Lecture Notes in Networks and Systems 25

Mostafa Ezziyyani Mohamed Bahaj Faddoul Khoukhi *Editors*

Advanced Information Technology, Services and Systems

Proceedings of the International Conference on Advanced Information Technology, Services and Systems (AIT2S-17) Held on April 14/15, 2017 in Tangier



Lecture Notes in Networks and Systems

Volume 25

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland e-mail: kacprzyk@ibspan.waw.pl

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Advisory Board

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

e-mail: gomide@dca.fee.unicamp.br

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Turkey

e-mail: okyay.kaynak@boun.edu.tr

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA and Institute of Automation, Chinese Academy of Sciences, Beijing, China

e-mail: derong@uic.edu

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada and Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

e-mail: wpedrycz@ualberta.ca

Marios M. Polycarpou, KIOS Research Center for Intelligent Systems and Networks, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus e-mail: mpolycar@ucy.ac.cy

Imre J. Rudas, Óbuda University, Budapest Hungary

e-mail: rudas@uni-obuda.hu

Jun Wang, Department of Computer Science, City University of Hong Kong

Kowloon, Hong Kong

e-mail: jwang.cs@cityu.edu.hk

More information about this series at http://www.springer.com/series/15179

Mostafa Ezziyyani · Mohamed Bahaj Faddoul Khoukhi Editors

Advanced Information Technology, Services and Systems

Proceedings of the International Conference on Advanced Information Technology, Services and Systems (AIT2S-17) Held on April 14/15, 2017 in Tangier



Editors Mostafa Ezziyyani Faculty of Sciences and Technologies Tangier Morocco

Mohamed Bahaj Faculty of Sciences and Technologies University Hassan 1st Settat Morocco Faddoul Khoukhi Faculty of Sciences and Technologies Mohammedia Morocco

ISSN 2367-3370 ISSN 2367-3389 (electronic) Lecture Notes in Networks and Systems ISBN 978-3-319-69136-7 ISBN 978-3-319-69137-4 (eBook) https://doi.org/10.1007/978-3-319-69137-4

Library of Congress Control Number: 2017957546

© Springer International Publishing AG 2018

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, express 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.

Printed on acid-free paper

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

Preface

This book provides an interdisciplinary collaboration in which scientists and professionals can share their research results. It brings also new challenges of insights into the fields of Modern Information Engineering Concepts and Communication Systems.

The selected papers involve great efforts to develop efficient techniques and tools for emerging areas related to Modern Information Engineering and Communication Systems. It also report new solutions for deployment, manipulation and exploitation of the advanced systems in Information Technology.

This book brings together the investigation result of researchers and practitioners from academia and industry to share novel ideas, results, experiences and work-in-process for establishing new collaborations in these areas.

It can be categorized by the involvement of five domains of treatment and management of information technology:

Advances in Software Engineering. The objective of this part is to communicate recent and projected advances in computer-based engineering techniques. This part cover the development and maintenance of software engineering to satisfy customer requirements on reliably and efficiently.

Advances in Web Technologies, Semantics, and Future Internet. This includes semantic Web and big data, semantic Web and conceptual knowledge representation, Web semantics and conceptual modeling, ontology engineering, learning and ontology evolution, semantic Web, information extraction, automatic, and semiautomatic generation of metadata, semantics for ubiquitous and cloud computing.

Advances Networking and Sensor Networks. This part is intended to discuss recent advances in the broad and quickly evolving fields of communication networks, and to highlight key issues and develop visions for networks of the future comprising sensors, actuators, mobile systems, and communicating objects.

Cloud, Parallel, Distributed, and High-Performance Computing. This part presents the latest fundamental advanced research result of cloud computing and identifies the emerging research topics related to distributed system and high-performance computing.

IR, Big Data, Business Intelligence, and Knowledge Management. This includes domain-driven data mining, postprocessing rules for actions, domain-driven customer analytics, big data, information retrieval, roles of human intelligence in AKD, maximal pattern-based cluster, and ontology mining.

We would like to thank the staff at Springer for their enthusiastic support and guidance during the preparation of the book. In particular, our special thanks to Dr. Thomas Ditzinger for kind support.

Mostafa Ezziyyani Mohamed Bahaj Faddoul Khoukhi

Contents

Advances in Software Engineering

3
2
21
36
50
59
59
78
36

Content-Based Image Retrieval Using Gabor Filters and 2-D ESPRIT Method	95
A Retirement Pension from a Supply Chain Side: Case of the Moroccan Retirement Pension Houda Mezouar and Abdellatif El Afia	103
Advances in Web Technologies, Semantics and Future Internet	
Creating Multidimensional Views from RDF Sources	119
An Ontology Based Approach to Organize Supplier and Transportation Provider Selection Negotiation in Multi-agent System Model Iman Achatbi, Khalid Amechnoue, and Saloua Aoulad Allouch	129
Deep Neural Networks Features for Arabic Handwriting Recognition Mustapha Amrouch and Mouhcine Rabi	138
SCH-WSD: A Semantic-Conceptual Hybrid Approach for Web Services Discovery Hicham Laabira, Khalid El Fazazy, and Redouane Ezzahir	150
Optimal Regulation of Energy Delivery for Community Microgrids Based on Constraint Satisfaction and Multi-agent System Mostafa Ezziyyani and Loubna Cherrat	162
Using Image Segmentation in Content Based Image Retrieval Method Mohamed Ouhda, Khalid El Asnaoui, Mohammed Ouanan, and Brahim Aksasse	179
Alignment of IT Frameworks for Corporate Governance	196
A Design Requirements Framework for Mobile Learning Environment	208
Knowledge Management in Business, A Multi-desciplinar Science and A State of Mind Ben Laadar Hajar and Cherti Ilias	216
Advances Networking and Sensor Networks	
Survey of Security in Software-Defined Network Nadya El Moussaid, Ahmed Toumanari, and Maryam El Azhari	227

Contents

Weakness in Zhang et al.'s Authentication Protocol for Session Initiation Protocol Mourade Azrour, Yousef Farhaoui, and Mohammed Ouanan	239
How Mobile Nodes Influence Wireless Sensor Networks Security and Lifetime	252
A Novel Smart Distribution System for an Islanded Region	269
Taxonomy of Routing Protocols in MANETs Younes Ben Chigra, Abderrahim Ghadi, and Mohamed Bouhorma	280
Cloud, Parallel, Distributed and High Performance Computing	
Allocation Strategy for Cloud Datacenter Based on Multi Agent and CP Approach	291
A Trusted Way for Encryption Key Management in Cloud Computing Saad Fehis, Omar Nouali, and Mohand-Tahar Kechadi	302
Use of Cloud Computing Technologies for Geographic Information Systems Ahmed Ziani and Abdellatif Medouri	315
New Real Time Cloud Telemedicine Using Digital Signature Algorithm on Elliptic Curves Asma Jebrane, Naima Meddah, Ahmed Toumanari, and Mohamed Bousseta	324
Scalable Lightweight ABAC Scheme for Secure Sharing PHR in Cloud Computing Naima Meddah, Asma Jebrane, and Ahmed Toumanari	333
IR, Big Data, Business Intelligence, and Knowledge Management	
Arabic Stemming Techniques as Feature Extraction Appliedin Arabic Text ClassificationSamir Boukil, Fatiha El Adnani, Abd Elmajid El Moutaouakkil,Loubna Cherrat, and Mostafa Ezziyyani	349
A Comparative Study of the Four Well-Known Classification Algorithms in Data Mining Safae Sossi Alaoui, Yousef Farhaoui, and Brahim Aksasse	362

Advanced SQL-to-SPARQL Query Transformation Approach Nassima Soussi and Mohamed Bahaj	374
Migration from Relational Databases to HBase:A Feasibility AssessmentZakaria Bousalem, Ilias Cherti, and Gansen Zhao	383
Big Data and IoT: A Prime Opportunity for Banking Industry Abdeljalil Boumlik and Mohamed Bahaj	396
Big Data Analytics Applied for Control Systems	408
Detecting Network Intrusions Using Multi-class Logistic Regression and Correlation-Based Feature Selection Taha Ait tchakoucht and Mostafa Ezziyyani	416
The Optimization of Search Engines to Improve the Rankingto Detect User's IntentSalma Gaou and Aissam Bekkari	427
Hybrid HMM/MLP Models for Recognizing UnconstrainedCursive Arabic Handwritten TextMouhcine Rabi, Mustapha Amrouch, and Zouhir Mahani	438
Reducing Crowding in Hospital Inpatient Unit Using Queuing Theory Sara Jebbor, Abdellatif El Afia, and Raddouane Chiheb	449
Hybrid Penguins Search Optimization Algorithm and GeneticAlgorithm Solving Traveling Salesman ProblemIlyass Mzili, Mohammed Essaid Riffi, and Fatiha Benzekri	461
The Particularities of the Counter Propagation Neural Network Application in Pattern Recognition Tasks Khatir El Haimoudi, Ikram Issati, and Ali Daanoun	474
Converting Temporal Relational Database into Temporal Object Relational Database Soumiya Ain El Hayat and Mohamed Bahaj	488
Implementing of a Binary Data Generator on a FPGA Card M. Benzaima, Mensouri Mohammed, Aaroud Abdessadek, and Ali El Hore	496
Towards a Hybrid Method of Construction of a Normalized Domain Ontology Used by Machine Teaching PERO2 Mostafa Chahbar, Ali Elhore, and Younes Askane	503
Author Index	517

Advances in Software Engineering

Factors Influencing the Adoption of Ambient Assisted Living Technologies by Healthcare Providers in the Kingdom of Saudi Arabia

Majid H. Alsulami^{1,2(18)}, Anthony S. Atkins², and Russell J. Campion²

¹ College of Community, Shaqra University, Shaqra, Saudi Arabia malsulami@su.edu.sa
² School of Computing and Digital Technologies, Staffordshire University, Stoke on Trent, ST24DE, UK {a.s.atkins,r.j.campion}@staffs.ac.uk

Abstract. The ageing population is considered to be a global challenge because of the reduction in fertility and the increase of life expectancy. In Saudi Arabia, the ageing population continues to age (>60 years of age) currently (5%) compared to other age groups. In 2050, it will rise rapidly to 20.9% of the Saudi population. Ambient Assisted Living (AAL) technology plays an important role in assisting elderly people to live in their home independently, longer, and improve their quality of life and health and in supporting their daily activities etc. The current research aims at examining the barriers that healthcare providers in the Kingdom of Saudi Arabia are experiencing in the adoption of AAL technologies among the elderly. The study identified some of the challenging issues with the increasing number of elderly people among the population in the country, which has highlighted the need to use AAL technologies to improve the quality of life among the elderly. The research involved a Community of Practice (CoP) study as a method of data collection where data collected was presented and discussed in line with the existing literature review findings. A lack of training, the high cost of AAL devices and the associated Management Information Decision Control System and cultural barriers were the main challenges identified in the research. The research suggests that awareness is important to encourage the elderly to accept the new technology and its potential in improving their quality of life. Training on the usage of these AAL devices should also be offered to ensure that self-care services are encouraged among the elderly who, in some cases, live away from their relatives.

Keywords: Quality of care \cdot AAL \cdot Healthcare \cdot Community of practice \cdot Effectiveness and efficiency \cdot KSA

1 Introduction

In the Kingdom of Saudi Arabia, there are several barriers and challenges that affect healthcare systems. These can be identified as the health of the workforce, finance issues and expenditure, varying patterns of diseases, accessibility to healthcare services, the health insurance system, deployment of electronic health (E-health) strategies and the development of a national system for healthcare information [1].

In particular, a study conducted by [2] stated the barriers that face adoption of E-Health in the Kingdom of Saudi Arabia (KSA) from professionals' perspective. The most important barriers outlined were connectivity of information systems, culture, security and privacy, and financial issues.

Another study illustrated the barriers facing health information system as follows: ICT Infrastructure, cost and time, national policies, social and cultural, educational, organisational and ethical barriers [3].

In this research, barriers in the healthcare sector in KSA will be reviewed. Although, healthcare is prioritised in KSA, there are gaps that should be addressed to improve its effectiveness and efficiency. Given the high commitment by the KSA government to both the quality of care and quantity of health facilities, barriers to efficient healthcare systems prevent the realisation of the strategic goals by the KSA Ministry of Education [4].

The current research will focus on the identification of these barriers, with emphasis on the quality of care, efficiency of healthcare services and performance indicators. The research will address the main barriers that prevent successful achievement of strategic measures to address the gaps in healthcare provision. Although, the country is improving its health services, there remain gaps in its performance due to the barriers that the research aims to understand, particularly with regards to elderly people. Identifying the barriers preventing effectiveness and efficiency in the healthcare sector in KSA will require assessment of cultural issues, distance to the nearest health facility, quality of care and competence and skill in the workforce.

This paper is structured as follows: Sect. 2 provides and examines a literature review and Sect. 3 describes the Community of Practice (CoP). Section 4 analyses the results conducted through the Community of Practice. The discussion is outlined in Sect. 5. The conclusion and future work are presented in Sect. 6.

2 Literature Review

2.1 Healthcare in KSA

According to the Saudi Ministry of Health (2015), KSA has experienced expansive growth in its healthcare system through an improvement in staff training, quality of care and integration of modern technology in the healthcare sector. From 1970 to 2015, KSA has been able to increase the number of hospitals from 74 to 462 with the number of beds correspondingly from 9,039 to 69,394 in the same period. There was an allocation of SR 62 billion (\$16.5 billion) in 2015 for an improvement in healthcare services [5].

The Ministry of Health report in 2015 indicated that there are 41,297 beds in 274 hospitals, which account for 58.5% of hospital beds in MOH, and the remainders are in the private and governmental sectors. In KSA, the total number of:

- Physicians was 86,756 (26% Saudi)
- Nursing staff was 172,483 (38.3% Saudi)

- Pharmacists was 23,626 (21% Saudi)
- Primary Health Care (PHC) centres were 2,282

After a World Health Organisation (WHO) assessment and review of global healthcare provision internationally, the KSA healthcare sector was ranked 26th out of 190 countries that were assessed, higher than Canada (30th), New Zealand (41) and Australia (32).

