Compared to oldest data collection methods, WSNs can provide continuous measurements of physical phenomena by sensor nodes.

KPCA used by outlier detection solution in WSNs considered as the main contribution of our work. So, based on real data from CRNS and SONEDE, the model is tested and results are interesting in terms of high detection rate and reduced false alarm.



Fig. 1 Visual inspection of damage water pipeline

The paper is made up as follows. Related work is presented in Sect. 2. Section 3 depicts different categories of outlier in WSNs. Section 4 describes mathematical fundamentals of kernel PCA. Obtained experimental results are displayed in Sects. 5, and 6 concludes the paper.

2 Related Works

Various applications like health, leak detection, military, agriculture, etc., are based on a key parameter. It can help efficiently for water leak in pipeline. For example, it prevents from certain events produced in forest by fire or climate change Based on the work presented by [6], they use solution based on KPCA. This solution detects the outlier and classifies data. Based on several equipments of sensors in wireless sensor networks (WSNs), they help to measure leak of water, temperature, pressure, fertility, etc.

Nowadays, kernel principal component analysis (KPCA) becomes an important algorithm which attracted researchers attentions because it produces a very high practical performance. It transforms data to a high-dimensional space [7] as described in Fig. 2.

Using KPCA, the volume of training data is considered higher compared to PCA. So, the amount of estimated principle components presents also a big number. The KPCA process shows efficiently in nonlinear systems compared with linear PCA process. Principal basic KPCA detail can be shown in [8]. Based on Scholkopf et al. work, KPCA is considered as efficient nonlinear dimension reduction technique of data. It transforms the input data into a higher-dimensional feature space. In feature space, by applying a linear operation, this latter can be applied in input space. Kernel PCA is considered as best dimensionality reduction technique [9]. So, it detects and classifies outlier data.

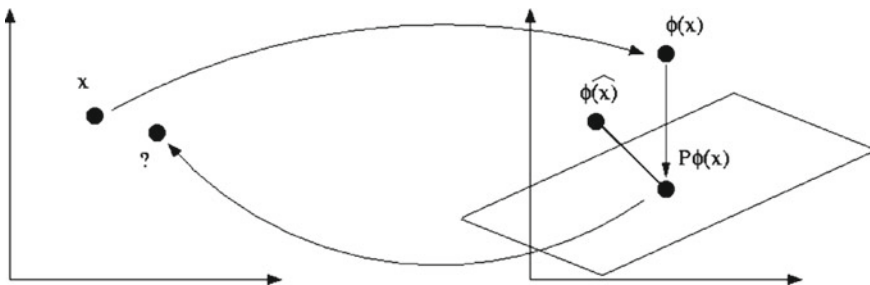
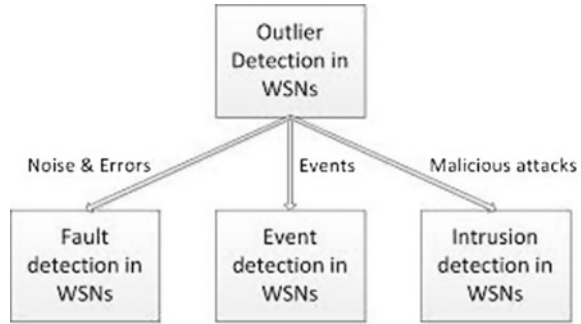


Fig. 2 Representation of KPCA techniques in WSNs

Fig. 3 Different sources of outliers in WSNs



3 Outlier Detection Technique

Outlier considered as specific word for inconsistent values, errors, noise, or duplicate data. System performance was reduced by these abnormal values that affect the nature of data. The data sources of outliers are divided into three types occurred in WSNs as follows: errors, events, and malicious attacks as shown in Fig. 3.

In several real-life applications, using outlier detection technique is important in different fields, such as environmental monitoring, fire monitoring, surveillance monitors, medical monitoring, and precision agriculture [10]. Sensors are considered as low energy and low cost in wireless sensor networks. So to improve robustness, outlier detection method is chosen as the best one [11]. Outlier detection method is evaluated in wireless sensor networks while maintaining to minimize the resource consumptions [12]. To have an important detection rate, we require the use of outlier detection solution while keeping a low false alarm rate. A receiver operating characteristic (ROC) has been used to show the detection rate and false alarm rate. However, many problems can be summarized to detect abnormal data described below: outlier detection application, identifying outlier source, dynamic network topology, etc. [13].

4 KPCA: Kernel Principal Component Analysis

A description of mathematical foundation of kernel principal component analysis (KPCA) is provided in this section. It supports the outlier detection scheme proposed in this work. Kernel PCA is an extension of standard PCA for distributions of nonlinear data [14]. This distribution is assumed that consists of (n) data points $x_i \in \mathbb{R}^d$. So, PCA is projected to (F) named higher-dimensional space.

$$X_i \rightarrow \Phi(X_i) \quad (1)$$

Using previous space, the PCA is applied and this latter has been computed. The vector $\Phi(X_i)$ appears within scalar products [15], and the mapping data has been omitted. So, based on $k(x, y)$, it replaces $(\Phi(x) \cdot \Phi(y))$ named scalar product and gives a good results. Based on KPCA, an eigenvector V of the covariance matrix in (F) is considered a linear combination of points $\Phi(X_i)$.

$$V = \sum_{i=1}^n \alpha_i \bar{\Phi}(X_i) \tag{2}$$

with

$$\tilde{\Phi}(X_i) = \Phi(X_i) - \frac{1}{n} \sum_{r=1}^n \Phi(X_r) \tag{3}$$

Firstly, the vectors $\tilde{\Phi}(X_i)$ are chosen after it centered in the origin of F . Secondly, (α_i) are noted as the components of a vector (α) . Finally, the vector is considered as an eigenvector of the matrix:

$$\tilde{K}_{ij} = \left(\tilde{\Phi}(X_i) \cdot \tilde{\Phi}(X_j) \right).$$

Chosen (α) length by taking into consideration that V (considered as a PC) have the length: $\|V\| = 1 \Leftrightarrow \|\alpha\|^2 = 1/\lambda$, such that λ is the eigenvalue of \tilde{K} wish correspond (α) value. However, $\tilde{\Phi}$ is changed by (3) to compute \tilde{K} that gives \tilde{K}_{ij} demonstrated below:

$$\tilde{K}_{ij} = K_{ij} - \frac{1}{n} \sum_{r=1}^n K_{ir} - \frac{1}{n} \sum_{r=1}^n K_{rj} + \frac{1}{n^2} \sum_{r,s=1}^n K_{rs} \tag{4}$$

5 Experimental Results

To validate the proposed models, an evaluation of the algorithms has been realized to analyze outliers in water pipeline. An experimental campaign has been used to validate the robustness of our solution. Also, a practical experimentation has been evaluated in Digital Research Center Techno-park Sfax (CRNS). The laboratory demonstrator equipment is presented in Fig. 4. It is composed of 25 m pipes which have 32 mm as shown in Fig. 4.

We use varied conditions to test our algorithm. So, damage in water pipeline is identified by data gathered from dynamic and static environments. The pressure data was performed in experimentation, and the database contains 694 samples.



Fig. 4 Testbed water pipeline demonstrator

Table 1 presents the different percentage obtained by KPCA and PCA validated using two databases named CRNS and SONEDE. For the first one, it is the data collected from the Testbed demonstrator realized in the Digital Center of Sfax in Tunisia, and for the second one, it is a national company for water distribution.

In our experimentation, the comparison results realized by kernel principal component analysis demonstrate that this latter is competitive for detecting outliers in water pipeline. For example, obtained value from CRNS dataset is 0.9924 and that obtained from SONEDE dataset is 0.9903 which are almost similar. Based on previous experiments, we conclude that KPCA algorithm gives a very interesting result compared to other methods specially in detecting outliers in pipeline. Because many potential outliers (which is the leak in our case) would not be detected, existing methods are considered unreliable for this kind of detection. So, an advantage offered by KPCA algorithm is that can be applied on a big datasets compared to PCA in detecting outliers as mentioned in Table 2.

Table 1 PCA and KPCA on the real datasets

	CRNS	SONEDE
KPCA	0.9924	0.9903
PCA	0.8588	0.8617

Table 2 False positive rate and detection rate-based KPCA on CRNS dataset

Nodes						
	N1	N2	N3	N4	N5	Average
DR (%)	99	100	99	98	98	99
FPR (%)	1	0	1	2	22	1

6 Conclusion

A useful solution is presented in this work that detects outlier based on KPCA in WSNs. After applying the KPCA on the collected data, the system determines the category of this point as a normal or abnormal one. However, locate the sensor that provides the wrong values.

We use in our experiment two real datasets named CRNS and SONEDE. After using KPCA, we conclude that this latter (as shown in Table 1) gives good results. However, KPCA is a perfect algorithm that detects outliers compared to other methods on water pipeline in wireless sensor network domains.