2.2 Ageing Population in KSA

The ageing population is a concern of the KSA government, especially in terms of access to healthcare services. As the country is experiencing a growth in economic and health sectors, the aged (> 60) are still at risk of health challenges if measures to address the situation are not considered [6, 7].

Ageing has recently become a phenomenon of population trends in KSA with improved healthcare conditions and standard of living. Research has identified those who are 60 years old or above are defined by the United Nation as elderly people.

In 2015, the life expectancy was 74.3 years, which is above the regional average by 6 years and by 4 years globally. The proportion of age group under 5 years was 10.06%, and the proportion of age group under 15 years was 29.12%, which is higher than the global rate (27%). The total fertility rate was 2.69 children per woman [5].

From the projections by the United Nations Department of Population, it is estimated that the population of the aged in KSA is expected to increase. In their estimates, the UN proposes that the aged population (60 years and above) will be 4.3 million people in Saudi Arabia by 2030. In addition, the number of elderly people is expected to increase to 20.9% in 2050, which will be translated as an increase from 5% to 20.9% of the total population [8].

The projections by the United Nations (2015) are determined through the consideration of both demographic and socioeconomic factors. Demographic factors illustrate that the fertility rate will decrease in KSA from 5.5 children per woman in 1990–1995 to an expected 1.8 in 2045 – 2050. On the other hand, life expectancy has also changed from 70 years in 1990–1995 to be estimated 78.4 years in 2045 – 2050 [8].

2.3 Ambient Assisted Living Technologies

Ambient Assisted Living (AAL) is a technological innovation that is intended to improve the life quality among the elderly and has the capability to support their requirements in their later years through the application of technology [9]. It is developed through the consideration of automation of their homes and through assistive domesticity. It provides people in this age group (elderly) with assistance in carrying out their daily activities, prolonging their life expectancy, and improves their social life and communication (Blasco, et al., 2014). In addition, research has shown that AAL tends to improve selfcare for patients with lifelong illnesses and conditions.

According to [6, 7], given the projected increase in the elder population, AAL is becoming an important health consideration for KSA. In addition, the high elderly

population projection is evidence that there will exist barriers to elderly care in KSA that may incur an increased financial expenditure on the Ministry of Health in meeting the needs of the elderly [10]. It is found that the elder population is prone to vulnerabilities, especially in healthcare, where research has shown that AAL technology requires close monitoring that will ensure the improved quality of life, while delivering to the elderly the crucial services that have a transformative benefit in their lives.

From the approach presented above, there is a gap in terms of monitoring the effectiveness of AAL that the current study intends to address effectively [11]. In addition, it is important to understand the implications, challenges, barriers and opportunities that arise in improving the quality of care among the elderly through the application of an effective mechanism of AAL [9]. Understanding the requirements, population trends and patterns will facilitate effective planning for the elderly and reduce the cases of suffering and poor quality of life among them.

3 Barriers Facing Healthcare in KSA

Assessment of the healthcare sector in KSA reveals significant barriers to optimal healthcare performance. The first factor influencing the health sector is the restrictive cultural setting, especially for women, who are bound by strict cultural limitations [9]. A lack of knowledge and information on health matters is an additional barrier that suppresses the commitment of KSA in promoting better healthcare outcomes [11].

An additional influential barrier is the lack of sufficient healthcare professionals who are required to take care of patients, resulting in a high patient-to-doctor/nurse ratio. Research indicates that the accessibility of healthcare facilities is also a barrier inhibiting successful implementation of healthcare strategies in KSA [10]. The distance a patient travels to access healthcare limits the number of patients served by the hospitals. A clash of culture, ineffectiveness in communication, a lack of desired skills and level of competence, and insufficient food, are the major barriers to accessing effective healthcare outcomes [12, 13].

4 Description of the CoP

The Community of Practice (CoP) is a 'group of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly' [14].

A CoP was conducted in KSA in August 2015. The Arabic language was preferred, and the session was recorded (with their permission). Four CEOs and an Executive Vice President of healthcare providers were involved and the size and complexity of the organisation are outlined as follows:

• Salam Home Health Care delivers a comprehensive range of home healthcare services. It has more than 100 staff, approximately 75 nurses, and more than 10 physicians. Its services deliver to more than 2000 people.

- Saudi Health Services Co has been in the health market since 1984. It provides a diversity of medical utilisations and equipment of the latest technology to all of Saudi Arabia.
- Sela Medical aims to improve health care services in Saudi Arabia by operating and managing facilities. It started in 2005 and promised to spend SR 20 million in each project for the International Renal Care Centres for treating end-stage kidney problems.

The researcher started with an introduction about AAL technologies and how AAL technologies support elderly people who are in good condition, but need some help to do some daily activities, such as shopping, cooking or cleaning, providing the definition of elderly people [15].

The researcher mentioned that two tasks should be performed in relation to this project. Firstly, a questionnaire should be designed to extract the attitudes and perceptions of Saudi Arabians' elderly people. Secondly, a CoP study should be conducted with healthcare providers to elicit the barriers and challenges that face them to deploy AAL technologies in KSA.

The researcher stated the objective of this project is to offer elderly people with a secure and private environment, providing quality of life and more time in their desired own home. This is because there is a high trend in the aged population in Saudi Arabia, which will be increasing in the next 50 years.

The researcher indicated that there are some barriers in the UK, which relate to privacy, security, cost, lack of awareness and lack of experience, etc. These barriers may or may not be the same in KSA; therefore, this CoP study is organised to share the barriers and challenges that face them according to their experience working in KSA in providing different technologies in relation to healthcare services to the elderly.

The participants mentioned that there is a challenge about the high percentage of the growing ageing population in KSA. One of them indicated, 'The Saudi government faces a huge trend in the ageing population in the future'. Therefore, we should notify people, organisations and delegations about this high increase of the ageing population. They said that the infrastructures are not ready to deploy such technologies (AAL technologies).

Some of the important concerns they shared are outlined as follows:

4.1 Concern 1

There is no continuum of care, but if it is found, it may be marginalised usually by a few organisations. It is scattered. In the long term, home care, day care and a hospice are not essential parts of the healthcare system; however, these have just started but will have difficulties to meet the future demands. In addition, the nursing home is not offered regardless of culture. For example, he said, 'My father had a stroke; he is in the ICU for a week. I own a company; thus, I provide him with nursing'. This proposed project should spread and be applied all over Saudi Arabia, because today, if we are unable to take care of our parents, who will take care of us in the future? There are many changes in demographics and economies, thus this project is needed.

4.2 Concern 2

The practice of elderly care by family members will change in the future. The parents will no longer live with their children. We notice that when children get married, they move to their own homes. While parents live in their homes independently, there are no services provided to them, and no one can take care of them. There are not enough specialists to work with the elderly and take care of them. We should graduate more students who major in this field. Further, healthcare providers may not be able to deliver technologies, such as these, to elderly people in KSA. This means we cannot bridge the gap of the needs of the elderly. Elderly people may not be able to pay for these technologies.

4.3 Concern 3

Smart watches are available in KSA, and athletes use them. Furthermore, there are more advanced technologies for those who suffer from some diseases, such as cardiac disease; however, the elderly and their families can use these kinds of technologies and we must study the cultural barrier regarding their use. We might find a problem in convincing families and the elderly to adopt them. There were concerns raised regarding the cost, but it will be great if the Ministry of Health (MOH) adopts this kind of service and technology and provides them to the elderly. Indeed, the MOH pays for whoever has diabetes to use some devices to measure the level of diabetes. This can be a national project that can be useful.

4.4 Concern 4

Elderly people who live alone and are supported by the Ministry of Social Affairs, which has a program for visiting the elderly, and are categorised into two groups: (1) Those who do not have family to take care of them but live in their homes, and the Ministry provides them with some services, such as cleaning the home, buying their groceries, giving rides to visit relatives and providing some entertainment etc.; (2) Those who are deserted and who move into social care homes for care. He suggests using this project to assist the elderly people to start using AAL.

4.5 Concern 5

The care of the elderly is limited, and the Social Affairs Ministry takes care of only the elderly who are so old and who do not have families or relatives. In KSA, we do not have enough information about the elderly, which causes a lack of information among the elderly, healthcare providers, the Ministry of Health and the Ministry of Social Affairs. Elderly people will face problems when using these technologies and unable pay for them; therefore, they should be trained on them and receive financial support from the Saudi government.

5 Results of the CoP

According to the description of the CoP, many barriers that face healthcare providers in KSA can be identified. The first barrier is the cultural barrier that may discourage the elderly from using the AAL devices to facilitate improving the quality of care. This finding is consistent with the findings in the literature review where cultural barriers were considered a major cause of resistance among the elderly in using AAL.

Another, critical challenge identified was the cost of the device, which most of the respondents agreed that most of the elder population are unable to afford to buy. It was suggested that the government should facilitate the elderly in purchasing AAL devices to reduce their burden. Self-care requires the assistance of family members, where another barrier occurs. Most of the elderly tend to be isolated from their relatives and children, resulting in the difficulties in taking care of their healthcare needs.

A lack of training among the elderly on the use and application of AAL is another barrier, to which a respondent noted that they ought to be provided with training on the use of these devices. These results indicate that there are plenty of gaps that have not been addressed in assessing healthcare needs among the elderly in KSA.

Awarness is mentioned as a barrier that faces the healthcare providers. There is a need to make the community, elderly people and practitioners aware about the importance of using AAL technologies, which can be a critical factor that increases the adoption of these technologies.

6 Discussion

Although the country is experiencing a high growth in technology and improvement of healthcare, there are gaps identified in the analysis of health status in the country, especially in the use and the application of AAL among the elderly [16]. A literature search and empirical results demonstrate that the barriers have reduced the effectiveness in the accomplishment of the desired goals. Cultural barriers are some of the challenges that face the integration of AAL technology in addressing the needs and quality of life among the elderly [2, 13].

Research has shown that the country has one of the best healthcare systems, which is improving in terms of quality of care and accessibility, but AAL integration in the elder population is still challenging. The high cost of purchasing AAL devices and the associated Management Information Decision Control System tends to be another significant barrier that should be addressed [17]. The moderately high growth rate, higher life expectancy and decreasing death rate and mortality rate tend to increase the elder population in the country. It is essential to analyse the increasing trends in the elder population in the country for better analysis later in the research. Government intervention measures should include increasing awareness on the use and application of AAL to encourage most elderly people to embrace this technology. The Ministry of Social Affairs should also gather data on the elderly to determine the best approaches that can be integrated into the healthcare system. Self-care programs should also be encouraged to ensure that the population comprising the elderly is considered.

7 Conclusion and Future Work

The research indicated that the healthcare sector in KSA is among the best, and was ranked 26th out of 190 countries. In addition, the MOH is addressing challenges in the healthcare sector to improve the quality of care among the patients. However, the elderly have been identified to experience challenges due to the number of barriers in the use and application of AAL technological devices aimed at improving the quality of life among this group [7, 18]. The gap is identified following an increasing trend in the elder population, resulting in the need to address the gaps in the care of the elderly [19]. The current research concludes that there is a need for training the elderly on the use of AAL devices, create awareness to reduce resistance due to cultural barriers, and present an effective database on the elderly to ensure that the Ministry of Social Affairs can take care of the elderly with improved effectiveness.

Future research should involve a high number of participants to draw diverse views on the role of healthcare barriers in AAL technology for healthcare providers. In addition, it would be important to include participants in the healthcare sector, as well as the Social Affairs Ministry, to analyse the main challenges they experience in the provision of care to the elderly. Moreover, it is important to design a model for training elderly people using AAL technologies.

References

- 1. Kuwait Finance House Research Ltd.: Saudi Arabia Healthcare "healthy living" (2014)
- Almuayqil, S., Atkins, A.S., Sharp, B.: Ranking of E-health barriers faced by Saudi Arabian citizens, Healthcare Professionals and IT Specialists in Saudi Arabia, pp. 1004–1013, July 2016
- Anwar, F., Shamim, A.: Barriers in adoption of health information technology in developing societies. Int. J Adv. Comput. Sci. Applications 2(8), 40–45 (2011)
- 4. Abusaaq, H.I.: SAMA Working Paper: Population Aging in Saudi Srabia (2015)
- 5. Ministry of Health.: MOH-Statistical Book (2015)
- Altamimi, T.: Healthy aging conceptualizations in Saudi Arabia: a systematic review. Int. J. Med. Sci. Public Health 5(4) (2016)
- Alsulami, M.H., Atkins, A.S.: Elderly Saudi Arabians' perceptions and attitudes towards using ambient assisted living technologies'. In: Proceedings of 64th The IIER International Conference, Barcelona, Spain, 4th March 2016, March 2016. ISBN:978-93- 85973-53-6
- 8. United Nations.: World Population Prospects (2015)
- Ansari, R.M., Dixon, J.B., Browning, C.J.: Self-management of type 2 diabetes in middleaged population of Pakistan and Saudi Arabia. Open J. Prev. Med. 4, 396–407 (2014)
- Ansari, R.M., Dixon, J.B., Browning, C.J.: Systematic review of diabetes self-management: focusing on the middle-aged population of Pakistan and Saudi Arabia. Open J. Prev. Med. 5, 47–60 (2015)
- Karlin, N.J., Weil, J., Felmban, W.: Aging in Saudi Arabia: an exploratory study contemporary older persons' views about daily life, health, and the experience of aging. Gerontol. Geriatr. Med. (2016)
- Yusuf, N., Al-sharqi, L., Durrani, F.: A determinant of healthy ageing women education in Saudi Arabia. Int. Bus. Econ. Res. J. 14(2), 355–366 (2015)

- Kronfol, N.M.: Access and barriers to health care delivery in Arab countries: a review. Eastern Mediterr. Health J. 18(12), 1239–1246 (2012)
- 14. Wenger, E.: Communities of practice, pp. 1–6. Cambridge University Press (2005)
- 15. Hossain, M.A., Alamri, A., Almogren, A.S., Hossain, S.K.A., Parra, J.: A framework for a context-aware elderly entertainment support system. Sensors, 10538–10561 (2014)
- Al Modeer, M.A., Hassanien, N.S., Jabloun, C.M.: Profile of morbidity among elderly at home health care service in Southern Saudi Arabia. J. Family Community Med. 20(1), 53– 57 (2013)
- 17. Al Saif, A., Waly, E., Alsenany, S.: The prediction of falls among older people in Saudi Arabia. J. Am. Sci. 2012 **8**(6) (2012)
- Alsulami, M.H., Atkins, A.S.: Factors influencing ageing population for adopting ambient assisted living technologies in the Kingdom of Saudi Arabia. Ageing Int. 41(3), 227–239 (2016). Mohammed
- Al-Doghether, H.: Prescribing in primary care for the older people. Saudi Med. J. 25(4), 488–492 (2004)

Continuous Improvement of Strategic Alignment Model

Akazzou Salaheddine $^{(\mathbb{K})}$ and Cherti Ilias

Department of Mathematics and Computers Science, Faculty of Sciences and Technologies (FST), University Hassan 1, Settat, Morocco salaheddine.akazzou@gmail.com

Abstract. Good management information system (IS) is the key of success of good IT governance and this can only be achieved if all strategic goals of the company are respected. That's why, it will be necessary to ensure alignment with general objectives, to respond to the business model of the company and to define and anticipate the orientations of technical and economic choices.

In this context, strategic alignment is applied to two elements: Business strategy and IT strategy and allows them to be linked [1]. However academics are more interested about the part concerning the strategic alignment [2, 3] and neglect the ideal conditions of its operation and continuous improvement.

This works presents a model gathering between the concepts of strategic alignment presented by the Strategic alignment model (SAM) and principles of lean management presented by PDCA method.

Keywords: Information system · Strategic alignment model · Lean management · IT strategy · PDCA method

1 Introduction

The contribution of Information Systems department in organizations is no longer viewed just a support function but also a major asset taking part to a successful strategy and the financial bottom line when properly aligned.

Our aim was to provide a more concrete and reliable model as guidelines for transforming organizations into a competitive and lean environment, that's why improvements should be made with an awareness of the effect that these changes will have on other aspects of the organization.

In this context, strategic alignment ensures that the operating elements of the company all work in harmony and that will allow organization to use information technology efficiently to achieve its business objectives. And in order to optimize evaluate and sustain improvements to operational performance we suggested to integrate the PDCA cycle during the implementation of SAM

2 Lean IT and Transformation in Organizations

The application of Lean principles, includes continuous improvement of the system, requires profound changes in organizations.

To solve problems on the ground, organization should involve both operational and managerial staff, that's the first major change, however change is perceived in large organizations as something "top-down", Where little autonomy is left for the operational ones in the taking of initiative.

In IT organizations (Information Systems department), deploying a lean approach will also require involvement of multidisciplinary teams (from user demand to daily operation of the service), this is the only model able to solve problems in short cycles [4]. This type of organization is in breach with the specialized organization of work (collection of needs, specifications, development, integration, qualification, exploitation) which is standard in the management information systems.

2.1 The Sustainable Lean Iceberg Model

In order to go Lean and stay lean, organizations should continually understand the competitive marketplace in which it lives and the needs of its customers and what they value.

To remain focused on these needs organization must define its value streams, its processes and also its supply chain so as to detect its different failures and wasteful activities which could block any positive change in strategies and harm a perfect strategic alignment affecting all of its departments.

Next organization has to find ways of:

- setting the direction
- fixing targets
- Seeing whether or not change is actually occurring.
- Seeing whether or not change is actually occurring.

It is gainful to consider the Lean process as an iceberg "Fig. 1,".The technology, tools and techniques that affect processes are those visible above the water. [5]

Nevertheless the most important part of the iceberg is under the surface and invisible. It is this part that makes the iceberg strong and heavy.

Managing all these elements of the iceberg is essential to ensure a successful, sustainable transformation. However, this constitutes only part of the lean maturity of which the team that will set up this lean must have it.

The dependency between all the components of the iceberg "above the water" and "below the water" forms a balance of this iceberg.

So effective strategy and alignment can only be implemented if there is good leadership, which in turn can only be successfully achieved in a positive organizational culture that accepts any form of learning and improvement.



Fig. 1. The sustainable lean iceberg model

2.2 Lean Tools and Techniques

As we mentioned the technology, tools and techniques "Fig. 2," that affect are those visible above the water and the fifth element of the sustainable Iceberg Lean [5].

Section	Lean tools and techniques	
Strategy and alignment	Policy deployment / Hoshin Kanri A3 planning and storyboards Catchball PDCA Visual management	
Leadership	Lean leadership	
Behaviours and engagement	7 Lean skills Team cultures Lean coaches Continuous improvement	
Process management	Big Picture mapping Four Fields mapping Pull systems Voice of the Customer insight tool	

Fig. 2. Lean tools and techniques

Those tools must be guided by the needs of the client, the companies and the people within the organization, they should be pulled, not pushed.

These different techniques can be applied to each section of the Lean Iceberg. It will be necessary to begin by inspecting these Lean tools and techniques and then examining the role of the technology before deploying them in the environment that we are trying to make it Lean.

In order to realize that, it's primordial to start by analyzing each section of the organization separately according to the sections of Lean iceberg.

In general, a deployment is planned and coordinated according to the company's strategy. It begins locally on pilot sites before generalizing it.

A Lean deployment is an approach giving priority to its tools and methods, it's not necessarily uniform, some groups let their subsidiaries making choice of details but impose the reference framework and a toolbox to unify and gather new standards to which all the entities must comply with.

2.3 The PDCA Method

The PDCA (plan-do-check-act), sometimes seen as (plan-do-check-adjust) is a repetitive four-step approach for continuous improvement in business process management.

The PDCA model is also known as the Deming circle/cycle, or plan-do-study-act (PDSA).

This approach is implemented to test various solutions to a problem to identify the most effective solution before implementation.

PDCA was popularized by Dr. W. Edwards Deming, an American engineer, statistician and management consultant. It can be used by any department, from supply chain to finance to IT department.

3 Lean Concept and Alignment Approach

The lean concept is necessary for an efficient management of organization, which seeks to redesign its overall strategy and technological development strategy so that they become in perfect harmony [6].

This implies a coherence of the general strategy with the administrative infrastructure on the one hand and with the infrastructure of the applications on the other hand.

As we mentioned before in order to implement this alignment approach, lean management uses methods and techniques such as the PDCA method, which is the most important one and which is considered as a framework represented in a PDCA cycle (design, implement, monitor, review, and continually improve), This cycle came out of the quality and continuous improvement field it is also integrated into the monitoring and evaluation process. [7, 8] which are consolidated in lean thinking.

Thus the key principles of lean [9] relate to the PDSA cycle although possessing specific meaning to lean thinking, which means defining value and planning for the flow of value with as little waste as possible in order to achieve the perfect optimization.

The creation and integration of a process model is the final part of this conceptual model, This is achieved by aligning lean processes with business processes and this is the key to the proper use of strategic alignment to the approaches of continuous improvement and Lean.

Anywise it is not surprising that the strategic alignment approach matches with lean management, since both had roots in the quality and continuous improvement systems.

3.1 Strategic Management Alignment

The word Strategic alignment "Fig. 3," has been studied by different academics and practitioners that's why several approaches were developed [10].



Fig. 3. Strategic alignment model, four domains of strategic choice

Among these approaches, Strategic Alignment Model (SAM) (Henderson and Venkatraman) [11] is the most popular and appreciated by academics.

The main features of this approach are:

- 1. To provide operational guidelines for achieving strategic alignment.
- 2. To distinguish the external perspective of information technology (IT strategy) from its internal development (IT infrastructure and process).

So, SAM changes the traditional role of IT from a support role of organization's activities to a more strategic role, and that's enables it to cover not only strategic alignment to its business plan, but also provides the tools which allow alignment extended to the environment.

3.2 Structure of Strategic Alignment Model

The SAM model is structured into three classes of different elements:

- Domains: Business and Information Technology (IT)

- Perspectives or Levels (which subdivide each domain): external strategy and internal structure.
- Components which structure and characterize each level:
 - Infrastructure, skills and knowledge, and processes for the internal level
 - Perimeter, skills and governance for the external level.

Thus, the Business domain is formed of two levels of strategic choices:

- Competitive strategy or Business Strategy, in the external level, concerning how to make decisions about the products and the position of the company in the market.
- Organizational structure and Processes in the internal domain.
- In the same way the IT domain is described by two levels.
- IT strategy, in the IT external level, concerning the technological perimeter, distinctive technological competencies and strategic technological alliances (technological governance).
- IT infrastructure and process which refers to the technical architecture of Information system design Processes, evolution, monitoring and also management of technological knowledge and skills.

3.3 Building Blocks of Strategic Alignment Model

The alignment of an Information system is conceptualized in the Strategic Alignment Model by two Building Blocks

- Strategic fit: between the external level and the internal level of the same domain.
- Functional integration: Between the external or internal levels of different domains (Business and IT in this case).

For functional integration, the SAM determinate two types of integration:

- A strategic integration: it takes place between the IT strategy and the competitive strategy in order to establish the IT potential at a strategic level. These opportunities are fundamental because IT is now seen as an important source of competitive advantage.
- Operational integration: it takes place between the internal aspects of the business and IT domains, that means between 'organizational structure and business processes' and 'IT infrastructure and processes'.

4 PDCA Cycle Adapted to Strategic Alignment Model

The Strategic alignment model (SAM) has been widely exploited in research related to the strategic dimension of IT.

As part of our work we use Strategic alignment model (SAM) as a framework for analyzing IS (information system) alignment with Business, and in order to make it a viable model accepting continuous improvement we applied to it the PDCA approach. The conceptual elements of the SAM are interesting because they provide:

- Components to structure and formalize the domains which must be aligned
- Building Blocks (strategic fit and functional integration) to build alignment prospects which must be verified.

Realizing a complete IT alignment in organization, leads to plan, implement, evaluate and adjust in order to eliminate discrepancies and respond in the best way to expectations and integrate many uses, as IS users are numerous and varied in terms of their skills, experience and relation with its Information System.

In addition, it is necessary to evaluate the different technological possibilities in order to choose the appropriate components for the development of the IS and configure the organization's system.

The structure of sectors and departments of organizations can be reproduced from the SAM by analogy of concepts. And since, the original SAM makes a distinction between the external and internal levels, we propose the same for any sector having an external level (strategy) and an internal configuration (infrastructure and process), in a similar way, each level of an organization should be include three elements: scope, competencies and governance in the external level and infrastructure, skills and knowledge, and processes in the internal level.

To be sure that the components of this structure are adapted to the SAM model, We propose to integrate strategic fit and functional integration in a PDCA cycle and to repeat its 4 steps (Plan - Do - Check - Act) until the expected level is reached "Fig. 4,". Let's explain each step individually.

– Plan

The "plan" process establishes objectives, targets, controls, processes and procedures in the external part which will be deployed using strategic fit for the program to deliver results in accordance with an organization's overall business and IT strategy.

To make this strategy actionable, organization should in operational terms translate it into objectives and measureable targets in order to accomplish its mission and achieve its vision

– Do

The "do" process implements and operates the business continuity policy, controls and procedures. This includes a number of actions and internal functional integration in order to understand, strategize, plan, and test organization's infrastructure skills and knowledge, and processes, for business continuity events.

Check

The "check" process monitors and reviews performance against established management system objectives and policies and reports the results to management for review. The program should be subject to internal review in internal level to measure program performance against pre-defined policies and objectives in external level.

– Act

The "act" process corrects the defects and make it comply to the specifications maintains and improves the program by taking preventive actions for all the root causes identified and implement the preventive actions and check whether the outcome is as expected. This includes updating and maintaining the corrective and preventative actions list and a post-incident review process.



Fig. 4. PDCA cycle applied to SAM

5 Recommendations for Good Implementation of Our Model

So that our model can be well adopted by organizations we suggest a formula of recommendations and directives allowing the implementation of the reliable strategies in order to satisfy the operational requirements and meet the needs of the continuous growing business.

- Evaluate the maturity of the business continuity's management to understand the current maturity level of the plans of continuity and the way to reach the desired state
- Make of the continuity a pertinent easy and simple one to be understood by all the main actors of the organization constituting the executive committee of inter functional management.
- Evaluate the risk and analyze its impact on the business and plan a communication related to the intervention in case of problems
- Launch programs of testing, training and exercise in order to raise awareness of the organization's actors.
- Be sure that the objectives of corporate governance are realistic and that its steering committee can make available all the necessary resources allowing to reach them.

6 Conclusion

The objective of this work was to explore how lean management methods are supportive and applicable to Strategic alignment model. It determines also the effectiveness of strategy development, implementation, and subsequent competitive success. Also a strategic approach that is aligned with IT guarantees that an organization's employees, skills, and abilities contribute to the achievement of its business goals.

As we have shown that it is possible to integrate lean management and strategic alignment, It is also possible to integrate other methods of lean management like Value stream mapping which is a lean-management method for analyzing the current state and designing a future state for the series of events that take a product or service from its beginning through to the customer. At Toyota, it is known as "material and information flow mapping". [12] It can be applied to nearly any value chain.