References

1. D. Wachla, P. Przystalka, W. Moczulski, A method of leakage location in water distribution networks using artificial neuro-fuzzy system. *IFAC-Papers OnLine* **48**(21), 1216–1223 (2015)
2. X. Deng, X. Tian, S. Chen, C.J. Harris, Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes. *Chemom. Intell. Lab. Syst.* **162**, 21–34 (2017)
3. Y. Zhang, N. Meratnia, P. Havinga, Outlier detection Techniques for wireless sensor networks: a survey. *IEEE Commun. Surv. Tutorials* **12**, 159–170 (2010)
4. D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1548–1560 (2011)
5. O. Ghorbel, M. Abid, H. Snoussi, Improved KPCA for outlier detection in Wireless Sensor Networks, in *1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (2014), pp. 507–511
6. Chakour et al., *Adaptive Kernel Principal Component Analysis for Nonlinear Time-Varying Processes Monitoring ICEECA2012*
7. H.K. Verma, V.S. Samparathi, Article: outlier detection of data in wireless sensor networks using kernel density estimation. *Int. J. Comput. Appl.*, (Published by Foundation of Computer Science, 2010), pp. 28–32
8. M.A. Rassam, A. Zainal, M.A. Maarof, An adaptive and efficient dimension reduction model for multivariate wireless sensor networks applications. *Appl. Soft Comput.* (2013)
9. K. Kapitanova, S.H. Son, K.D. Kang, Event detection in wireless sensor networks, in *Second International Conference, ADHOCNETS2010*, Victoria, BC, Canada, August 2010
10. Y. Zhang, N. Meratnia, P. J.M. Havinga, Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine. *Ad Hoc Networks*, December 2012
11. T. Nakanishi, A generative wireless sensor network framework for agricultural use, in *Makassar International Conference on Electrical Engineering and Informatics (MICEEI)* (2014), pp. 205–211
12. N. Chitradevi, V. Palanisamy, K. Baskaran, U.B. Nisha, Outlier aware data aggregation in distributed wireless sensor network using robust principal component analysis, in *International Conference on Computing Communication and Networking Technologies* (2010), pp. 1–9
13. Y. Zhang, N.A.S. Hammb, N. Meratnia, A. Steinb, M. Voorta, P.J.M. Havinga, Statistics-based outlier detection for wireless sensor networks. *Int. J. Geogr. Inf. Sci.* **26**(8), 1373–1392 (2012)

14. M. Ding, Z. Tian, H. Xu, Adaptive kernel principal component analysis. *Signal Process*, 1542–1553 (2010)
15. H. Hoffmann, Kernel PCA for novelty detection. *Pattern Recogn.*, 863–874 (2007)
16. T. Naumowicz, R. Freeman, A. Heil, M. Calsyn, E. Hellmich, A. Brandle, T. Guilford, J. Schiller, Autonomous monitoring of vulnerable habitats using a wireless sensor network, in *Proceedings of the Workshop on Real-World Wireless Sensor Networks, REALWSN'08. Glasgow, Scotland* (2008)

Data Extraction and Exploration Tools for Business Intelligence



Mário Cardoso, Tiago Guimarães, Carlos Filipe Portela
and Manuel Filipe Santos

Abstract Business intelligence (BI) has undergone constant changes currently, due to the increasing emergence of new technologies, which are introduced to improve the processes inherent in decision-making in organizations. However, not all users are familiar with the tools of a typical BI system, so there is a heavy reliance on the assistance of information technology (IT) technicians in the area of data extraction and exploitation (DEE), for ad hoc analyses. In this article, we intend to analyze some DEE tools on the market and their applicability to resolve and help these user's issues in their work environment. For this purpose, literature survey of these type of users and their requirements was done; six DEE tools were selected, analyzed, and experimented; a topology was defined to evaluate the DEE tools in order to identify the one that best applies to business data extraction and exploitation from data warehouses and data marts, associated with BI system and responds to the requirements of these users.

Keywords Business intelligence · Data analytics · Data extraction · Data exploration

M. Cardoso · T. Guimarães · C. F. Portela · M. F. Santos (✉)
University of Minho, Guimarães, Portugal
e-mail: mfs@dsi.uminho.pt

M. Cardoso
e-mail: a66349@alunos.uminho.pt

T. Guimarães
e-mail: tsg@dsi.uminho.pt

C. F. Portela
e-mail: cfp@dsi.uminho.pt

1 Introduction

Decision-making process is becoming a data-driven process [9]. In this sense, the analysis and adaptation capabilities of the users, in contrast with the technologies adopted, is a decisive factor for success. Aiming to help these users, to make the high-grade decisions, after a meticulous literature revision, an experimental study was conducted, by identifying some types of these users and classifying their styles of interaction with data extraction and exploration (DEE) tools connected with data repositories and business intelligence systems for ad hoc analysis. Likewise, elaborating a topology for the evaluation and cataloging of DEE tools.

2 Background

Negash [8] points out that one of the fundamental purposes of business intelligence systems is the ability to convert data into useful information and through human analysis into knowledge. In the context of business intelligence, it is important to distinguish data from information, since both are mutually interconnected, but come from different processes and sources. Data are symbols that represent the properties of objects or events [1, 3]. Davenport and Prusak [3, p. 2] also add that in an “organizational context, data is more properly described as structured transaction records”. Accordingly, for Bellinger et al. [2, p. 5], data represents a fact or statement of an event unrelated to other concepts/facts and that the transition from data to information, information to knowledge, and knowledge to wisdom, happens according to their understanding. For this work, the structured data is imperative, since the tools that deal with these, were evaluated in an ad hoc context and in connection with BI systems. So, this type of data is presented as data organized in records with simple data values (categorical, ordinal, and continuous variables) and stored in the database management systems [9]. Examples: numbers, financial transactions, dates, etc.

3 Data Users

3.1 Users Classification

In order to accomplish the goals of this project, it was necessary to identify what type of users there is in a data environment and what are their needs. According to Dyché [4] users can be classified as: Inventors—non-traditional findings; Explorers—are not sure what they are going to find, but they know where to look; and Casual users—rely on information circulated on a regular basis via standard reports. Eckerson [5] presents a distinct classification, grouping users into two categories: Information producers and information consumers. In the context of this work, it is important

Table 1 Casual user classifications

Class	Description	Role	Analytical needs	Layout preferences	Channels
Viewer	View static reports and dashboards	Executives, responsible for sales and employees	Questions anything to the analysts	Supplementary tables and charts, static presentation	Email, PDF documents, mobile devices
Navigator	Navigates and performs operations on the data contained in the reports and dashboards, looking for more detail	Managers who need information about business performance	Operations on the data (<i>drill-down, pivot, ranking, modify, etc.</i>) and requests support from analysts	Complex and dynamic data tables and charts	Web platforms and mobile devices
Explorer	Explores data from the semantic layer of BI systems and elaborates simple reports	Analysts	Ad hoc exploration and elaboration of simple reports	Semantic layer and interfaces of <i>point-and-click</i>	Desktop Computers

Adapted from [6]

to evaluate the needs of casual users, because they are the end users, who depend more on the support of IT technicians, to access, solve, and work the data with the data extraction and exploration tools. Table 1 presents the Eckerson classification [6] strategies to identify the needs and requirements of these users.

3.2 Types of Interaction

The way users interact to analyze, manipulate, and share the data and can vary according to the infrastructure, capabilities, and user group. For the purposes of this work, it is considered the term style, such as the form/intent of user analysis, as in Lauer et al. [7], where only the following styles were considered: self-service analysis and reporting—empower users who do not have specific skills, to explore the data and manipulate the available information; self-service data mashups—data mashups are created by combining data from multiple sources; and scorecarding—a style that describes highly summarized visualizations with key performance indicators (KPIs) for predefined goals, such as a balanced scorecard.

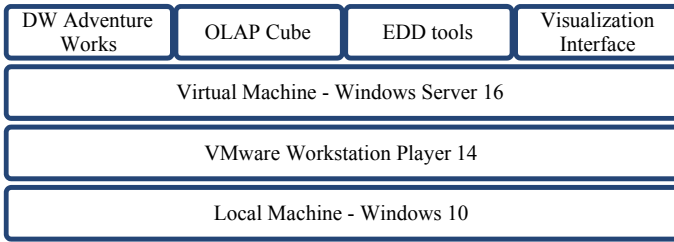


Fig. 1 Architecture of the environment for experimental study

4 Environment for Experiments

For the testing and evaluation of the set of tools selected, it was necessary to have a planned context of data to perform data extraction and exploration tasks. By following the Kimball life cycle, the Adventure Works data warehouse was implemented through Microsoft (MS) SQL Server 2016 and Visual Studio data tools, incorporated in a virtual Windows Server 2016 installed in VMware Workstation Player 14. Figure 1 presents the infrastructure created. The dimensional model of Adventure Works was implemented in MS SQL Server 2017 as the OLAP Cube through Visual Studio—Analysis Service Tools. After this, the data extraction and exploration tools were tested and in visualization interface the MS SharePoint Server 2013 Enterprise Edition was explored in a collaborative role, to present the data extraction and exploration activities elaborated in MS tools.

5 Data Extraction and Exploration Tools

For this study, the most common available tools were considered for evaluation: Microsoft Excel, Microsoft Power View; Microsoft Power Pivot; Microsoft SharePoint; Performance Point; Microsoft Power BI; Tableau. For the evaluation and cataloging of those tools, it is necessary to determine a set of assumptions/characteristics required, so that the requirements identified, in this case, of casual users are fulfilled. These were established accordingly to the styles of interaction with the data, presented at Lauer et al. [7], where the following are highlighted: Self-service analysis and reporting; self-service data mashups; scorecarding.

6 Results

According to the requirements of each tool, a classification based on an ordinal qualitative scale of the interval 0–3 was applied to each tool, according to its correspondence with the requirements: 0—the requirement is not evident or is missing; 1—the requirement exists, but do not responds; 2—the requirement exists and responds fairly; and 3—the requirement exists and responds fully.

Also, each class of casual users has a weight, depending on their ability to use the tools: Viewer (V)—Executives—30%; Navigator (N)—Managers—60%; Explorer (E)—Analysts—10%; All (VNE)—All—100%. Table 2 presents a topology of the DEE considered in the study according to the principal features that are expected to be supported.

6.1 Evaluation

The classification of each tool was evaluated according to the metrics and values obtained. This assessment covers two perspectives:

Individual—taking into account the classes of casual users, awarded to the evaluation of the tools, the topology will allow defining the appropriate tool for each class, consequently for each type of user. The expression 1 presents the individual perspective, where N is the number of requirements/features and X is the scale;

$$\sum_{i=1}^n Xi \tag{1}$$

Global—how each characteristic is associated with the classes of the users and therefore the weight. When assigning a scale to a tool, it is multiplied by the weight (of the user class), i.e., the final score is determined by the sum of the product-scale and weight. Equation 2 transcribes this perspective, where N is the number of features/functionalities, X is the scale and Y is the weight of casual user classes.

$$\sum_{i=1}^n Xi * Y \tag{2}$$

Table 2 Topology

Features	Class	Excel	Excel & PV	Excel & PP	Point	Power BI	Tableau
Allows it to be used without users being familiar with query languages, such as SQL and MDX (multidimensional expressions)	V	3	2	2	2	3	3
Allows analysis of large amounts of data	NE	1	3	3	3	3	3
Allows advanced creation of reports using SQL, MDX, or other query languages	NE	0	0	3	3	2	2
Allows quick creation of reports and data visualizations	NE	3	2	2	2	3	3
Allows custom aggregations and KPIs	NE	1	1	3	3	3	3
Allows you to elaborate reports that provide scores for the performance of an organization, department, or individual	VV	2	3	1	3	3	3
Allows KPIs that target navigation to other report styles in dashboard	NE	1	1	1	2	2	3
Allows access to business, departmental, external, and personal data sources	NE	3	1	3	1	3	3
Allows the use of interfaces drag-and-drop for report design	NE	2	2	2	2	3	3
Allows information to be distributed regularly	VNE	0	0	0	3	3	1
Allows data to be updated automatically	NE	0	2	3	3	4	0
Allows reports to be soaked in other business applications	VNE	3	2	2	2	1	2
Allows reports to be exported in multiple formats with PDF, Excel, Microsoft Word, and HTML	VNE	3	3	0	2	1	1

(continued)

Table 2 (continued)

Features	Class	Excel	Excel & PV	Excel & PP	Point	Power BI	Tableau
Allows users to do activities data-driven (drill-down, drill-up, filtering, and pivot) and present the data in tables, charts, and other visualizations	VN	2	3	3	3	3	3
Allows users to include their reports and visualizations in presentations or share informally with other co-workers	VN	3	1	1	3	3	3
Allows users to work independently from IT technicians	VNE	3	2	1	2	3	3
Allows reports that present the business objectives and KPIs discriminated hierarchically, and can be filtered to assist in the identification of “outliers”	VNE	1	2	3	3	3	2

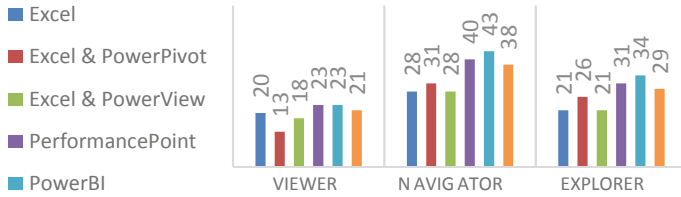


Fig. 2 Individual perspective results

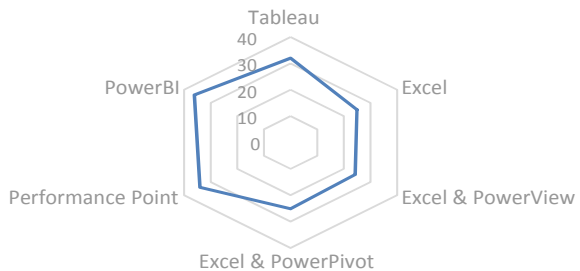
6.2 Individual Perspective Results

After applying the cataloging topology of the tools, we analyzed the scores obtained, in line with the classes of casual users (viewer, navigator, and explorer). This produced results that point the way, to identify the appropriate tool for the requirements of these users. The results obtained according to the classes, presents the tools for each casual user class, and in the first class occur a technical tie between two tools, Performance Point and Power BI, and in relation to the other classes, a clear advantage of Power BI, as depicted in Fig. 2.

7 Global Perspective

Within this perspective is considered the variables that influence the widespread use of tools, that is, a use made by various types of users simultaneously, in order to highlight the most favorable in the global scope. Figure 3 illustrates the results obtained, where a favorable score for the Power BI tool can be verified.

Fig. 3 Global perspective results



8 Conclusions and Future Work

This paper is intended to facilitate the identification and analysis of the data requirements of users. Also, to help planning and implementing data extraction and exploration tools (DEE), finding those that best fits user's needs. A topology for DEE assessment has been introduced. From this topology, an individual and global perspective was proposed in order to evaluate DEE.

Future work will include the integration of the topology in a DEE recommendation component, based on the topology evaluation parameters and scales. This component will be part of the DEM platform in order to help end users selecting the most appropriated DEE.

Acknowledgements This article is a result of the project Deus Ex Machina: NORTE-01-0145-FEDER-000026, supported by Norte Portugal Regional Operational Program (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

References

1. R.L. Ackoff, From data to wisdom. *J. Appl. Syst. Anal.* **16**(1), 3–9 (1989). <http://doi.org/citeulike-article-id:6930744>
2. G. Bellinger, D. Castro, A. Mills, *Data, Information, Knowledge, and Wisdom* (2004), pp. 5–7
3. T.H. Davenport, L. Prusak, Working knowledge—how organizations manage what they know. **21**(8), 395–403 (Harvard Business School Press, Boston Massachusetts, 2000)
4. J. Dyché, *Categorizing Business Intelligence Users*, 4 (2007). Retrieved from <https://searchbusinessanalytics.techtarget.com/news/2240036691/Categorizing-business-intelligence-users>
5. W.W. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*, 2nd edn. (Wiley, Hoboken, 2010). <https://doi.org/10.2514/6.2008-3494>
6. W.W. Eckerson, *Classifying Business Users* (2013). Retrieved February 10, 2018, from http://www.b-eye-network.com/blogs/eckerson/archives/2013/09/classifying_bus.php
7. J. Lauer, S. Cameron, J. Nelson, V. Rocca, How to choose the right reporting tools for your instrument control system. Microsoft (2012). Retrieved from [https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/jj129615\(v=msdn.10\)](https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/jj129615(v=msdn.10))
8. S. Negash, Business intelligence. *Commun. Assoc. Inf. Syst.* **13**, 177–195 (2004). <https://doi.org/10.1002/9781118915240.ch7>
9. E. Turban, R. Sharda, D. Delen, D. King, J.E. Aronson, *Business Intelligence: A Managerial Approach*, 2nd edn. (Prentice Hall, 2010). Retrieved from <https://books.google.com/books?id=IvZORAAACAAJ&pgis=1>

Systems and Methods for Implementing Deterministic Finite Automata (DFA) via a Blockchain



Craig S. Wright

Abstract We present a novel technology for the establishment and (discretionary) automatic execution of (financial) contracts based on the realisation of the commitments of the different parties, and other clauses and provisions, as a non-deterministic finite automaton (NFA) embodied in a computational and record-keeping structure on the (Bitcoin) blockchain. In particular, the process provides methods for constructing non-deterministic finite-state automata in Bitcoin script. The “best” method produces a one-to-one relation between the definition and the state table of the automaton.

Keywords Automata · DFA · Computation · Bitcoin · Blockchain

1 Introduction

The automation of (financial) contracts has been a topic of continued academic research (see, e.g., [1] and references therein) and practical interest since the realisation that an electronic version of the essence of such contracts can be better defined (e.g. avoiding ambiguities and interpretations of current legalese as well as potentially costly and long litigations) and executable and enforceable by computers, and consequently cheaper and more reliable.

Among the different approaches which have been proposed in the literature (see Chap. 1 of [1] for a short review), it has been shown that a deterministic finite automata (DFA), also known as deterministic finite-state machines, have a rich enough structure to represent a wide range (if not all) of imaginable financial agreements, and other kinds of contracts [2, 3].

A DFA is a mathematical model of computation conceived as an abstract machine that can be in one of a finite set of states and can change from one state to another (transition) when a triggering event or condition occurs. Its computational capabilities are more than those of combinational logic but less than those of a stack machine.

C. S. Wright (✉)
CNAM, Paris, France
e-mail: c.wright@nchain.com

Going beyond existing literature, we develop a technological innovation, which facilitates the incarnation of a DFA in a practical way on the existing infrastructure of the (Bitcoin) blockchain. The states of the machine are defined and recorded on the permanent ledger which is the blockchain, and blockchain transactions work as agents changing the machine from one state to another. When the DFA defines a contract, the innovation provides with a mechanism for the automatic execution and enforcement of the commitments of the different parties, and other clauses and provisions.

Key Elements. The blockchain-based DFA has the following key features: legalise and litigation-free by construction; executable and enforceable automatically by computers; and permanent record of contracts, their execution, and outcomes.