References

- 1. LE CIGREF, Alignement stratégique du système d'information, Comment faire du système d'information un atout pour l'entreprise? (2002)
- Rahimi, F., Møller, C., Hvam, L.: Alignment Between Business Process Governance and IT Governance, Technical University of Denmark (2014)
- 3. Avila, O., Goepp, V., Kiefer, F.: Vers une extension du SAM (Strategic Alignment Model) pour les systèmes d'information de production (2008)
- 4. Poppendieck, M., Poppendieck, T.: Implementing Lean Software Development: From Concept to Cash. Pearson Education, London (2003)
- 5. Hines, P., Found, G., Griffiths, G., Harrison, R.: Staying Lean-Thriving. Not Just Surviving. Lean Enterprise Research Centre, Cardiff (2008)
- ISO/DIS 31000, ISO/DIS 31000: 2009-Risk Management—Principles and Guidelines on Implementation, International Organization for Standardization, Geneva, Switzerland (2009)
- 7. Moen, R., Norman, C.: Evolution of the PDCA Cycle, Associates in Process
- Womack, J.P., Jones, D.T.: Lean Thinking: Banish Waste and Create Wealth in your Corporation, 1st edn. Productivity Press, New York (1996)
- 9. Deming, W.E.: Out of the Crisis. MIT Press, Cambridge (1986)
- Henderson, J., Venkatraman, N.: Strategic Alignment a model for organizational transformation via information technology, November 1990
- 11. Henderson and Venkatraman (1999), Papazoglou and van den Heuvel (2000), Scheer and Nuttgens (2000), Wegmann (2003)
- Rother, M., Shook, J.: Learning to See: Value-Stream Mapping to Create Value and Eliminate Muda. Lean Enterprise Institute, Brookline (1999)

A Comparative Simulation Study on the Performance of LDPC Codes and 3Dimensional Turbo Codes

Mensouri Mohammed^(⊠), Aaroud Abdessadek, and El Hore Ali

Department of Computer Science, Faculty of Sciences, El Jadida, Morocco mensourimohl@hotmail.com, a.aaroud@yahoo.fr, aelhore@gmail.com

Abstract. Low-density parity-check (LDPC) codes and convolutional Turbo codes are two of the most powerful error correcting codes that are widely used in modern communication systems. This paper provides an overview of the basic concepts employed in LDPC codes and convolutional Turbo codes with 3 dimensions, and compare the performance of these codes. A description of both classes of codes will be given. The LDPC codes and 3 dimensional turbo code are coupled with receive diversity techniques and are employed as the error correction scheme over Additive White Gaussian Channels (AWGN) by employing Binary Phase Shift Keying (BPSK) modulation scheme. The performance of newly obtained codes is evaluated in term of bit error rate (BER) for a given value of Eb/No.

Keywords: Channel coding \cdot Turbo code \cdot 3 dimensional turbo codes \cdot LDPC codes \cdot Iterative decoding \cdot AWGN channel \cdot BER

1 Introduction

In 1963, Gallager introduced a family of error correcting codes constructed from the matrix of low density parity, called LDPC (Low Density Parity Check Code) [1]. These were forgotten with time, but these codes are good ideas particularly relevant operator for the construction of good codes. Thus, Gallager uses random permutations between parity codes to construct an efficient code of low complexity that imitates the random coding. Since 1993 and the introduction of turbo codes, LDPC codes have been largely rediscovered and are used in various.

We in 1993 Berrou et al. [2] proposed a new class of convolution codes called turbo codes whose performance in terms of Bit Error Rate (BER) are close to the Shannon limit. The conventional turbo code is a parallel concatenation of two identical recursive systematic convolutional encoders separated by a pseudo-random interleaver. Most conventional turbo codes using 8-state constituent encoders suffer from a flattening around a frame error rate (FER) of 10^{-5} due to a poor minimum distance dmin. To improve the performance in terms of Bit Error Rate (BER), one could either design a better interleaver, use more powerful constituent encoders, or increase the dimension, i.e. the number of constituent encoders.

Modern wireless communication standards have already adopted these types of codes for channel coding applications. For example, the Turbo codes are used in the 3GPP Universal Mobile Telecommunications System (UMTS) [3] and its Long Term Evolution (LTE) [4] system. On the other hand, LDPC codes can be found in applications ranging from wireless Local/Metropolitan Area Networks LAN/MAN) (IEEE 802.11n [3] and 802.16e [6]) and high-speed wireless personal area networks (PAN) (IEEE 802.15.3c [7]) to Digital Video Broadcast (DVB-S2 [8]). Furthermore, these codes are currently being proposed for next generation cellular and mobile broadband systems.

This paper compares the error performance of LDPC codes with message passing decoding and 3 dimensional turbo-codes.

The remainder of the paper is organized as follows. Section 2 describes the overview of LDPC codes which include the basic representation and types of LDPC codes. In this section, describes the encoding method of LDPC codes and the various decoding schemes of LDPC codes. Section 3 presents the concepts of turbo code with 3 dimensions. Experimental results are given in Sect. 4 that compares the error performance of LDPC codes with Turbo codes with 3 dimensions. Finally the conclusion and further work is given in Sect. 5.

2 LDPC Codes

2.1 Construction of LDPC Codes

Kindly LDPC code of parameters (N, j, l) is a linear block code of length N such that the parity check matrix H has j'1' in column and l'1' in a row. The numbers j and l are very small compared to the length of code to provide a low density matrix.

Figure 1 shows the parity check matrix of an LDPC code. This matrix H has therefore Nj/l rows, i.e. code is constituted Nj/l parity equation. The rate of the code then verifies:

$$R \ge 1 - \frac{j}{l} \tag{1}$$



Fig. 1. Parity check matrix of an LDPC code (N, j, l).

The parity check matrix H, as presented by Gallager in his thesis [7], can be divided into j sub-matrix $H^1, ..., H^j$, each containing an only '1' by column. The first sub matrix H^1 is a kind of identity matrix in which each '1' would be replaced by l'1', and whose number of columns is therefore multiplied by a factor l. The j – 1 other sub matrix $H^2, ..., H^j$ are obtained by applying j – 1 random permutation $\pi 2, ..., \pi j$ on the columns of the sub-matrix H^1 . Thus the matrix of LDPC code with (N = 20, j = 3, l = 4) is given in Fig. 2.



Fig. 2. Parity check matrix of an LDPC code (20, 3, 4)

2.2 Representations of LDPC Codes

There are essentially two different possibilities to represent LDPC codes. Like all linear block codes, they can be described by matrices. The second possibility is a graphical representation. These two representations are completely equivalent, and we choose one or the other depending on the level of simplification it brings to resolve the problem.

Graphical Representation

The first representation of LDPC codes is a bipartite graph (see Fig. 3), also called Tanner graph [9]. A graph is bipartite if there are two sets of nodes U and V vertices and a set of edges such that each edge connects a node U to node V.

LDPC code can be defined from a Tanner graph whose set of nodes left, denoted V (variable nodes) represents the symbols of the codeword, and all right nodes, denoted C (constraint nodes) represents the constraints. A sequence of symbols then constitutes a valid code word if and only if for each node constraint, the sum of symbols corresponding to the variable nodes is zero. Figure 4 represents the Tanner graph of LDPC code.



Fig. 4. Tanner graph of LDPC code

For example, consider the following parity check matrix of an LDPC code with rate 1/2 and producing four redundancy bits:

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

The equations associated with that parity matrix and a codeword $x=(x_0,\,x_1,\ldots,x_7)$ are:

$$\begin{cases} x_0 + x_4 = 0\\ x_1 + x_4 + x_5 = 0\\ x_2 + x_5 + x_6 = 0\\ x_3 + x_6 + x_7 = 0 \end{cases}$$



Fig. 5. Tanner graph of LDPC code

Figure 5 shows the graph of the parity check matrix defined in the previous example.

Representation of LDPC code in the form of a Tanner graph is particularly useful when one wants to represent the message passing mechanism that intervenes in the iterative decoding. In addition, this representation allows using the powerful tools of the graph theory to study and design the LDPC codes.

Matrix Representation

Like any linear code, the parity check matrix H defines the code completely. H is the matrix of size $(M \times N)$ with M = N - K. More precisely the parity check matrix defines relations between the symbols of the code words in the form of a linear system:

$$Hx = 0 \tag{2}$$

When he introduced LDPC codes, Gallager specified that this matrix should have a low number of non-zero elements. The matrix low density then each relationship will concern a small number of symbols. In the case of a systematic code, each column corresponds to a symbol, the first K columns corresponding to the information symbols and last N - K columns for redundancy symbols. For example an LDPC code with N = 8 and K = 4 is defined by the parity matrix:

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

2.3 Classifications of LDPC Code

LDPC codes can be classified into two general categories: regular codes and Irregular codes.

A regular LDPC code is defined by a parity check matrix H which has the following properties:

- each row has *l* values 1
- each column has j values of 1
- the number of 1 in common between two columns, denoted by μ , is not greater than 1 (hence $\mu = 0$ or $\mu = 1$)

Numbers 1 and j have values smaller compared to the code length and the number of rows of the matrix H.

An irregular code has parity check matrix H with different weights of rows and columns. This means that in the Tanner graph, nodes encoded bits have multiple degrees and parity nodes. Irregular LDPC codes are most commonly constructed on the basis of their Tanner graphs according to the distribution A(x) degrees of variable nodes and the distribution B(x) control nodes, its error performance is strongly linked to the values of these two parameters. These distributions are usually represented by polynomials. The degree distributions of node the point of view are characterized by the polynomial (A, B) and have the following form:

$$A(x) = \sum_{n \ge 1} A_n x^n \tag{3}$$

$$B(x) = \sum_{n \ge 1} B_n x^n \tag{4}$$

Where A_n and B_n are respectively the fractions variable nodes and check nodes of degree n, corresponding respectively to those portions of the columns and rows of the parity check matrix having a degree equal to n.

This is also referred to as degree distribution of a point view edges. Polynomials are associated (λ, ρ) and have the following form:

$$\lambda(x) = \sum_{n \ge 1} \lambda_n x^n \tag{5}$$

$$\rho(x) = \sum_{n \ge 1} \rho_n x^n \tag{6}$$

Where λ_n and ρ_n are the fractions edges connected respectively to nodes symbols and control nodes of degree n. These two pairs of polynomials are related by the following relationships:

$$\lambda(x) = \frac{A'(x)}{A'(1)} \tag{7}$$

$$\lambda(x) = \frac{B'(x)}{B'(1)} \tag{8}$$
These polynomials also determining the rate R of code:

$$R = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx}$$
(9)

2.4 Encoding Processes for LDPC Codes

The introduction of systematic LDPC codes by Macky et al. [10] was motivated primarily by reducing the encoding complexity. The problem of encoding LDPC codes is determined codeword x from the vector b containing the information bits and the parity check matrix H defined by the following equation:

$$Hx = 0 \tag{10}$$

If G is a generator matrix of the LDPC code to encode we have the following relationship:

$$x = Gb \tag{11}$$

Note that this equation introduced in (10) we obtain the following relationship:

$$HG = 0 \tag{12}$$

The method that seems to be simpler for encoding is to build a matrix encoding H' upper triangular by applying the Gaussian elimination on the matrix H. As shown in Fig. 5. The x codeword is then cut into a systematic part containing the information bits, b_n with n = 1, ..., k and the part containing the redundant bits r_m with m = 1, ..., M determining. Then by identifying form the bottom up, it is easy to calculate the redundancy bits. However, the use of the Gauss pivot of the sparse matrix H only allows to obtain a semi-encoding matrix H' upper triangular, with important density '1' making the number of the operation to be performed when the encoding of the order of N^2 while their direct competitor, that is to say, the turbo codes have a linear complexity $\theta(N)$. This encoding step is the principal drawback in the practical implementation of LDPC codes (Fig. 6).

To reduce this complexity two approaches have been developed. The first is to constrain the matrix parity checking to reduce this complexity, while the second is based on the construction of a matrix encoding semi-triangular the low density as possible.

Regarding the first approach, we note the work of Spielman et al. [11, 12] which consist of concatenated LDPC codes with short length (sub-code) then the resulting code is incredible with linear complexity. However, given the size of each block (sub-code) is very small compared to the size of the resulting block, the performance of these types of codes are inferior to those obtained with a standard LDPC code of the same length. Another approach that was introduced by Macky et al. [10] is to force the parity check matrix to be symmetrical (either directly encoding) i.e. of the set of codes

27



Fig. 6. Encoding LDPC codes

is limited by the fact that the matrix H either the semi-triangular as described in Fig. 3. Given that the parity matrix is low density, this restriction ensures linear complexity. It generates, in general, some small losses in performance.

Methods of reducing the complexity of encoding in the form of the parity check matrix have been also developed in [13], Richardson and Urbanke use semi-triangular matrices for encoding. These matrices obtained only by simple permutation of rows and columns, which can retain ownership of the sparse matrix. The complexity of encoding depends on a parameter reflecting the difference g between the matrix resulting semi-triangular and a triangular matrix for which the complexity is linear. The second approach, introduced in [14] is to construct a triangular matrix encoding is as hollow as possible. This is effected by applying at the Gaussian elimination and the criterion Markovitez of which comprises a pivot about "1" on its entire line and column of less than "1" line. Resulting complexity is then $\theta(Nt_{rmoy})$ where t_{rmoy} is the number average the "1" by lines.

2.5 Decoding Processes for LDPC Codes

Compared to other types of codes, the decoding of LDPC codes do not pose many problems for researchers other than their construction. The most difficult work is to find the best methods to build efficient LDPC codes. LDPC code can be decoded by several methods like:

Maximum Likelihood Decoding

Decoding techniques by maximum likelihood or decoding ML ("Maximum Likelihood"). ML decoding returns the codeword closest (in the case of the BSC channel distance is expressed in terms of Hamming distance) to a given sequence of symbols. The code word returned is actually one that was most likely sent to the channel. When performing ML decoding, it exploits the correction capabilities of code. However, this optimality is at a cost in complexity, which may be prohibitive for long.

Iterative Decoding of LDPC Codes

LDPC codes can be decoded iteratively by the sum of product algorithm, called the message passing algorithm the LDPC codes. The MAP decoding algorithm achieves performance, provided that the code graph is cycle free. Although it is closer to the algorithms decision decoding based on the work of Gallager [7]. The sum-product algorithm [15] is applied to the code graph where the check nodes and variable nodes exchange messages which mainly extrinsic reliability values associated with each code symbol. One can find other algorithms for decoding LDPC codes [16].