And it offers the following benefits inherent to the Bitcoin blockchain: inherently secure by design (the Bitcoin protocol requires no trusted parties); distributed (so avoids a large single point of failure and is not vulnerable to attack); easy to manage and maintain (the Bitcoin network is straightforward to use); inexpensive (just a small transaction fee is usually expected under the Bitcoin protocol); global and can be used at any time by anyone with access to the Internet; transparent—once data has been written to the blockchain, anyone can see it; immutable—once data has been written to the blockchain, no one can change it; privacy is maintained, as no personally identifying information is involved.

Section Organisation. The contribution is broken up into two main sections: a functional specification, offering a high-level outline, and explanation of the matter and nature of the proposed solution; and a technical specification, outlining the technical possibilities and novelties involving the innovation, ending with an example.

2 Functional Specification

At a high level of abstraction, the system we are proposing consists of the DFA itself and a closely related system of agents (Botnet) which write the transactions and submit them to the blockchain. We will concentrate here on the description of the DFA mechanism, and possible realisations of the transactions, leaving the specification of the system of agents for parallel companion work [4].

An illustration of the complete system can be found in Fig. 1. There the Botnet¹ interacts with the world (humans or other computers) to receive instructions, e.g. which contract to create and execute, as indicated by the connection lines. The specification of the contract itself can be provided in any format, e.g. xBRL [5], and stored in a secure and decentralised manner, for example, in a distributed hash table (DHT) on the torrent network. From the specification of the contract, a Botnet agent

¹We will refer generally to the Botnet, often without specifying, which agent actually carry out the actions described, this could be one of the lower-level bots, a bot manager (Botman) or any other appropriate entity as specified in [4].

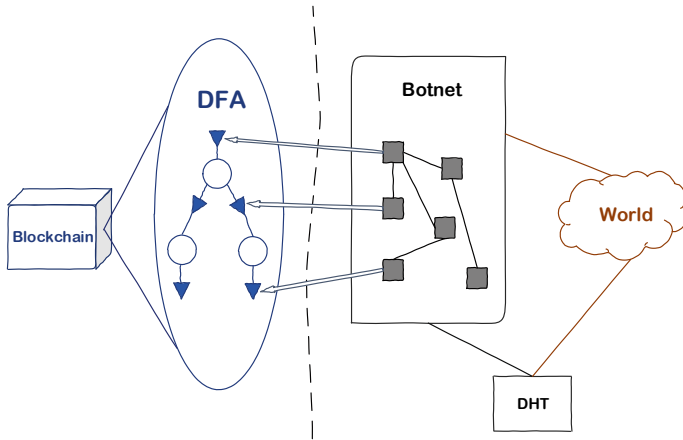


Fig. 1 Diagram describing the blockchain-based DFA and Botnet systems

constructs the DFA, which is subsequently incarnated on the blockchain by Botnet agents.

The DFA itself is specified as a finite set $\{S, I, t, s_0, F\}$, where S stands for the (finite) set of possible states in which the contract/DFA can be; I is a (finite) set of inputs (also known as the alphabet), which in our context means any event or condition which can occur in relation to the contract, e.g. a payment is made, the maturity of the instrument is reached, a counterparty defaults, etc., in our mechanism, these input signals are received/produced by Botnet agents, which then determine the next state of the system (possibly the same one).

The third component of a DFA is a transition function $t: S \times I \rightarrow S$. Deterministic refers to the uniqueness of the decision: given a state and an input there is only one new state (possibly the same one); thus, given an initial state (s_0) and a history of inputs the outcome of the calculation (contract) is unique, one among the set of all possible final outcomes ($F \subseteq S$). Once all these elements have been established, the DFA is completely defined by a transition table, specifying the future states for all possible current states and input signals. The states of the DFA are themselves associated with unspent transaction outputs (UTXO) on the blockchain. Note that the Bitcoin network continuously tracks all available UTXO. The mechanism by which the DFA moves from one state to another is incarnated in our proposal by Bitcoin transactions; effectively they spend the UTXO associated with one state (an input of the transaction) and create the UTXO associated with the next state (an output). This constitutes a key inventive element of our proposal.

To illustrate these ideas, we are going to consider a discount (zero-coupon) bond, which is a simple debt instrument usually bought at a price (normally at a discount with respect to its face value), then held for some time, until its principal is returned at maturity. The possible states we will consider are $S = \{s_0, f_0, f_1\}$, indicating, respectively, the holding state (s_0), the normal conclusion of the contract (if it follows the happy path) or happy ending (f_0), and a state (f_1) in which things go wrong, e.g.

Table 1 Transition matrix for a DFA representing a zero-coupon bond

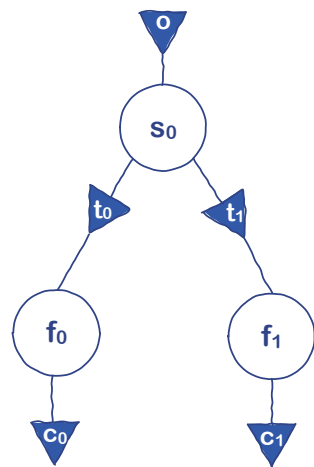
t	r	d	e
s ₀	f ₀	f ₁	f ₁
f ₀	–	–	–
f ₁	–	–	–

litigation; the final states of the system are thus $F = \{f_0, f_1\}$. The alphabet we will consider is $I = \{r, d, e\}$, indicating, respectively, repayment of the principal at (or before) expiration (r), default of the issuer at (or before) expiration (d), and expiration of the contract without repayment (e). The transition matrix for this simple contract is presented in Table 1. Note that for the final states represent the completion of the contract, thus no further states need to be specified from them (currently noted as ‘–’ in the transition table, although those lines could be omitted).

Figure 2 represents the embodiment of the zero-coupon bond DFA on the (Bitcoin) blockchain. The states are represented by circles and the Bitcoin transactions which move the machine from one state to the other by the blue triangles. Note that the inputs received by the Botnet agents are omitted in this diagram, however, in each state one or other transition should occur according to these inputs, which is reflected in the diagram by the construction of one or other Bitcoin transaction (e.g. t₀ or t₁ in state s₀); no transactions are required for transitions which do not change the state, thus they have been omitted. In addition to the transitional transactions of the DFA (t_i), an initial origination transaction (o) and transactions corresponding to the completion of the contract (c_i) are considered.

A last point to discuss before turning to the technical specification of the invention is the flow of funds in the transactions (originations, transitions and completions). An important observation is that because of the finite nature of the DFA, and of (financial) contracts, they would be completed after a number of transitions. This

Fig. 2 Diagram describing a blockchain-based DFA



necessarily implies (assuming some finite fees for the Botnet agents involved and the Bitcoin miners) that the maximum costs of the establishment and execution of the contract is bound and can be determined in advance, e.g. at the point of establishment of the DFA. It is given by the total amount of funds required to execute the contract following the longest imaginable path. This, of course, excludes the possibility of infinite loops in the execution, note, however, that this is not relevant for current (financial) contracts; even contracts such as perpetuities are bound to be completed at some point in future, despite their name.

The particular distribution of the fees, how much each agent receives for their work, although of obvious practical significance is not fundamental to the invention described here. Continuing with our example of a streamlined zero-coupon bond, we will arbitrarily assume a fee of 3 mBitcoin for the origination transaction (o), a transition fee (ti transaction) of 1 mBitcoin and a 2 mBitcoin fee for the completion transaction (ci); note that these fees are automatically included in the transactions themselves. Together with 3 mBitcoin of total mining fees for the 3 transactions, this results in a total maximum cost of 9 mBitcoin. In our example, the length of all paths is equal, thus the final cost of the contract will certainly coincide with its maximum cost. Since this need not be the case in general, in order to completely illustrate a general flow of funds, we will assume that the funds provided for the execution of the contract are 10 mBitcoin, and that 1 mBitcoin is returned to the same source of funds after completion (this takes the place of eventual unused funds at completion, i.e. if maximum and used funds were to differ).

We will assume that the funds for the establishment and execution of the contract (10 mBitcoin) are initially provided by some funding source referred to as the originator (as mentioned, in our example, this source will also receive 1 mBitcoin unused funds after completion). In principle one could also include additional inputs and outputs of funds in the transactions, e.g. the price paid for the zero-coupon bond, the repayment of the principal, or any other imaginable transfer of funds. Although this may be of practical interest, at this point it would only help to obscure the essential elements of blockchain-based DFA processing. For the sake of clarity, we have decided not to include such details in the examples below; note, however, that the structure is completely general and such possibilities are not excluded.

Once the functional specification of the technology has been established, the remaining components of the system ought to be specified at a technical level, in particular, the inner working and the flow of information and funds of the Bitcoin transactions which constitute a fundamental innovation of our proposal.

3 Technical Specification

There are several options here, depending to a large extent on the particular configuration of the Botnet system [4], e.g. whether the particular agents involved are known in advance or not. We will explicitly consider two options here. If participation on the contract is kept fairly open, so that a variety of agents can participate, a structure

based on the standard transaction of the type transaction puzzle [6] is feasible. Conversely, if the transactions for each state can be assigned in advance to a particular agent (or group of agents), a pay-to-script-hash (P2SH) [7] transaction can be used. Of course, one can imagine a number of possibilities in between, e.g. a group of agents which change in time, a hierarchy of agents, private information (keys) could be securely interchanged just before each transaction is written, etc. The possibilities are many and, as mentioned, depend to a large extent on the particular configuration of the Botnet to be discussed elsewhere [4]; we will only specify explicitly the two options mentioned.

Note, however, that which particular type of transaction is used, as specifically who provides/receives the funds in the transactions (the issuer, the purchaser, the payee, etc.), is ultimately not fundamental to the invention described in this paper. The essence of the proposal is that at any point in time the state of the contract is well defined within the blockchain, and that we describe a mechanism (whatever its detailed concrete realisation) which generates the contract, executes it on the blockchain, and enforces the appropriate outcome according to the sequence of events that occurs. Note that, because the whole mechanism is embodied on the blockchain, it automatically provides a permanent immutable record of the history and outcome of the contract, among many other advantages.

3.1 *Transaction-Puzzle-Based*

One of the configurations of the Botnet contemplated in [4] regards the system as an open network of computers in which anyone with some Internet-connected processing power may join, provide some processing power to the system (e.g. a provider of the establishment and execution of the blockchain-based DFA contracts) and get rewarded for their resources (see [4] for details). Thus, we are compelled to assume that it is impossible to know in advance which particular agent will submit the transaction to the blockchain, i.e. it is not possible to use any specific information on the agent (e.g. its public keys); however, a transaction-puzzle type is still feasible. The general locking/unlocking mechanism of this type of transactions is [6]:

Locking Script : OP_HASH256 <state s_i puzzle > OP_EQUAL

Unlocking Script : <puzzle s_i solution >

where <state s_i puzzle> = HASH(<state s_i puzzle solution>) and the puzzle solution itself can include any desired information, including a code for the contract, the label of the state, and any other desired information, for example, an extra bit of *salt* (for added security [8, 9]):

<state s_i solution > = HASH(<contract code; state s_i ; other data; salt >)

The sequence of actions is as follows. First, the Botnet system (the Botman or another agent, as specified in [4]) creates the DFA structure, stores the transition table externally (e.g. in a DHT), create the puzzles for each possible state of the DFA and distributes them securely to the agents which are allowed to participate on the execution of the contract. There are a number of possibilities for this flow of information but, as before, since it is not fundamental to the process, we will omit such details here.

Next, a Botnet agent (the same or a different one [4]), creates the origination transaction as specified in Table 2. At this stage, the contract is incarnated as a structure on the blockchain and is in its first state s_0 , i.e. there is an UTXO associated with the state s_0 of the particular contract on the blockchain.

As can be deduced by studying the transaction, the original funding for the contract is received (from the originator in the example) by a P2PKH-type transaction, *output 0* (puzzle-based) can be unlocked by any Botnet agent in possession of the puzzle solution, and *output 1* (P2PKH) pays the required fee to the Botnet agent which has succeeded in placing the transaction on the blockchain.

Successive transitions *on the execution of the contract* are carried out by Botnet agents in a similar fashion as exemplified in Table 3. They need to get the puzzle

Table 2 Example of an origination (to state s_0) transaction (o). The funds allocated to the contract have been assumed to be 10 mBitcoin, the Botnet fee 3 mBitcoin, and the outgoing funds for further processing of the contract 6 mBitcoin (1 mBitcoin mining fee is implicit)

Transaction identifier			origination o
Version number			<version number>
Number of inputs			1
Input	Previous transaction	Hash	<10 mBitcoin from originator>
		Output index	00
	Length of signature script		<unlocking script length>
	Signature script		<originator signature> <originator public key>
	Sequence number		<sequence number>
Number of outputs			2
Output 0	Value		600000
	Length of public key script		<locking script length>
	Public key script		OP_HASH256 <state s_0 puzzle> OP_EQUAL
Output 1	Value		300000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG
Locktime			0

Table 3 Example of a transition (from state s_0 to state f_f) transaction (t_f). The incoming funds have been assumed to be 6 mBitcoin, the Botnet fee 1 mBitcoin, and the outgoing funds for further processing of the contract 4 mBitcoin (1 mBitcoin mining fee is implicit)

Transaction identifier		transition t_f	
Version number		<version number>	
Number of inputs		1	
Input	Previous transaction	Hash	<6 mBitcoin from origination o >
		Output index	00
	Length of signature script	<unlocking script length>	
	Signature script	<state s_0 puzzle solution>	
	Sequence number	<sequence number>	
Number of outputs		2	
Output 0	Value	400000	
	Length of public key script	<locking script length>	
	Public key script	OP_HASH256 <state f_f puzzle> OP_EQUAL	
Output 1	Value	100000	
	Length of public key script	<locking script length>	
	Public key script	OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG	
Locktime		0	

solution corresponding to the current state (s_0), interact with the world (or some other computer on the Botnet) in order to receive the appropriate input, read the transition table (or just the part of them corresponding to the current state) and get the puzzle corresponding to the appropriate next state (f_f). They can then submit the transaction to the blockchain, if they succeed in placing it, they will get their fee and the DFA will be in the state f_f .

In order to conclude the technical description of the mechanism, we need to define the structure of the last kind of possible transactions, those completing the execution of the contract (c_f). This is shown in Table 4, where the input part follows the same transaction puzzle logic as before, while *output 0* pays the *unused funds* back to the originator (1 mBitcoin, as discussed above), and *output 1* pays the fee to the Botnet agent (Table 5).

Table 4 Example of a completion (from state f_i) transaction (cf). The incoming funds have been assumed to be 4 mBitcoin, 1 mBitcoin unused funds are returned to the originator, and the Botnet fee is 2 mBitcoin (a 1 mBitcoin mining fee is implicit)

Transaction identifier			completion c_f
Version number			<version number>
Number of inputs			1
Input	Previous transaction	Hash	<4 mBitcoin from transition t_f >
		Output index	00
	Length of signature script		<unlocking script length>
	Signature script		<state f_i puzzle solution>
	Sequence number		<sequence number>
	Number of outputs		
Output 0	Value		100000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <originator public key hash> OP_EQUALVERIFY OP_CHECKSIG
Output 1	Value		200000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG
Locktime			0

3.2 P2SH Based

Instead of being open to a large number of a priori unknown participants, another possible configuration of the Botnet could be that of a limited number of acknowledged computers. In this case, transactions of the type P2SH [7] might be more appropriate since they can be configured to include the public keys of the agents, thereby providing an extra layer of security. In the case of a single acknowledged agent, a feasible locking/unlocking mechanism for the transactions is:

Locking Script: OP_HASH160 <state s_i redeem script hash> OP_EQUAL

Unlocking Script: OP_0 <agent signature> <state s_i redeem script>

Redeem Script: OP_1 <state s_i metadata> <agent public key> OP_2 OP_CHECKMULTISIG

Note that this can be trivially extended to a larger number of acknowledged agents by including their signature as well. As before, the metadata of state a state s_i can include any desired information, for example:

<state s_i metadata> = HASH (<contract code; state s_i ; other data>)

Table 5 Example analogous to that in Table 2 but using the P2SH transaction type for the DFA states

Transaction identifier			origination o
Version number			<version number>
Number of inputs			1
Input	Previous transaction	Hash	<10 mBitcoin from originator>
		Output index	00
	Length of signature script		<unlocking script length>
	Signature script		<originator signature> <originator public key>
	Sequence number		<sequence number>
Number of outputs			2
Output 0	Value		600000
	Length of public key script		<locking script length>
	Public key script		OP_HASH160 <state s_0 redeem script hash> OP_EQUAL
Output 1	Value		300000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG
Locktime			0

The sequence of actions is similar as in the previous case. First, the Botnet system creates the DFA structure, stores the transition table externally (e.g. in a DHT), determines which agents will process the transactions and retrieve/generate their public keys, which are then included in the redeem scripts for each possible state of the DFA, note that these scripts can be stored externally and need not be transmitted securely. Next, a Botnet agent creates the origination transaction as specified in Table 2. At this stage the contract is incarnated as a structure on the blockchain and is in its first state s_0 .

The only change in the transaction with respect to that in Table 2 is *output 0*, which now is of type P2SH and includes the public keys of the acknowledged agents as discussed above. Although the changes are completely analogous, for completeness we present in Tables 6 and 7 examples of transition transactions and completion transactions based on the P2SH transaction type. The flow of funds is the same as before.

Table 6 Example analogous to that in Table 3 but using the P2SH transaction type for the DFA states

Transaction identifier			transition t_f
Version number			<version number>
Number of inputs			1
Input	Previous transaction	Hash	<6 mBitcoin from origination o >
		Output index	00
	Length of signature script		<unlocking script length>
	Signature script		OP_0 <Botnet agent signature> <state s_0 redeem script>
	Sequence number		<sequence number>
	Number of outputs		
Output 0	Value		400000
	Length of public key script		<locking script length>
	Public key script		OP_HASH160 <state f_f redeem script hash> OP_EQUAL
Output 1	Value		100000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG
Locktime			0

3.3 Example

Imagine Alice has some savings and wants to invest them in a discount bond from a certain bond issuer. She can contact a service provider of blockchain-based DFA contracts, provides the appropriate (discounted) price funds (possibly through another Bitcoin transaction which can be integrated in the DFA as well), and have it create and automatically execute the contract. If everything goes well (happy path) she will automatically receive the face value of the bond back at maturity. If, for example, the bond issuer happens to default while Alice holds the bond, the Botnet system will automatically take the appropriate action as defined in the contract.