3 3Dimensional Turbo-Code

Generally, a Turbo code with 2 dimensions is composed of a concatenation a series or in parallel [17] of two codes (C₁, C₂), that is called code components, and an interleaver π (see Fig. 7). While the first code component encodes the information in the original order, the second receives the information in a permuted order. In all norms, the convolutional codes are used as component codes. For more information about turbo-code with 2 dimensions and turbo-series, refer to [8].



Fig. 7. Structure of parallel turbo codes with 2 dimensions

3.1 Coding Structure of 3 Dimensional Turbo Code

In the future, most of digital transmission systems require low error rates that can go up to 10^{-8} dB [18]. Improving performance at very low error rates by raising the minimum hamming distance may involve using component encoders with a larger number of states, devising more appropriate internal permutations, or increasing the dimension of the turbo code, i.e. the number of component encoders [19]. In this work, we are interested in the third case: increasing the dimension of the turbo code.

Figure 8 shows TC-3D using the three components codes C_1 , C_2 and C_3 in parallel that will be used throughout this article to illustrate some fundamental concepts. The three codes are recursive convolutional in nature [20], with a constraint length L = 3 (i.e. memory = 2). The overall code is a rate to equal $\frac{1}{4}$ code with four output streams. One of the output streams is the information sequence u uncoded. The other three output streams X_1 , X_2 and X_3 in this example are parity sequences corresponding to



Fig. 8. 3 Dimensional turbo code constitute from three codes C_1 , C_2 and C_3 which are grouped in parallels.

three codes C_1 , C_2 and C_3 . These three parity streams would be identical if permutations $\pi 1$, $\pi 2$ were not used.

The sequence u information of length K bits is encoded by TC-3D. This code is realized by the parallel concatenation of three identical coders: $C_1 = C_2 = C_3$. They are recursive convolutional in nature, with eight states and the generator polynomial is:

$$G1 = [1 \ 1 \ 1]$$

 $G2 = [1 \ 0 \ 1]$

The information bits u are encoded initially by the encoder C_1 to provide the first redundancy bits X_1 , then they are interleaved by $\pi 1$ before being encoded by the encoder C_2 and delivering the second redundancy X_2 . The message u is then interleaved by $\pi 2$ before being encoded by the encoder C_3 and delivering the third redundancy X_3 .

For the canal entrance, we send a message u and data redundancy X_1 , X_2 and X_3 generate with three elementary codes C_1 , C_2 and C_3 . The originality of the encoder is performing an interleaving ($\pi 1$; $\pi 2$) on the data u before treatment C_2 and C_3 encoders so that errors are not corrected by the first encoder will generally be different the errors not corrected by the second and third codes. For more information about TC-3D refer to [21].

Finally, the information sequence u and the code sequences X_1 , X_2 and X_3 are multiplexed to form the codeword u^{Ch} of length N bits, they are transmitted to the channel. Note that the total code rate of the TC-3D is $R = \frac{K}{N}$, and we can write:

$$\mathbf{u}^{\rm Ch} = (\mathbf{u} \ \mathbf{X}_1 \, \mathbf{X}_2 \, \mathbf{X}_3) \tag{13}$$

3.2 Decoding of 3 Dimensional Turbo-Code

In the AWGN channel, the binary information sequence $u^{Ch} = (u \ X_1 \ X_2 \ X_3)$ N-dimensional bits as input to the channel, produces a sequence $u^{Ch'} = (u' \ X'_1 \ X'_2 \ X'_3)$ the N dimensional bits, the relationship between u^{Ch} and $u^{Ch'}$ for the AWGN channel is:

$$\mathbf{u}^{\mathbf{C}\mathbf{h}\prime} = \mathbf{u}^{\mathbf{c}\mathbf{h}} + \mathbf{B} \tag{14}$$

31

Where B is a random sequence representing the "noise" or "error" additive.

The 3 dimensional turbo code can be decoded using the turbo principle. The decoder of 3 dimensional turbo code consists of three soft-input soft-output (SISO) [22] decoders 1, 2, and 3 corresponding to the three constituent encoders C_1 , C_2 , and C_3 , respectively. A decoding iteration consists of a single activation of SISO decoder 1, SISO decoder 3 and SISO decoder 2, in this order. This process continues iteratively until the maximum number of iterations is reached or an early stopping rule criterion is fulfilled.

These three decoders exchange extrinsic information on the systematic bits u' and the parity bits X'_1 , X'_2 and X'_3 This extrinsic information exchange is referred as turbo principle. Figure 9 shows a possible decoder realization. The three component decoders perform a maximum a posteriori probability (MAP) decoding on bit. They use the BCJR [23] algorithm transformed in the logarithmic domain, the so called Log MAP algorithm [22], to decrease the implementation complexity.



Fig. 9. Schema of 3Dimensinal turbo decoder.

4 Simulation

The comparison of LDPC codes and Turbo codes with 3 dimensions in terms of performance will be given in this section. In order to give a fair comparison of the codes, we use codes of the same input word length when comparing. The rate of both codes is R = 1/2. All simulations are performed assuming code words are transmitted over an additive white Gaussian noise (AWGN) channel with zero mean and variance $N_0/2$ via binary phase shift keying (BPSK) signaling.

Figure 10 shows the error performance of the LDPC code with the message passing decoding as a function of the E_b/N_0 .



Fig. 10. Performance of LDPC

The performance of the 3dimesional Turbo Code was assessed by means of simulation. In Fig. 11 we report frame error rate results for typical block sizes, K = 1024 bits. In all simulations, a maximum of 6 iterations are assumed. The component decoding algorithm is the simple Max-Log-MAP algorithm.

Figures 12 show the performance of turbo code with 3 dimensions and the LDPC codes with information length 1784. The turbo codes with 3 dimensions defined in [20] were used for obtaining the performance curves. The turbo-code with 3 dimensions is better until a certain S/N is reached, respectively Eb/N0 = 1 and 4 dB. This characteristic 'error floor' was also typical of turbo-codes in earlier years, but this has become less of a problem lately due to the definition of good permutation matrices.



Fig. 11. Performance, in Bit Error Rate, of the 3dimesional turbo code



Fig. 12. Performance comparisons between of LDPC codes and Turbo codes with 3 dimensions.

5 Conclusion

In this paper, we compared the performances of a turbo-code with 3 dimensions and codes LDPC (Fig. 12). We presented in the first time an introduction on the LDPC codes and the various methods of coding/decoding of this kind of codes. The code LDPC has a remarkable performance on AWGN channels with the iterative decoding SPA based on the propagation of confidence. Then, we described the coding and the decoding of the turbo-code with 3 dimensions. We also studied with the simulation, the performances of a turbo-code in 3 dimensions which produced a performance very good in term of the binary rate of errors by contribution turbo-code with 2 dimensions (Fig. 12).

References

- 1. Gallager, R.G.: Low Density Parity Check Code. MIT Press, Cambridge (1963)
- Berrou, C., Glavieux, A., Thitimajshima, P.: Near Shannon LimitError-correcting coding and decoding: turbo codes. In: IEEE International Conference on Communication, pp. 1064– 1070, May 1993
- 3. 3GPP UMTS: General UMTS Architecture. 3GPP TS 23.101 version 7.0.0 (2007)
- 4. 3GPP LTE: Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved UniversaTerrestrial Radio Access Network (EUTRAN). 3GPP TS 36.300 (2008)
- 5. IEEE-802.11n: Wireless LAN Medium Access Control and Physical Layer Specifications: Enhancements for Higher Throughput. P802.11n-2009, October 2009
- IEEE-802.16e: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. P802.16e-2005, October 2005
- IEEE-802.15.3c: Amendment 2: Millimeter-wave-based Alternative Physical Layer Extension. 802.15.3c-2009 (2009)
- ETSI DVB-S2: Digital video broadcasting, second generation. ETSIEN 302307, vol. 1.1.1 (2005)
- 9. Tanner, R.M.: A recursive approach to low complexity codes. IEEE Trans. Inform. Theory 27(5), 533–547 (1981)
- Mackay, D., Wilson, S.T., Davey, M.C.: Compression of constructions irregular codes. IEEE Trans. Commun. 47(10), 1449–1454 (1999)
- Spielman, D.: Linear-time encodable and decodable error-correcting code. IEEE Trans. Inf. Theory 42(6), 1723–1731 (1969)
- Luby, M., Mitzenmacher, M., Shokrollahi, A., Spielman, D., Stemann, V.: Partical loss resilient codes. In: Proceedings of the 30th Annual ACM Symposium on Theory of Computing, pp. 150–159 (1997)
- Richardson, T.J., Urbanke, R.L.: Efficient encoding of low-density parity check code. IEEE Trans. Inf. Theory 47, 638–656 (2001)
- Debaynast, A., Declercq, D.: Gallager codes for multiple access. In: Proceedings. of the ISIT 02 Symposium, Lausanne, Switzerland, July 2002
- 15. Amador, E.: Aspects des Décodeurs LDPC Optimisés pour la Basse Consommation. thèse, Université TELECOM ParisTech, Soutenue le 31 Mars 2011
- 16. Gorgoglione, M.: Analyse et construction de codes LDPC non-binaires pour des canaux à évanouissement. Thèse, Université de Cergy Pontoise, Soutenue le 25 October 2012

- Boutros, J.J.: Les turbo codes parallèle séries décodage SISO itératif et performances ML, October 1998
- Rosnes, E., Graell, A., Amat, I.: Performance analysis of 3-dimensional turbo codes. IEEE Trans. Inf. Theory 57(6), 3707–3720 (2011)
- Berrou, C., Graell i Amat, A., Ould-Cheikh-Mouhamedou, Y., Douillard, C., Saouter, Y.: Adding a rate-1 third dimension to turbo codes. In: Proceedings of the IEEE Information Theory Workshop, Lake Tahoe, CA, pp. 156–161, September 2007
- Aaroud, A., Mensouri, M.: Performance analysis of 3-dimensional turbo codes. Int. J. Comput. Inf. Technol. (IJCIT) 4(1), 17–24 (2012)
- Mensouri, M., Aaroud, A.: Weight distribution and bounds of turbo-code with 3 dimensions. SIJ Trans. Comput. Netw. Commun. Eng. (CNCE) 1(1), 18–23 (2013). The Standard International Journals (The SIJ)
- Robertson, P., Villebrun, E., Hoeher, P.: A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log-domain. In: Proceedings of the 1995. International Conference on Communications (ICC 1995), Seattle, Washington, USA, pp. 1009–1013, June 1995
- 23. Bahl, L., Cocke, J., Jelinek, F., Raviv, J.: Optimal decoding of linear codes for minimizing symbol error rate. IEEE Trans. Inf. Theory **20**, 284–287 (1974)

Energy Efficiency Approach for Smart Building in Islanding Mode Based on Distributed Energy Resources

Youssef Hamdaoui^(IM) and Abdelilah Maach

Department of Computer Science, Mohammedia Engineering School (EMI), Mohammed V University, Rabat, Morocco youssefhamdaoui@research.emi.ac.ma, maach@emi.ac.ma

Abstract. Smart buildings represent a prototypical cyber-physical system with deeply coupled embedded sensing and networked information processing that has increasingly become part of our daily lives. The notion of cyber physical is very important because the cyber word collects information from the various physical parameters through sensors and smart devices and on the basis of this information the main controller makes decisions and an energy distribution plan during outage period. Through context-aware sensing and computation the cyber-physical systems will be able to acquire contextual awareness information from the physical world and use this information in an intelligent way to build an effective energy management system based on smart distribution of power produced from renewable resources. The proposed approach allow a dynamic selection based on real time consumption measurement and information collected by different controllers installed in different entities inside building to decide the entities that will be powered during outage, the idea presents a way for optimization of restored energy and a intelligent strategy to response to customer demands based on their classification and their needing priority. An algorithm is developed to implement our approach with a simulation in Java environment to validate our efficiency management system through on-site distributed generators within buildings to operate in islanding mode. Finally, The objective of this paper is to propose a solution for new building generation that represent future smart cities concept and to have an self-supply based on distributed generation without needing to be connected to the main power grid.

Keywords: Smart building · Efficiency management system · Cyber physical system · Outage · Islanding · Distributed energy resources · Smart distribution

1 Introduction

The fast development of urban area mostly in the economic cities, Growth in energy demand, integration of distributed renewable resource (DER) [1], renewable energy based generation, distributed storage and controllable loads such as Electric Vehicles (EV), as alternative resources, evolution of IT technologies that allow a fast communication and real measurements, context-aware sensing and computing and especially the high reliability between physical world and virtual world and its impact on human process and aspects, all these factors forces us to rethink our smart grid concept in our

building and presents a new several challenge to ensure a trade-off between energy consumption and occupant's comfort and demand. The objectives is to respond to the growing demands, economic growth, security, quality of life of habitants and increased quality of life specially in buildings and generally in cities, thus to make them smarter. Cyber-physical systems can collects the awareness information from the various physical parameters through sensors and process this information in the cyber-world and control actuators, it is efficient to manage, control, monitor, and query on machinery, equipment, and personnel state, a smart building is such a good scope of application.

This paper looks modern buildings as a cyber-physical energy system with deeply coupled embedded sensing and networked information processing because through context-aware sensing and computation the cyber-physical systems (CPS) will be able to acquire contextual awareness information from the physical world and use this information in the cyber-world and in an intelligent way to take the right decision and ensure the best energy distribution management by priority and level of emergency. Integrating context awareness into the system increases the efficiency in terms of energy savings and was observed to be significant, around 30% and these systems will be able to anticipate needs and situations and react to the environment around them. In this paper, we present an autonomous distribution approach based context aware parameters and entities classification for smart buildings (SB) through cyber physical system models focusing on diverse requirement satisfaction, including response time, energy efficiency, real time sensor measurement, continuous supply, reliability and users' needs. In this sense, the smart energy management system presented considers an smart distribution platform that gathers real-time data from a physical sensors and meters [2] connected to a house controller connected to the cyber physical system manager (CPSM) to determine appropriate control actions on individual building appliances, in order to satisfy comfort conditions of houses and saving energy. The communication between building component is supposed done with a combination of wireless and wired network.