4 Summary and Conclusion

The invention relates to a technique for implementing, controlling and automating a task or process on a blockchain such as, but not limited to, the Bitcoin blockchain. The invention is particularly suited for, but not limited to, automated execution of

Table 7 Example analogous to that in Table 4 but using the P2SH transaction type for the DFA states

Transaction identifier			completion c_f
Version number			<version number>
Number of inputs			1
Input	Previous transaction	Hash	<4 mBitcoin from transition t_f >
		Output index	00
	Length of signature script		<unlocking script length>
	Signature script		OP_0 <Botnet agent signature> <state f_f redeem script>
	Sequence number		<sequence number>
	Number of outputs		
Output 0	Value		100000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <originator public key hash> OP_EQUALVERIFY OP_CHECKSIG
Output 1	Value		200000
	Length of public key script		<locking script length>
	Public key script		OP_DUP OP_HASH160 <Botnet agent public key hash> OP_EQUALVERIFY OP_CHECKSIG
Locktime			0

contracts such as smart contracts for financial agreements. However, other types of tasks and non-financial contracts can be implemented. The invention can be viewed as the implementation or incarnation of a state machine or DFA on a blockchain by using the unspent outputs of blockchain transactions to represents the states of the machine, and spending of those outputs as the transition of the machine from one state to another. The invention provides a technical realisation and implementation of a mathematical model of computation conceived as an abstract machine that can be in one of a finite set of states and can change from one state to another (transition) when a triggering event of a finite set (called input) occurs. The invention comprises compilation and codification techniques for the DFA implementation.

As detailed, this method is a means to implementing a DFA on a blockchain, comprising the listed steps. It is a method of implementing a DFA on a blockchain, comprising the steps of associating a portion of data in the locking script of an unspent output (UTXO₁) of a blockchain transaction (Tx₁) with a given state of the DFA.

A method further comprising the step of using a further transaction (Tx₂) to make a transition from the state of the DFA to a further state by spending the output (UTXO₁) of the transaction (Tx₁); wherein the further state is associated with a portion of data provided within a locking script of an unspent output (UTXO₂) of the

further transaction. The process is completed using a portion of code to implement or represent at least one state transition trigger which, when executed, causes a further transaction (Tx_2) to spend the output ($UTXO_1$) of the transaction (Tx_1) and thus move the DFA to another state.

The process is a method wherein the portion of code comprises a machine-testable condition which provides a Boolean result based upon an input. In this, the input is determined at run time and is used by the portion of code to determine whether or not the unspent output ($UTXO_1$) should be spent so as to move the DFA to the further state such that the portion of data in the unspent output ($UTXO_1$) is provided in a locking script and data is a tag, label or a portion of metadata. In the model, the DFA is a model of a machine-executable smart contract.

This is conducted successfully as the unspent output ($UTXO_1$) comprises a locking script which includes a hash of a puzzle, the solution of which must be provided by an input of a further transaction in order to spend the output ($UTXO_1$) and transition the DFA to another state.

In the system, the unspent output ($UTXO_1$) comprises a locking script which includes a hash of a redeem script which must be provided by an input of a further transaction in order to spend the output ($UTXO_1$) and transition the DFA to another state. Further, the redeem script comprises a cryptographic key. From this, we have a step of further comprising the step of using one or more computing agents to perform the step of any preceding state.

The DFA is constructed on a blockchain, the method comprising the steps of executing a program which is arranged to monitor for and/or receive an input signal and, responsive to the input signal, generate a blockchain transaction Tx_2 which comprises an unspent output ($UTXO$) and spends an output of a previous transaction Tx_1 ; wherein the output of previous transaction Tx_1 comprises a locking script which includes an identifier associated with a first state of the DFA, and the unspent output ($UTXO$) of transaction Tx_2 comprises a locking script which includes a further identifier associated with a further state of the DFA.

A method allows for the implementing a DFA on a blockchain, comprising the steps using at least one input signal to execute at least one condition and, based on the outcome of the execution of the condition, perform an action in accordance with a state transition table for the DFA, wherein performance of the action is identifiable from the state of a blockchain ledger.

This system comprises of a blockchain platform and at least one computing agent arranged to implement the DFA via the blockchain.

References

1. T. Hvitved, *Contract Formalisation and Modular Implementation of Domain-Specific Languages*. Faculty of Science, University of Copenhagen, Ph.D. Thesis (2012)
2. C. Molina-Jimenez et al., Run-time monitoring and enforcement of electronic contracts. *Electron. Commer. Res. Appl.* **3**, 108 (2004)

3. M.D. Flood, O.R. Goodenough, *Contract as Automaton: The Computational Representation of Financial Agreements*. OFR Working Paper 1504 (2015)
4. Botman, *nCrypt: Umbrella Document*, WP0238 (2016)

The following resources provide background material relating to the technological background of the present process

5. XBRL Homepage. <https://www.xbrl.org/>. Last accessed 26 Jan 2019
6. On Bitcoin script. Bitcoin Wiki Homepage. <https://en.bitcoin.it/wiki/Script>. Last accessed 26 Jan 2019
7. On BIPs. Github Homepage. <https://github.com/bitcoin/bips/blob/master/bip-0065.mediawiki>. Last accessed 26 Jan 2019
8. On “salt.” Aspheute Homepage. <http://www.aspheute.com/english/20040105.asp>. Last accessed 26 Jan 2019
9. On “salt.” Jasypt Homepage. <http://www.jasypt.org/howtoencryptuserpasswords.html>. Last accessed 26 Jan 2019

Closest Fit Approach Through Linear Interpolation to Recover Missing Values in Data Mining



Sanjay Gaur, Darshanaben D. Pandya and Deepika Soni

Abstract Data in the dataset is always remaining as the basic building blocks for any query and further task and decisions. If basis data is incomplete or dataset have missing values then one cannot assume about well up to date final reports. In data mining, missing values recognition and recovery is still major issue with irregular data. To overcome from such situation, there is need of statistical or numerical techniques to recover the missing values in the dataset. Missing values in the dataset or database always cause of ambiguity and its affects final results, accuracy of query and reduce decision-making capacity. The present paper is an attempt to recover missing values using closest fit approach through linear interpolation. There is application of the concept of linear approach is used to recover the missing values.

Keywords Data mining · Attribute · Missing values · Closest fit · Approach

1 Introduction

In general, all the reports and queries are performed by the help of database. Data in the database remains in the tabular form, or we can say that in the form of dataset. Dataset are basically attributes of the concern relation, whereas the record set is combination of various fields. It is clear that data in the dataset remains as basic facts and these are used for any query and further task and decisions. Due to various reasons, sometimes there is unavailability of complete data in the dataset. If dataset is incomplete or dataset have missing values, it directly affects the final reports. In data mining, missing values recognition and recovery are still most important

S. Gaur (✉)

Jaipur Engineering College and Research Center, Jaipur, India
e-mail: sanjay.since@gmail.com

D. D. Pandya

Madhav University, Pindwara, Sirohi, Rajasthan, India
e-mail: pandya_darshana@rediffmail.com

D. Soni

University of Technology, Jaipur, India

© Springer Nature Singapore Pte Ltd. 2020

X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_44

issue. Missing values in the dataset or database always cause of ambiguity and its affects final results, accuracy of query and reduce decision-making capacity. It is mandatory to solve such problem before moving toward any query and report preparation. To conquer such state of affairs, there is requirement of statistical or numerical techniques to recover the missing values in the dataset.

Linear interpolation is numerical method which is used to generate the values by the help of available data in the dataset in efficient manner. The present paper is an attempt to recover missing values using closest fit approach through linear interpolation. There is application of the concept of linear interpolation approach is used to recover the missing values.

2 Formulation of Problem

The present approach considers the closest fit approach for missing data recovery. Here, we initially assess the complete dataset values for missing value situations.

At this point, there are two variable X and Y denoted as year and dataset value. Variable X (year) is fixed for other attributes and Y, which have missing values. Attributes for Y are changeable, whereas X is stable for present study.

At the starting, examine the entire table dataset including variable X and Y. Here, it is noticeable that Y is the variables which consist of missing data value set. The search indicator used to point out the missing data place in the variable Y, the initial value of the pointer is $Y[0]$ and the last value or end one is symbolized by $Y[n-1]$.

When search pointer identifies the missing element for attribute, then subscript of the variable/element will be recorded by the pointer. The NULL value and space in the attribute are identified by the search pointer. Once such types of value identified in the dataset/attribute, the pointer remain stay at the subscript and process the activities to recover the missing values. The present element or subscript is denoted by (X_i) . At this moment, it is noticeable that (Y_i) is adjacent to (X_i) . Preceding data as (Y_0) from the missing value data and first subsequent value as (Y_1) , we consider (X_0) to store the value of preceding element of X_i and (X_1) to store the value of succeeding of X_i . At this time apply a loop 'for $i = 0$ to $i < n-1$ ' passes.

$$Y_0 = \text{value}(Y_{i-1}) \quad (1)$$

Y_0 , preceding value from Y_i

$$Y_1 = \text{value}(Y_{i+1}) \quad (2)$$

Y_1 succeeding value from Y_i

$$X_0 = \text{value}(X_{i-1}) \quad (3)$$

X_0 preceding element from X_i

$$X_1 = \text{value}(X_{i+1}) \quad (4)$$

X_1 succeeding value from X_i

$$X_i = \text{value}(X_i) \quad (5)$$

where, $Y_0, Y_1, X_0, X_1 \neq \text{'NULL'}$

At next stage, initialize the variables of this approach

$$\text{Sum} = 0 \quad (6)$$

Now, Initialize the Sum variable from 0, and storing final result after processing. So, initially variable $\text{Sum} = 0$. Then, we make calculation of Sum for estimated value.

$$\text{Sum} = Y_0 + ((Y_1 - Y_0) * (X_1 - X_0) / (X_1 - X_0)) \quad (7)$$

Then, assign sum to Y estimated value.

$$Y_{\text{est}} = \text{Sum} \quad (8)$$

Then, after assign estimated value to missing value place so assigning is done.

$$\begin{aligned} \text{value}(Y_i) &= Y_{\text{est}} \\ \text{counter } i &= i + 1 \\ \text{loop is running until } i &< n. \end{aligned} \quad (9)$$

counter $i = i + 1$

loop is running until $i < n$.

3 Algorithm

```

Attribute  $X = \{ X_1, \dots, X_n \}$ ,  $Y = \{ Y_1, \dots, Y_n \}$ 
Where  $X = X_{\text{obs}} + X_{\text{mis}}$ 
 $X_{\text{obs}} = \{ X_1, \dots, X_k \}$  // Attribute values observed
 $X_{\text{mis}} = \{ X_{k+1}, \dots, X_n \}$  // Attribute values missing
 $Y = Y_{\text{obs}} + Y_{\text{mis}}$ 
 $Y_{\text{obs}} = \{ Y_1, \dots, Y_k \}$  // Attribute values observed
 $Y_{\text{mis}} = \{ Y_{k+1}, \dots, Y_n \}$ 
array(X) == array(Y)
Read  $X = \{ X_1, \dots, X_n \}$ ,  $Y = \{ Y_1, \dots, Y_n \}$ 
for i=0 to n-1 do // for scanning data
  if (value(Yi) == 'NULL') then
     $Y_0 = \text{value}(Y_{i-1})$  // previous value from Yi.
     $Y_1 = \text{value}(Y_{i+1})$  // value of succeeding of Yi.
     $X_0 = \text{value}(X_{i-1})$  // preceding element of Xi.
     $X_1 = \text{value}(X_{i+1})$  // succeeding from Xi.
     $X = \text{value}(X_i)$ 
  where  $Y_0, Y_1, X_0, X_1, X \neq \text{NULL}$ 
  Sum=0 // Initializing variables
  Sum =  $Y_0 + ((Y_1 - Y_0) * (X - X_0) / (X_1 - X_0))$  // Estimated value calculation
   $Y_{\text{est}} = \text{Sum}$  // predicted value
  value (Yi) =  $Y_{\text{est}}$  // transfer predicated value
  i = i+1 // i counter increment
  repeat until (i < n)
end loop
stop.

```

4 Discussion of Results

Analysis [mean]: According to Table 1 the average value of carbon emissions from coal, oil and natural gas are 2109, 2262 and 879, respectively. In the missing value condition, values are recorded as 2129 for coal, 2276 for oil and 901 for natural gas. After filling of missing values from the calculated estimated values, the results are 2109 for coal, 2258 for oil and 879 for natural gas, respectively. Here, it is found that after estimation of missing value by proposed method, values are very close to original value and central tendency values are almost equal to the original set values.

Standard Deviation: Here, it is originate that later than generation of missing value by proposed method, values are very close to original value and value of the standard deviation are almost equal to the standard deviation of original set values.

Coefficient of Variation: it is found that after estimation of missing value by proposed method, values of the coefficient of variation are not very or we can say, CV is similar to CV of original dataset.

Table 1 Closest fit approach through linear interpolation: global carbon dioxide emission from fossil burning by fuel type (Carbon emission in million tones)

S. No.	Year	Standard dataset				Missing value dataset				Recovered dataset			
		Coal	Oil	Natural Gas		Coal	Oil	Natural Gas		Coal	Oil	Natural Gas	
1	1960	1410	849	235		1410	849	235		1410	849	235	
2	1961	1349	904	254		1349	-	254		1349	<u>919</u>	254	
3	1962	1351	980	277		1351	980	277		1351	980	277	
4	1963	1396	1052	300		1396	1052	300		1396	1052	300	
5	1964	1435	1137	328		1435	1137	-		1435	1137	<u>326</u>	
6	1965	1460	1219	351		1460	1219	351		1460	1219	351	
7	1966	1478	1323	380		1478	1323	380		1454	1323	380	
8	1967	1448	1423	410		1448	-	410		1448	<u>1437</u>	410	
9	1968	1448	1551	446		1448	1551	446		1448	1551	446	
10	1969	1486	1673	487		1486	1673	487		1486	1673	487	
11	1970	1556	1839	516		1556	1839	-		1556	1839	521	
12	1971	1559	1946	554		1559	1946	554		1559	1946	554	
13	1972	1576	2055	583		1576	2055	583		<u>1570</u>	2055	583	
14	1973	1581	2240	608		1581	-	608		1581	<u>2150</u>	608	
15	1974	1579	2244	618		1579	2244	618		1579	2244	618	
16	1975	1673	2131	623		1673	2131	623		1673	2131	623	
17	1976	1710	2313	650		1710	2313	-		1710	2313	<u>636</u>	
18	1977	1766	2395	649		1766	2395	649		1766	2395	649	
19	1978	1793	2392	677		1793	2392	677		1827	2392	677	
20	1979	1887	2544	719		1887	-	719		1887	<u>2407</u>	719	

(continued)

Table 1 (continued)

Standard dataset		Missing value dataset				Recovered dataset				
S. No.	Year	Coal	Oil	Natural Gas	Coal	Oil	Natural Gas	Coal	Oil	Natural Gas
21	1980	1947	2422	740	1947	2422	740	1947	2422	740
22	1981	1921	2289	756	1921	2289	756	1921	2289	756
23	1982	1992	2196	746	1992	2196	–	1992	2196	751
24	1983	1995	2177	745	1995	2177	745	1995	2177	745
25	1984	2094	2202	808		2202	808	2096	2202	808
26	1985	2237	2182	836	2237	–	836	2237	2246	836
27	1986	2300	2290	830	2300	2290	830	2300	2290	830
28	1987	2364	2302	893	2364	2302	893	2364	2302	893
29	1988	2414	2408	936	2414	2408	–	2414	2408	933
30	1989	2457	2455	972	2457	2455	972	2457	2455	972
31	1990	2409	2517	1026		2517	1026	2399	2517	1026
32	1991	2341	2627	1069	2341	2627	1069	2341	2627	1069
33	1992	2318	2506	1101	2318	2506	1101	2318	2506	1101
34	1993	2265	2537	1119	2265	2537	1119	2265	2537	1119
35	1994	2331	2562	1132	2331	2562	–	2331	2562	1136
36	1995	2414	2586	1153	2414	2586	1153	2414	2586	1153
37	1996	2451	2624	1208		2624	1208	2447	2624	1208
38	1997	2480	2707	1211	2480	2707	1211	2480	2707	1211
39	1998	2376	2763	1245	2376	2763	1245	2376	2763	1245

(continued)

Table 1 (continued)

S. No.	Standard dataset				Missing value dataset				Recovered dataset				
	Year	Coal	Oil	Natural Gas	Coal	Oil	Natural Gas	Coal	Oil	Natural Gas	Coal	Oil	Natural Gas
40	1999	2329	2716	1272	2329	2716	1272	2329	2716	1272	2329	2716	1272
41	2000	2342	2831	1291	2342	-	1291	2342	2779	1291	2342	2779	1291
42	2001	2460	2842	1314	2460	2842	1314	2460	2842	1314	2460	2842	1314
43	2002	2487	2819	1349	2487	2819	1349	2487	2819	1349	2487	2819	1349
44	2003	2638	2928	1399	2638	2928	1399	2638	2928	1399	2638	2928	1399
45	2004	2850	3032	1436	2850	3032	1436	2850	3032	1436	2850	3032	1436
46	2005	3032	3079	1479	3032	3079	1479	3032	3079	1479	3032	3079	1479
47	2006	3193	3092	1527	3193	-	1527	3193	3083	1527	3193	3083	1527
48	2007	3295	3087	1551	3295	3087	1551	3295	3087	1551	3295	3087	1551
49	2008	3401	3079	1589	3401	3079	1589	3401	3079	1589	3401	3079	1589
50	2009	3393	3019	1552	3393	3019	1552	3393	3019	1552	3393	3019	1552
	MEAN	2109	2262	879	2129	2276	901	2109	2258	879	2109	2258	879
	S.D	567.89	621.13	400.27	586.60	602.54	410.80	568.04	618.02	400.41	568.04	618.02	400.41
	C.V	0.27	0.27	0.46	0.28	0.26	0.46	0.27	0.27	0.46	0.27	0.27	0.46

Source www.earth-policy.com

Analysis of Variance: Testing of assumption that

H0 $\mu_1 = \mu_2 = \mu_3$ against the alternative

H1 at least two μ different

For testing the hypothesis following arrangement have been done:

ANOVA test result for Coal

Source of variation	SS	df	MS	F	P-value	F crit
Between groups	11,634.56	2	5817.281	0.017,674	0.982,484	3.060292
Within groups	46,409,790	141	329,147.4			
Total	46,421,425	143				

Observed value at 5% Level of Significance = 0.0176, the F critical value is 3.06, so hypothesis/assumption is accepted.

ANOVA test result for Oil

Source of variation	SS	df	MS	F	P-value	F crit
Between groups	8346.521	2	4173.261	0.011051	0.98901	3.06076
Within groups	52,868,005	140	377,628.6			
Total	52,876,352	142				

Observed value at 5% Level of Significance = 0.0110, the F critical value is 3.06, so hypothesis/assumption is accepted.