This paper provides an approach for smart buildings with the focus on energy saving in islanding case. We consider smart buildings as cyber-physical systems and describes the most able technologies and theories and their impacts to arrive a best way of energy management system. The rest of the paper is organized as follows: Sect. 2 present a literature of deferent concepts used by reviewing recent related works, describes our proposal architecture and concept for intelligent distribution based building context aware through cyber physical system (CPS), Sect. 4 show our scheme Model Algorithm description, case study and a simulation based on dynamic parameters stored in database and Finally we conclude with discussion of results and decision generated for outage case studied.

2 Literature and Definitions

2.1 Cyber-Physical System

The CPS is the main controller and consists of hardware and software and is an intelligent, programmable environment capable of performing metering, computations, numerical processing, running optimization subroutines, establishing bi- directional communication with the smart grid Control centre and building controllers to make decisions based on the specified real time constraints. It would also have direct control capability of the electrical appliances. CPS is a system which interacts also with humans. It extends the physical world trough computing power, by processing information received from different sensors in the physical world and affects it by controlling actuators [3]. An actuator for example can be used to open a window or to increase the temperature of the heating. The distributed components of a CPS are connected each other, to process information or execute tasks because a cyber-physical system uses information from the physical world to process and control different actuators, For example when we are in rain case we can detect rain with sensor and the CPS make a decision base on this context aware information to close all windows in building through controllers and actuators. Table 1 present some important properties for an autonomic CPS [4].

Property	Description
Self-adaptation	To react and adapt to the continuously changing context
Self-organization	In order to provide seamless data exchange throughout highly heterogeneous networks, the CPS have to be able to reorganize itself against this evolving topology
Self-optimization	Optimal usage of the constrained of CPS devices is necessary for sustainable CPS deployments
Self-configuration	To avoid hand configuration of different controllers, sensors and actuators
Self-protection	Due to its wireless and ubiquitous nature, CPS will be vulnerable to numerous malicious attacks
Self-healing	To detect and diagnose problems as they occur and to immediately fix them in an autonomous way
Self-description	Devices and resources should be able to describe their characteristics in order to allow other communicating objects to interact with them.
Self-discovery	CPS devices/services should be dynamically discovered and used by the others in a seamless and transparent way
Self-energy-supplying	To realize and deploy sustainable CPS energy management solutions

Table 1. CPS properties

2.2 Smart Building Concept

Smart Building is new concept which accompanies the cities evolution in renewable energy integration and the IT evolution. Smart building has a lot of benefits for both countries in environment context and to avoid power production/importation because buildings account for 30–40% of total energy consumption [5], and also for citizens to produce his own power needing and consequently reduce costs. In the future, most of buildings will be an independent building with an auto management system that control the distribution of power in all building with a best management that take in consideration all building context aware real time parameters before make decisions through a

CPS [6]. Research in the field of home and building energy management Systems is discussed in [7, 8].

2.3 Smart House Based on Smart Component

The power aware smart house is in interaction with internal and external environments. The external environment consists of all the entities belonging to the Smart Grid. The internal consists of all appliances and devices belonging to the smart home, which are centrally managed by an smart controller [9]. Controller is a smart sophisticated component which is responsible for execution actions making by the CPS and sending different state of house appliances and real time sensors/meters measurements. Controllers ensure information exchange between the CPS (Main controller) and all sensors and meters belonging the smart house [10]. Smart sensors, meters or actuators are used to collect data about the power of current consumed by each appliance, temperature and the light of the environment to the overall load profile of the house and transmit it to the user and controller as shown in Fig. 1. The purpose of the smart plug is not only to send information, but to also receive commands and execute them through actuators. Simple home become an interactive smart home with a storm of information and measurements, this information is important to control and optimize the energy consumed and make a decisive actions to secure house from dangers, create a easy live for humans and smart selflearning space based on daily humans activities A very new concepts is done thanks to this small smart device that can be used in different context [10].



Fig. 1. Smart house with smart components

2.4 Context Aware Definition

The ability to take into account the digital and physical environment and the context of the user makes a space smart. In this work, Context is one solution for building energy efficiency that to consider Context-Aware aspect in energy control operations which allow adapting to the context and catering to highly dynamic environments, environment parameters means any information used to describe indoor environment of a space. A space that contain some controllers, sensors and devices which allow collecting information about user's activities [11] and give a real time measurement. Context-aware systems must have a promising way to automatically model and represent a user's context to help end users obtain their desired services [12].

2.5 Microgrid

Micro-grids are integrated energy systems consisting of inter connected loads and distributed energy resources which as a system can operate in parallel with the grid or in an intentional island mode. The Micro-grid (MG) is considered to be a distribution grid incorporating local generation, storage devices and responsive loads [13, 14].

Through the future change of paradigm in the mobility sector, Electric Vehicles (EV) will become another resource to be integrated in the MG system, which can behave either as flexible loads or as mobile energy storage devices [13, 14]. In order to be flexible and controllable, the MG power infrastructure is supported by a communication and information system constituting the MG technical management and control system [13–15]. The local intelligence is responsible for coordinating and controlling local resources and enable innovative self-healing operating strategies.

2.6 Islanding

Islanding is a condition in which a portion of the power grid, which contains both load and DG is isolated from the remainder of the utility system resulting from extreme weather or other emergency situations or When the demand in the peak hours exceeds the supply. In this case some studies are done to present the potential of Distributed energy resources (DER) to response emergency demand and have a dynamic power balancing based on some parameters [16, 17].

Islanding detection Methods (IDM) advantages and disadvantages are presented in Table 2. Generally, all islanding techniques detect the absence of the main generation and stop automatically production through measurement the voltage or frequency parameters on the micro grid side. However, when the DER generation and loads within the island segment are well balanced it is difficult to detect the utility absence [15]. Islanding must be detected in the instant at which the grid is cut off from the utility the outage in order for the DG system to change between grid-connected and intentional-islanding modes.IDM are generally divided into local and remote methods [18, 19] and we can have a Hybrid detection methods consist of combination of different IDMs.

2.7 Hybrid Distributed Energy Renewable

The majority of studies confirm that using a single renewable energy generation, such as solar, wind, geothermal and hydropower generation is difficult to provide a stable continuous power supply all the time, for example PV systems have two big problems that the efficiency of electric-power generation is very low and depends on weather conditions. Wind turbines are also depending on the weather conditions and also affected by system fault at t = 52,000 s that means power output of wind generation is reduced

to zero. Consequently, the hybrid energy system with a DER combination can be effective solution to unstable effects of electricity supply. There are various combinations for hybrid renewable energy, such as wind/PV, PV/biomass, wind/hydropower, wind/PV/ biomass, etc. A hybrid solar-wind-battery system can provide 100% of power supply for consumers, thus greatly decreasing the energy costs and increasing the continuity and reliability of power supply [20]. The hybrid combination used in our study case and simulation is solar PV and batteries.

3 Smart Building Concept

3.1 Smart CPS Architecture

The Cyber-Physical Energy Systems represent a new class of widely distributed and globally interconnected energy systems that integrate computation processes, communication processes and control processes [21]. The CPS layered architecture should include an information-centric protocol stack to support data fusion for making the data into the network and converting to high-level information for applications as shown in Fig. 2 [21].



Fig. 2. CPS architecture

The CPS Model consists to do an iterative 4 actions: Monitor, Analyze, Plan, and Execute. The objective is to continue monitoring different controllers, sensors or actuators status and collect information, second step is to analyze all context aware parameters, the analysis should be done in quasi real-time, Third action is to gives decisions

for actions to take in order to attain the high-level objectives defined CPS manager and the last action is to schedules and executes the decided actions on the managed element. They mainly consist of re-configuration of building appliances in order to obtain the state desired by the CPS.

3.2 Smart Building Based DER

In this section we present a smart building architecture based on some DER and powered also by utility grid to response to loads in the case the energy needing is not covered by the energy produced by solar PV and by the Storage batteries. Smart building contains some DER in the floor, some electrical vehicles (EV), HVAC, and some comfortable habitant needs to have a best management through the CPs Manager In this paper, we suppose a residential building that contains as combined DER composed of PV and Storage to power building and some EVs that do a daily travail (50 miles/day) and some apartments with residential appliances.

3.3 Smart Home Architecture Based Controllers and Sensors

Smart house contain some smart components like sensors, controllers and actuators. All sensors and actuators are connected each other's and also connected to the main house controller which is in communication with the CPS manager to transmit information, all components state and also execute all actions generated by the CPS manager.



Fig. 3. Smart building model

4 Methodology

4.1 Building Scheme Model

We concept a new model for smart building connected to the main electrical grid that contain a limited numbers of smart house as shown in Fig. 3. SB have a outside controller C0 connected to the CPSM with a smart meter and smart switcher to decide using main grid power of not, the decision is done by the CP and depends on the level of the batteries charge, the production of solar PV and the building loads. Different controllers (C) are installed front of each house's door, each house controller is connected to sensors (S) and actuators (A) inside the house to collect information and execute actions planned by the CPSM. All information collected from different controllers is transferred to CPS to manage, analyze, plan and Wireless Communication networks (Wi-Fi, Bluetooth, and Zigbee), play an important role for continuously and seamlessly monitoring the building energy use, which gather the information on user behavior and its interaction with appliances from the building environment. Zigbee is the best option for a reliable communication between the smart plugs and the gateway in the home management system. Indeed, it allows a communication that offers a larger range of communication approximately 100 feet than Bluetooth. Moreover, Zigbee is low cost, low power and easier to implement [22] send actions for a smart energy distribution based on real time state. Figure 3 present the model with different controllers, sensors, actuators installed in the SB, on the floor we see the DER installed connected to a CPSM level and CO present the main controller that contains a switch to use utility power if we are in the case of needing energy and not use the utility energy if we have enough power to cover all loads, obligatory loads and facultative loads because we have a classification of each demand in building. The classification mean the level of emergency because we should start with houses with high emergency like health needing and let at the end the component with low level of emergency. Electrical power sources are available on a scale (kW range) that makes them suitable for on-site generation in buildings.

5 Smart Distribution Approach

5.1 Smart Distribution Based on Smart Selection Approach

The idea in our approach is to define the critical houses that need a important energy continually without interruption and define it as special node on the building grid, like habitants with a health difficulty, the idea is when the main grid is down or if we want to use just the energy powered by local DER, the CPSM have to consider the emergency process, CPS collect data from a deferent controllers installed in the door of each house to collect sensors information and based on this information, Selection algorithm have to define the islanded houses that can't be covered by our selection and the houses that we can response to his load. The advantage in our method is that it is dynamic and not static and can change each outage case because the context and state of the energy produced or stored change and not static. We support that each entities have a bus connected that collect the house power state and can switch from in energy use or out

energy use; the decision is to CPS manager to take and not the habitants. The smart selection method and smart distribution is based on the deferent small sources DGs installed in the building considered as small micro grid. All house loads is devised between obligatory and facultative load.

The Distribution network contains different nodes as explain in the last paragraph and each node has his weight into the building grid. The selection construction model consist on the state of some parameters that we see more important and decisive in the selection construction model, there is others parameters like house maturity, environment, weather context, etc [23]. But in this paper we limit our study to three parameters.

- Costumer Category (C).
- Costumer emergency Demand (P).
- Costumer facultative demand (Q).

The dynamic selection consists to start with the more emergency and pass to the medium and the last house with low level of priority. The classification can change because the level of priority and the costumer need is dynamic It's an advantage that we work on to have an adaptive algorithm to the context and the off grid area. We will have an event islanding processing lunched by the CPSM each time we have new important information.

5.2 Import/Export Energy

Given the variable nature of renewable energy resources, including solar, energy storage is a necessary component for a distributed PV system to provide reliable power and batteries are the most commonly used and well-suited storage technology for residential building.

Battery storage systems are often provided with a power rating in kilowatts (kW). Storage batteries for a grid connected solar PV storage system are typically around 1 kW to 7 kW. This is the capability of the battery to provide power. A battery's stated electricity capacity, as expressed in kilowatt-hours (kWh)1 is generally larger than the battery's actual useable capacity, because: • all batteries lose some energy in charging and discharging, though some have better 'charge-discharge efficiency' than others. • Most batteries are not designed to be routinely fully discharged (can reduce battery life). Some have deeper discharge capability than others.

A building can have sometimes an extra power and sometimes can have a need of power to response to local demand. We cited some CPS:

- Case 1: Not using all the electricity that the solar PV system is generating, then the system will ensure that any surplus energy is used to charge the battery.
- Case 2: When the batteries are full charged then the system will ensure that any surplus energy is used to charge the EV batteries (Vehicle–to-grid technology).
- Case 3: When the parked EV batteries are full charged then the system will ensure that any surplus energy is exported to the grid.
- Case 4: In the evening or at time of low solar generation, the solar PV panels have a reduced or zero output and once the batteries are discharged then system can

use energy stored in EV parked in the basement building (Vehicle-to-building technologies) [24, 25].

Case 5: if we need to use more electricity, then the system is in the obligation to make decision to C0 to switch on for using main grid power.

5.3 Entity Classification

Consumer or end user is an important stakeholder in the micro-grid and have a specific weight. The difference is based on the activities done by costumer and we devise the demand to two types: Obligatory and optional. Obligatory is the emergency needing like people with health problems and optional is the comfortable activities like washing clothes. Classification parameter is decisive parameters in our distribution algorithm and we categorize three classification: C1 high level of emergency, C2 medium level emergency and C3 low level emergency.

The classification is not static and we take real time information or the newest information stored in database, if an entity is C2 today, it can change to C1 category tomorrow if there is an emergency healthy problem with time.

5.4 Selection Scheme Algorithm

Objective is to build a dynamic smart selection of different entities connected to the micro grid and response to all loads by emergency and level of priority and have an order in our distribution of energy to reduce costs and manage our local energy produced by building DER. the general condition is to not use the main grid power and be in an autonomous building powered and to satisfy this condition we have to control the power generated with the loads, total loads must not exceed the generation capacity of DERs. The power existing P(total) as local resource is the sum of different energy produced by solar PV panels and energy existing in batteries. The concept of the smart distribution and the dynamic selection is to start to satisfy the obligatory demand (P) and secondly satisfy the facultative demands, When all P demand are satisfied we start then satisfy the Q demand.

The searching of neighboring entities is done with arbitrary searching because we have limited houses and through wired/wireless communication and controllers we have access to real time parameters information. For example if we suppose that our islanded building contain different entities classified as E1(C1, P1, Q1), E2(C3, P2, Q2 = 0), E3(C1, P3, Q3) and E4(C2, P4, Q4), if we suppose that we can cover all demands during outage time then the distribution algorithm result will be [E1, E3, E4, E2], E1 and E3 are the same and we can start with E3.

The dynamic approach consists to collect information in real time from different controllers to know the existing power and the entities belonging to the building area, after updating data information the first thing is to identify the entities with high level of emergency. The idea is to obtain a one vector of classified components (V[n]) where the first components are entities declared with high priority followed by medium and at last the low classified. The concept is described below and by steps:

- 1- The CPSM Collect information from different Bus connected to DER to know the energy available.
- 2- The CPSM Collect information from different Bus connected to entities to know load of each entity (P, Q and C).
- 3- Identify the special entity with high level of emergency E(C1). DO
 - Verify if the P(total) > P(E(C1)), if ok add entity to selection list in the first column and do same for each E(C1) in a new row and give flag = 1 to avoid double selection.
 - Update P(total) = P(total)-P(E(C1))
- 4- If the P(Total) > 0 and entities is not all covered.
- 5- Repeat 4 until response the P load of C1 entities.
- 6- If the P(Total) > 0 and components is not all covered then start response obligatory load of entities with classification C2 until response all demand or the P existing will be depleted.
- 7- If the P(Total) > 0 and components is not all covered.
- 8- Start response obligatory load of entities with classification C3 until response all demand or the P existing will be depleted.
- 9- If the P(Total) > 0 and components is not all covered.
- 10- Start response facultative load of entities with classification C1 until response all demand or the P existing will be depleted.
- 11- If the P(Total) > 0 and components is not all covered.
- 12- Start response facultative load of entities with classification C2 until response all demand or the P existing will be depleted.
- 13- If the P(Total) > 0 and components is not all covered.
- 14- Start response facultative load of entities with classification C3 until response all demand or the P existing will be depleted.
- 15- Finally, when the rest of P(Total) can cover all demands building then we stop the process of algorithm or when cover all entities.

6 Simulation and Analyses

The islanding scheme model implementation is done with open source software (ECLIPSE, JAVA and MYSQL).

6.1 Case of Study

Table 2 present parameters of different entities in the simulation. Simulation is done for 10 entities with different classification and different loads. The test system adopts reactive local compensation, so all power loss is ignored. To be in a dynamic context, we use the random function to generate a different weight factor entities with different classification, entity with weight factor under 0.02 is C1 category, entity with weight factor between 0.02 and 0.05 is categorized C2 and entity weight factor more than 0.05

is categorized C3, the total DERs active power/Kwh is 400 KW with combination of renewable resource PV/Batteries.

Entity	P (Kwh)	Q (Kwh)	Classification
E(1)	10	5	C3
E(2)	10	5	C3
E(3)	100	300	C2
E(4)	10	5	C3
E(5)	100	300	C2
E(6)	10	5	C3
E(7)	10	5	C3
E(8)	10	5	C3
E(9)	10	5	C3
E(10)	200	500	C1

Table 2. Load characteristics

6.2 Simulation Results and Discussion

According to the selection scheme algorithm proposed in this paper, the simulation steps and results are presented as follows.

The building contains 10 houses with different classification and contains some DERs (PV/Batteries). After generation of different classification with a random function and collecting the P and Q parameters from controller's house, we begin filling our vector with selected entities. In our case we have only E(10) classified with high priority V[P(E10)]. P(total was 400Kwh and after selection of E10 we have P(total = P(total) -P(E10) = 200kwh. Entities with C2 Classification are E5 and E3 with P = 100 for both, we start with E5 then V[P(E10), P(E5)] and update P(total) = P(total)-P(E5) = 100kwh and after we control the possibility to add E3 if w can cover E3 we add to the vector and we will have V[P(E10), P(E5), P(E3)] with P(total) = P(total)-P(E3) = 0.we ignore all

Entity	P (Kwh)	Q (Kwh)	Classification	FLAG
E(1)	10	5	C3	Null
E(2)	10	5	C3	Null
E(3)	100	300	C2	1
E(4)	10	5	C3	Null
E(5)	100	300	C2	1
E(6)	10	5	C3	Null
E(7)	10	5	C3	Null
E(8)	10	5	C3	Null
E(9)	10	5	C3	Null
E(10)	200	500	C1	1

Table 3. Simulation results

loses that can be exist by different causes. When we arrive to depth all existing P, we stop selection. The collection, analyze and selection algorithm take a few seconds. Table 3 present the results in Mysql database after running program, each entity selected is flagged to not be selected twice.

7 Conclusion

This paper focuses on energy efficiency in smart building through a CPS based on distributed generation and smart distribution systems using intelligent micro grid and we have proposed a new smart islanding selection with dynamic parameters to control and perform the distribution network in an islanded building during outage and to self building powered. A next step of work is to validate our algorithm with others similar algorithm of smart distribution and compare results to validate our approach. The approach of smart islanding building algorithm is developed and implemented to resolve an autonomous management of energy with a smart CPS that monitor, analyze, plan and execute decision to all controllers and actuators in a regular intervals and to response the maximum of demand and have a continuous power supply by taking actions that depends on the context and historic information.

References

- Wang, Q., Zhang, C., Ding, Y., Xydis, G., Wang, J., Stergaard, J.: Review of real-time electricity markets for integrating distributed energy resources and demand response. Appl. Energy 138, 695–706 (2015)
- Weiss, M., Helfenstein, A., Mattern, F., Staake, T.: Leveraging smart meter data to recognize home appliances. In: International Conference on Pervasive Computing and Communications (PerCom), pp. 190–197. IEEE (2012)
- Gill, H., Baheti, R.: Cyber-physical systems. In: The Impact of Control Technology, pp. 161– 166 (2011)
- Gurgen, L., Gunalp, O., Benazzouz, Y., Gallissot, M.: Self-aware cyber-physical systems and applications in smart buildings and cities. In: Conference on Design, Automation and Test in Europe, pp. 1149–1154 (2013)
- 5. Rajkumar, R., Lee, I., Sha, L., Stankovic, J.: Cyber-physical systems: the next computing revolution. In: The 47th Design Automation Conference, pp. 731–736. IEEE (2010)
- 6. Kleissl, J., Agarwal, Y.: Cyber-physical energy systems: focus on smart buildings. In: Design Automation Conference, pp. 749–754. IEEE (2010)
- Cao, L., Tian, J., Zhang, D.: Networked remote meter-reading system based on wireless communication technology. In: IEEE International Conference on Information Acquisition, pp. 172–176 (2006)
- Kushiro, N., Suzuki, S., Nakata, M., Takahara, H., Inoue, M.: Integrated residential gateway controller for home energy management system. IEEE Trans. Consum. Electron. 49(3), 629636 (2003)
- Komninos, N., Philippou, E., Pitsillides, A.: Survey in smart grid and smart home security: issues, challenges and countermeasures. IEEE Commun. Surv. Tutor. 16(4), 1933–1954 (2014)

- 10. Khanna, A.: Smart grid, smart controllers and home energy automation—creating the infrastructure for future. Smart Grid Renew. Energy **03**(03), 165–174 (2012)
- Madkour, M., Benhaddou, D., Khalil, N., Burriello, M., Cline Jr., R.E.: Living campus: towards a context-aware energy efficient campus using weighted case based reasoning. In: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
- 12. Madkour, M., Maach, A.: Ontology based context modeling for vehicle context_aware services. J. Theor. Appl. Inf. Technol. (2013)
- Peças Lopes, J.A., Silvan Polenz, A., Moreira, C.L., Cherkaoui, R.: Identification of control and management strategies for LV unbalanced microgrids with plugged-in electric vehicles. Electr. Power Syst. Res. 80(8), 898–906 (2010)
- Peças Lopes, J.A., Soares, F.J., Almeida, P.M.R.: Integration of electric vehicles in the electric power system. Proc. IEEE 99(1), 168–183 (2011)
- Guerrero, J.M., Vasquez, J.C., Matas, J., Vicuna, L.G., Castilla, M.: Hierarchical control of droop-controlled AC and DC microgrids—a general approach toward standardization. IEEE Trans. Industr Electron 58(1), 158–172 (2011)
- Hamdaoui, Y., Maach, A.: An intelligent islanding selection algorithm for optimizing the distribution network based on emergency classification. In: 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), pp. 1–7. IEEE (2017)
- Hamdaoui, Y., Maach, A.: A smart approach for intentional islanding based on dynamic selection algorithm in microgrid with distributed generation. In: 2017 International Conference on Big Data, Cloud and Application (BDCA), pp. 1–7. ACM (2017)
- Hamdaoui, Y., Maach, A.: Smart islanding in smart grids. In: 2016 IEEE Smart Energy Grid Engineering (SEGE), pp. 175–180. IEEE (2016)
- Kuang, Y.: A review of renewable energy utilization in islands. Renew. Sustain. Energy Rev. 59, 504–513 (2016)
- Ma, T., Yang, H.X., Lin, L.: A feasibility study of a stand-alone hybrid solar-windbattery system for a remote island. Appl. Energy 121, 149–158 (2014)
- Chen, H., Xiong, P., Schwan, K., Gavrilovska, A., Xu, C.: A cyber-physical integrated system for application performance and energy management in data centers. In: 2012 International Green Computing Conference (IGCC), pp. 1–10 (2012)
- Korkua, S.K., Thinsurat, K.: Design of ZigBee based WSN for smart demand responsive home energy management system. In: 2013 13th International Symposium on Communications and Information Technologies (ISCIT), pp. 549–554 (2013)
- Hamdaoui, Y., Maach, A.: Dynamic balancing of powers in islanded microgrid using distributed energy resources and prosumers for efficient energy management. In: 2017 IEEE Smart Energy Grid Engineering (SEGE). IEEE (2017)
- Pang, C., Dutta, P., Kezunovic, M.: BEVs/PHEVs as dispersed energy storage for V2B uses in the smart grid. IEEE Trans. Smart Grid 3, 473–482 (2012)
- Kiviluoma, J., Meibom, P.: Methodology for modelling plug-in electric vehicles in the power system and cost estimates for a system with either smart or dumb electric vehicles. Energy 36, 1758–1767 (2011)

Genetic Algorithm for Reusable Containers Management Problem

Mohammed Rida Ech-Charrat^{1(⊠)}, Khalid Amechnoue¹, and Tarik Zouadi²

 ¹ National School of Applied Sciences, 90000 Tangier, Morocco charrat.mohammed@gmail.com
 ² BEAR Lab, Rabat Business School, International University of Rabat, 10000 Rabat, Morocco

Abstract. This paper deals with the reverse flow management. We address the dynamic assignment problem of reusable containers in the supply chain (e.g. gas bottles, pallets, maritime containers, etc.). The objective is to optimize the collect, reloading, storage and redistribution operations taking into account the environmental constraints. We propose a newly generic mathematical model, which describes the studied problem. This mathematical model is solved following IBM ILOG CPLEX platform; although this method gives exact solutions, it is very time consuming, therefore, we adapted a hybrid approach based on a genetic algorithm to solve the problem at a reduced time. The numerical results show that the developed hybrid approach gives near optimal solutions in a moderate time.

Keywords: Reverse logistics \cdot Collect \cdot Return flow \cdot Hybrid algorithm \cdot Reusable container

1 Introduction

Managing reverse flow is becoming a crucial element of a supply chain management and often is a profit generating activity. We are interested in this study in reusing activities, mainly the dynamic assignment problem of reusable containers.

Producers in several countries are facing increasing market pressures to use reusable containers. Along the same lines, many studies in the literature propose models that could help companies to manage effectively those reusable containers. Goudenege et al. [1] proposes a generic model for reverse logistics management focused on reusable containers. The authors adapt the model to the specific requirements of companies and they focuses on a precise and real-life industrial application at a luxury goods company. Accorsi et al. [2] propose an original conceptual framework for the integrated design of a food packaging and distribution network. The framework can be applied to different food manufacturing and distribution supply chains. Atamer et al. [3] present a study that focus on pricing and production decisions for the utilization of reusable containers with stochastic customer demand. Our model considers a manufacturer that sells a single product to the customers in reusable containers with two supply options brand-new containers or reused containers. The returned quantity of the

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_5

used containers depends on both customer demand and the acquisition fee determined by the manufacturer. The authors also consider resource restrictions on the production operations.

On the other side, evaluating the environmental impact of the distribution, production and energy used in the supply chain activities is one of the aims of the green supply chain. The signed Kyoto protocol engage many countries and companies to a carbon emission quota. Several methodologies are used to calculate carbon emissions [4]. Many studies in the literature propose many technics to integrate the emission constraints in the quantitative models. Absi et al. [5] propose four types of carbon emission constraints (Periodic, cumulative, global and rolling) for the problem of the multi-sourcing lot-sizing, which can be used and adapted to several other cases.

In this paper, we propose an integrated planning for the collect, reloading, storage and redistribution of the reusable containers in a close-loop supply chain under carbon emission constraints.

2 Paper Preparation

2.1 Problem Definition

The objective of this research is to propose a containers exploitation policy that would minimize the holding, reloading and transportation costs for the warehouses and the stores, satisfy the customer's need and respect the carbon emission restrictions. The proposed model determines on each period; (1) the delivered quantities from each warehouse to different client; and (2) the collected quantities from each client to their destinations. Delays related to goods consumption and reloading are incorporated in this optimization, but also delays related to the transportation from and to the clients (Fig. 1).