ANOVA test result for Natural Gas

Source of variation	SS	df	MS	F	P-value	F crit
Between groups	14,795.25	2	7397.625	0.045424	0.955606	3.060292
Within groups	22,962,912	141	162,857.5			
Total	22,977,708	143				

Observed value at 5% Level of Significance = 0.045, the F critical value is 3.06, so hypothesis/assumption is accepted.

Decision and Conclusion: Given that F (Observed/Calculated) < 3.06 for Coal, Oil and Natural gas ANOVA (One way) test. In case, hypothesis is accepted in all cases, therefore it is considerable that no significant difference found between groups regarding mean value.

5 Conclusion

In the present study, numerical techniques linear Interpolation is used in applied nature to recover the missing values in the dataset. In the database, three dataset namely coal, Oil and natural Gas are taken, here, along with Year as common for all. After applying proposed algorithm, suitable values are received as estimated/recovered value at the place of missing values.

According to measurement of central tendency, SD and CV result are significant. One way ANOVA test also gives significant result with acceptance of hypothesis. So it can be said that the results are statistically significant. Finally, it can be said that proposed techniques are significant for small database which consist of linear trends in the dataset.

References

1. P.D. Allison, Estimation of Linear Models with Incomplete data, Social Methodology (Jossey Bass, San Francisco, 1987), pp. 71–103
2. P.D. Allison, *Missing data* (Sage Publication, Thousand Oaks CA, 2001)
3. S.F. Buck, A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *J. Royal Statistical Society, Series B* **2**, 302–306 (1960)
4. L. Chen, M.T. Drane, R.F. Valois, J.W. Drane, Multiple imputation for missing ordinal data. *J. Mod. Appl. Stat. Methods* **4**(1), 288–299 (2005)
5. S. Gaur, M.S. Dulawat, A perception of statistical inference in data mining. *Int. J. Comput. Sci. Commun.* **1**(2), 653–658 (2010)
6. S. Gaur, M.S. Dulawat, Univariate analysis for data preparation in context of missing values. *J. Comput. Math. Sci.*, **1**(5), 628–635 (2010)
7. S. Gaur, M.S. Dulawat, A closest fit approach to missing attribute values in data mining. *Int. J. Adv. Sci. Technol.* **2**(4), 18–24 (2011)
8. S. Gaur, Closest fit approach to handle odd size missing block values. *Int. J. Math. Arch.* **3**(7) (2012)
9. J.W. Grzymala-Busse, Data with missing attribute values: generalization of in-discernibility
10. Realtion and rules induction, Transactions of rough sets. *Lect. Notes in Comput. Sci. J. Subline*, **1**, 8–95 (2004). (Springer-Verlag)
11. D.B. Rubin, Inference and missing data. *Biometrika* **63**, 581–592 (1976)
12. S. Sharma, S. Gaur, Contiguous agile approach to manage odd size missing block in data mining. *Int. J. Adv. Res. Comput. Sci.* **4**(11), 214–217 (2013)

Correction to: Process Driven Access Control and Authorization Approach



John Paul Kasse, Lai Xu, Paul de Vrieze and Yuewei Bai

Correction to:
Chapter “Process Driven Access Control and Authorization Approach” in: X.-S. Yang et al. (eds.),
Fourth International Congress on Information and Communication Technology,
Advances in Intelligent Systems and Computing 1041,
https://doi.org/10.1007/978-981-15-0637-6_26

In the original version of the book, the author name has been updated from “Paul deVrieze” to “Paul de Vrieze” in the Chapter “Process Driven Access Control and Authorization Approach”. The chapter and book have been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-15-0637-6_26

© Springer Nature Singapore Pte Ltd. 2020
X.-S. Yang et al. (eds.), *Fourth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1041, https://doi.org/10.1007/978-981-15-0637-6_45

Author Index

A

Abbod, Maysam, [431](#)
Abdulkader, Omar, [375](#)
Abid, Mohamed, [481](#)
Addaim, Adnane, [259](#)
Adigun, Matthew O., [397](#)
Alabdullah, Ahmed, [481](#)
Alfred, Rayner, [127](#)
Al Ghamdi, Bandar, [375](#)
Almaadeed, Noor, [333](#)
Al-Maadeed, Somaya, [333](#)
Al-Saggaf, Ubaid M., [303](#)
Alshammari, Hamoud, [481](#)
Al-Taie, Inas, [73](#)
Amarasinghe, Thushara Madushanka, [385](#)
Anikin, Anton, [365](#)
Aponso, Achala Chathuranga, [343](#), [353](#), [385](#)
Aseeri, Mohammed, [481](#)
Astorga, Gino, [459](#)
Ayyappaa, Nithin, [285](#)
Azeez, Nassr, [73](#)

B

Bai, Yuewei, [313](#)
Bamhdi, Alwi M., [375](#)
Basbrain, Arwa, [73](#)
Basukoski, Artie, [343](#)
Beranek, Marek, [201](#)
Bolsunovskaya Marina, V., [171](#)
Bouزيد, Aicha, [229](#)
Buqi, Raol, [41](#)
Burnashev, Rustam A., [191](#)

C

Calin, Mihnea Andrei, [323](#)
Carcioffi, Sara, [41](#)
Cardoso, Mário, [489](#)
Chandradeva, Lakshika Sammani, [385](#)
Chinna Veeresh, A., [271](#)
Clark, Adrian, [73](#)
Coetzee, Marijke, [293](#)
Crawford, Broderick, [459](#)

D

Damjanović, Slaven, [419](#)
Dascalu, Mihai, [323](#)
de Vrieze, Paul, [313](#)
Domb, Menachem, [113](#)

E

Elharrouss, Omar, [333](#)
Enikeev, Arslan I., [191](#)

F

Fawzy, Heba, [15](#)
Frolova, Elena, [179](#)

G

Gardella, Orlando, [459](#)
Gaur, Sanjay, [513](#)
Ghorbel, Oussama, [481](#)
Ghulam, Mohi-Ud-Din, [471](#)
Gimhani, Gayakshika, [353](#)
Greaves, Brian, [293](#)
Gretete, Driss, [259](#)

Guimarães, Tiago, 489

H

Hari Prasad, S. V., 271
 Härting, Ralf-Christian, 33
 Hasan, Hasanein, 431

J

Jambi, Kamal, 375
 Jiang, Zejun, 471

K

Kalansuriya, Chamalka Seneviratne, 343
 Kasamani, Bernard Shibwabo, 241
 Kasse, John Paul, 313
 Khan, Shujaat, 303
 Korshunov, Gennady, 179
 Kovar, Vladimir, 201
 Krey, Mike, 1
 Krishnarajah, Naomi, 353, 385
 Kultsova, Marina, 365

L

Lee, Samuel Sangkon, 157
 Lemus-Romani, José, 459
 Leshem, Guy, 113
 Leung, Wai Sze, 293
 Litovkin, Dmitry, 365
 Litunya, Duncan, 241
 Liu, Zhiqiang, 471
 Lun, Chew Ye, 127
 Luo, Jianchao, 471

M

Magdi, Dalia A., 15
 Mahroo, Atieh, 41
 Makarov Aleksei, S., 171
 Makarova, Irina, 219
 Manrique-Losada, Bell, 449
 Mba, Ijeoma N., 397
 Militaru, Gheorghe, 323
 Mohammadi Ziabari, S. Sahand, 91
 Moinuddin, Muhammad, 303
 Montalvo-Garcia, Jhon, 449
 Murali, Rahul, 285
 Muthu Manikandan, M., 285

N

Naseem, Imran, 303

Nazarevich, Stanislav, 179
 Nigro, Christian, 55
 Nigro, Libero, 55
 Nikolaev, Timur, 219
 Nolich, Massimiliano, 41
 Nxumalo, Mandisa N., 397

P

Pandya, Darshanaben D., 513
 Patel, Ahmed, 375
 Patrono, Luigi, 419
 Pavlidou, Meropi, 407
 Perković, Toni, 419
 Portela, Carlos Filipe, 489

Q

Quintero, Juan Bernardo, 449

R

Raikov, Alexander, 147
 Rakotomalala, Hery Frédéric, 83
 Reichstein, Christopher, 33
 Rodrigues, Joel J. P. C., 419
 Ruseti, Stefan, 323

S

Saadoon, Raed, 431
 Sacco, Marco, 41
 Sadiq, Alishba, 303
 Saidi, Hanane, 259
 Sakat, Raid, 431
 Salas-Fernández, Agustín, 459
 Santos, Manuel Filipe, 489
 Sciammarella, Paolo F., 55
 Shanableh, Tamer, 137
 Shubenkova, Ksenia, 219
 Silva De, Minoli, 385
 Sirbu, Maria-Dorinela, 323
 Smidi, Ghaya, 229
 Smirnov, Vladimir, 179
 Šolić, Petar, 419
 Soni, Deepika, 513
 Soto, Ricardo, 459
 Soundarakumar, M., 271
 Spoladore, Daniele, 41
 Sriram, P. R., 285

T

Thayananthan, Vijey, 375
 Totohasina, André, 83

Trausan-Matu, Stefan, [323](#)
Treur, Jan, [91](#)
Tyagi, Vipin, [271](#)

U

Usman, Muhammad, [303](#)

V

Vakhitov, Galim Z., [191](#)
Venkata Siva Prasad, S., [271](#)

W

Wang, Lifang, [471](#)

Wright, Craig S., [499](#)

X

Xu, Lai, [313](#)

Y

Yahya, Wafa, [73](#)
Yan, Xuanlin, [471](#)

Z

Żabiński, Krzysztof, [219](#)
Zhu, Ye, [471](#)
Zioutas, George, [407](#)