Fig. 1. Generic model scheme

We consider that the full delivered containers are consumed by the stores and recuperated empty thereafter by the warehouses to be reloaded and redistributed again to the stores on a finite planning horizon.

In this configuration (Fig. 1), the following is assumed: (1) the client demands are deterministic over a finite planning horizon; (2) each warehouse or store has its own holding cost for full and empty containers; (3) each warehouse may serve many clients on each period; (4) returnable containers quantities split is allowed for each client; (5) transportation costs may be different for each client/warehouse; (6) each warehouse has its own returnable containers maintenance cost; (7) warehouses and stores possess a limited storage capacity; (8) the transportation is achieved using limited capacity vehicle; (9) consumption and reload delays are fixed by the stores and the warehouses.

2.2 Mathematical Model

The main parameters incorporated in this model are

- (N, M, T): (Number of warehouses, Number of clients, Number of periods).
- $Z_{i,j,t} : \qquad \mbox{Binary variable indicating if a client purchases full containers from a warehouse in period t}$
- $\begin{array}{ll} Y_{i,j,t} \hbox{:} & & \\ Binary \text{ variable indicating if a warehouse recovers empty containers from} \\ a \text{ client in period } t \end{array}$
- recov_{i,j}: Recovering cost from a client i to a warehouse j.
- pur_{i,j}: Ordering cost of a client i from a warehouse j.
- XF_{i,j,t}: Full containers delivered from a warehouse i to a client j in period t.
- XE_{i,j,t}: Empty containers delivered from a warehouse i to a client j in period t

$$\begin{split} \text{TransC} &= \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{M} \sum\nolimits_{t=1}^{T} \begin{pmatrix} \left(Z_{i,j,t} \, * \, \text{pur}_{i,j} \right) + \left(Y_{i,j,t} \, * \, \text{recov}_{i,j} \right) + \\ \left(\left(XF_{i,j,t} \, + \, XE_{i,j,t} \right) \\ * \, \text{trans}_{i,j} \, * \, \text{dist}_{i,j} \end{pmatrix} \end{pmatrix} \\ \text{HoldC} &= \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(\left(\text{IFW}_{i,t} \, * \, h_{f_{wi}} \right) + \left(\text{IEW}_{i,t} \, * \, h_{e_{wi}} \right) \right) + \\ \sum\nolimits_{j=1}^{M} \sum\nolimits_{t=1}^{T} \left(\left(\text{IFC}_{j,t} \, * \, h_{f_{cj}} \right) + \left(\text{IEC}_{j,t} \, * \, h_{e_{cj}} \right) \right) + \sum\nolimits_{i=1}^{N} \begin{pmatrix} \left(\text{IOEW}_{i} \, * \, h_{e_{wi}} \right) + \\ \left(\text{IOFW}_{i} \, * \, h_{f_{wi}} \right) \end{pmatrix} \\ \text{ProdC} &= \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(Q_{i,t} \, * \, \text{Prod_cost}_{i} \right) \end{split}$$

The integrated planning model is shown in the following:

$$Min (TransC + HoldC + ProdC)$$
(1)

Subject to:

$$IFW_{i,t} = IFW_{i,t-1} + Q_{i,t-Delay} - \sum_{j=1}^{M} (XF_{i,j,t}) \quad \forall i, (\forall t > Delay, \text{ if not } Q = 0)$$
(2)

$$IEW_{i,t} = IEW_{i,t-1} - Q_{i,t} + \sum_{j=1}^{M} (XE_{i,j,t}) \forall i$$
(3)

$$IFC_{j,t} = IFC_{j,t-1} + \sum_{i=1}^{N} \left(XF_{i,j,t-d_{j,i}} \right) - dem_{j,t} \qquad \forall j, \forall t > 1, (\forall d_{j,i} < t \text{ else } XF = 0)$$
(4)

$$IEC_{j,t} = IEC_{j,t-1} - \sum_{i=1}^{N} \left(XE_{i,j,t+d_{j,i}} \right) + dem_{j,t-1} \forall j, \forall t > 1, \left(\forall d_{j,i} \le T - t \text{ else } XE = 0 \right)$$

$$(5)$$

$$\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{t=1}^{T} \left(\left(XF_{ijt} + XE_{i,j,t} \right) * CO2 * dist_{i,j} \right) \le Emax * \sum_{j=1}^{M} \sum_{t=1}^{T} dem_{j,t}$$
(6)

$$XF_{ijt} \le H * Z_{i,j,t} \& XE_{ijt} \le H * Y_{i,j,t} \quad \forall i, \forall j, \forall t$$

$$(7)$$

The objective function (1) calculate the solution fitness. It minimizes the transportation, reloading, investment and holding cost of reusable containers between warehouses and stores. Constraints (2 & 3) are respectively the inventory flow conservation equations for full and empty reusable containers in the warehouses. Constraints (4 & 5) are respectively the inventory flow conservation equations for full and empty reusable containers in the warehouses. Constraints (4 & 5) are respectively the inventory flow conservation equations for full and empty reusable containers in the stores. Constraint (6) presents the unitary carbon emission over the whole horizon that cannot be larger than the maximum unitary environmental impact allowed. Constraints (7) guarantee the cancellation of full and empty reusable containers deliveries when no delivery is programmed. There are also constraints concerning the capacities, the initial state of the various operations and the positivity of the decision variables.

3 Hybrid Approach

Researchers use several approaches to improve the general efficiency and the quality of the proposed resolution methods, such as the hybrid approach; in this contribution, we propose a hybrid approach based on the genetic algorithm followed by an exact resolution method.

The genetic algorithm (GA) determines the binary decisions values related to the full and empty containers deliveries and/or collected. While the exact method determines the integer variables (quantities of the full and empty reusable containers delivered, collected, and reloaded).

3.1 Solution Encoding

We use a binary encoding where each chromosome is represented using a binary string. Figure 2 illustrates the proposed encoding.

The proposed binary encoding presents a 3-dimensional vector, with the period of the planning horizon "T" in the first axe, the warehouses number "N" on the second axe, and a double of the number of clients on the last axe "M" (for the deliveries and collects).

53



Fig. 2. Solution encoding

3.2 Crossover

The initial population is 3-dimensional vectors. So we propose a crossover operator able to cross two 3-dimensional vectors. Figure 3 illustrates the proposed crossover operator.

The proposed crossover operator consists of splitting the 3-dimensional vectors (Parent A and B in Fig. 3) into multiple 2D vectors over the axe "T" (Parent A1 and B1 in Fig. 3). Each one of these 2D matrices corresponds to the deliveries and collection launching decisions of the two period of the planning horizon (T). we use the crossover operator proposed by Toledo et al. [6] to cross tow 2D matrices.

The crossover operator consists on crossing the first half of the matrix Parent A1 (presenting the deliveries quantities decisions launching) with the first half of the matrix Parent B1, and the second half of the Parent A1 (presenting the collection quantities decisions launching) with the second half of the Parent B1to have the off-spring 1. The same procedure is launched to cross the Parent A2 with Parent B2 to get offspring 2 and so on. The two offspring 1 and 2 are assembled in a 3-dimensional vector presenting the resulted offspring.

In order to insure the mutation of the decisions matrixes, we propose the implementation of four matrix mutation operators proposed by Teledo et al. [6].

The first operator consists on changing a value randomly chosen. The second operator, two values are randomly chosen and inverted from the same column. The third operator two values are randomly chosen and inverted from the same row. The last operator consists on applying the first mutation operator twice. One of these four mutations is randomly selected.



Fig. 3. Crossover operator

3.3 Exact Resolution to Determine Quantities

At each iteration of the Hybrid algorithm, the generated offspring determine the binary decisions variable values of full and empty reusable containers deliveries or collects. These propositions generated by the offspring are integrated in the mathematical model that is then solved by CPLEX.

The solution returned by CPLEX defines optimal reusable containers quantities to deliver or to collect according to the binary decisions proposed by the offspring structure resulting from the crossover and mutation operator.

4 Result and Discussion

The hybrid method employ parameters that requires fine tuning. Based on a large number of runs, these parameters was finally selected (β ; μ ; α ; Γ) = (500; 50; 0.7; 0.3). β , μ , α and Γ denote respectively the number of crossing, the number of individuals in the initial population, the probability of crossover and The probability of mutation

The instances are derived from the article of Teunter et al. [7] and Zouadi et al. [8]. Five different types of demand and return patterns (stationary, linearly increasing, linearly decreasing, seasonal (peak in the middle), and seasonal (valley in the middle)), 6 horizon lengths, 4 set of warehouses, and 8 sets of clients are considered.

Tests are performed on 960 instances; The notations used are: (1) AS(average solution); (2) AT (average time per second); (3) AS (average solution); HA (Hybrid algorithm); GHAC (gap between Hybrid method and CPLEX); MaxG (The maximum gap found).

4.1 Grouped According to the Number of Periods of the Planning Horizon

The HA gives solutions that are close by 2.56% to the optimal solutions (obtained by CPLEX). The quality of solutions depends on the length of the planning horizon. For small horizons (5 periods), the method gives near optimal solutions, while for long horizons the solution is pretty far from being close to optimal. Regarding the computational time, CPLEX is rather efficient to prove optimality for small size instances,

Periods	Cplex		HA			
	AS	AT	AS	GHAC	AT	MaxG
5	80889.34	66	81989.44	1.36%	73	7.00%
10	193663.88	332	198931.54	2.72%	211	17.27%
15	315144.58	3425	324977.09	3.12%	396	8.10%
20	433879.89	8409	447156.61	3.06%	501	13.38%
25	577111.09	13353	592923.93	2.74%	779	9.10%
30	747121.18	24438	764529.10	2.33%	968	11.80%

Table 1. The results obtained according to the number of periods.

but on larger ones the computational time becomes more important and exceeds half an hour on many instances without finding an optimal solution (Table 1).

4.2 Grouped According to the Number of Periods of the Planning Horizon

Table 2 show that when the number of the client or the warehouses increases the gap between the hybrids approach and Cplex.

Warehouses	2	4	8	12
Clients	GHA1C	GHA1C	GHA1C	GHA1C
2	0.11%	0.92%	2.53%	3.12%
4	1.17%	1.68%	1.93%	2.98%
6	1.43%	1.90%	2.27%	3.29%
8	1.37%	2.73%	2.19%	3.34%
10	1.96%	2.36%	2.62%	3.89%
20	2.43%	2.97%	3.21%	3.68%
30	2.34%	3.23%	3.47%	4.27%
40	2.86%	3.22%	3.82%	4.07%

Table 2. The results obtained according to the number of warehouses and clients.

4.3 Grouped According to the Number of Periods of the Planning Horizon

Table 3 shows that the HA perform better when the demand is constant. However, when the demand is positive or negative trend, the approaches are less efficient. When demands are seasonal peak in middle and seasonal valley in middle, a small gap is noticed.

Type of demands	Cplex		HA		
	AS	AT	AS	GHAC	AT
Stationary	289630.48	7371	295799.60	2.13%	319
Positive trend	506910.90	8964	521966.15	2.97%	573
Negative trend	472475.81	8563	485846.87	2.83%	597
Seasonal (peak in middle)	356277.15	6931	365255.33	2.52%	517
Seasonal (valley in middle)	331213.96	6024	339063.73	2.37%	434

Table 3. The results obtained according to the type of demand.

5 Conclusion

In this contribution, an assignment model for reusable containers distribution and collection is proposed and a hybrid approach was developed to solve the problem. The numerical results show that the developed hybrid approach generates high-quality solutions in a moderate computational time. This contribution presents a base to develop more generic models by considering several constraints, and more meta-heuristics such as improving the hybrid approach with a local search procedure.

References

- 1. Goudenege, G., Chu, C., Jemai, Z.: Reusable containers management: from a generic model to an industrial case study. Supply Chain Forum Int. J. **14**(2), 26–38 (2013)
- Accorsi, R., Cascini, A., Cholette, S., Manzini, R., Mora, C.: Economic and environmental assessment of reusable plastic containers: a food catering supply chain case study. Int. J. Prod. Econ. 152, 88–101 (2014)
- Atamer, B., Bakal, İ.S., Bayindir, Z.P.: Optimal pricing and production decisions in utilizing reusable containers. Int. J. Prod. Econ. 143(2), 222–232 (2013)
- 4. RetelHelmrich, M.J., Jans, R., van den Heuvel, W., Wagelmans, A.P.M.: The economic lot-sizing problem with an emission capacity constraint. Eur. J. Oper. Res. (2014)
- Absi, N., Dauzère-Pérès, S., Kedad-Sidhoum, S., Penz, B., Rapine, C.: Lot sizing with carbon emission constraints. Eur. J. Oper. Res. 227(1), 55–61 (2013)
- Toledo, C.F.M., Ribeiro de Oliveira, R.R., França, P.M.: A hybrid multi-population genetic algorithm applied to solve the multi-level capacitated lot sizing problem with backlogging. Comput. Oper. Res. 40, 910–919 (2013)
- 7. Teunter, R.H., Pelin Bayindirand, Z., Van Den Heuvel, W.: Dynamic lot sizing with product returns and remanufacturing. Int. J. Prod. Res. 44, 4377–4400 (2006)
- 8. Zouadi, T., Yalaoui, A., Reghioui, M., El Kadiri, K.E.: Lot-sizing for production planning in a recovery system with returns. RAIRO Oper. Res. **49**(1), 123–142 (2015)

Multi-agent Modeling of Resource Allocation Under Competence and Emergency Constraints in the Hospital Environment

M. El Hankouri^(K), M. Kharbach, and M. Ouardouz

Mathematic Modeling and Control, Faculty of Sciences and Techniques, B.P. 416, Tangier, Morocco mohammed.elhankouri@gmail.com, skharbach@yahoo.fr, ouardouz@gmail.com

Abstract. The management of hospitals is characterized by a high degree of diversity and daily complexity of the various healthcare activities, reflecting both resource constraints and patient satisfaction. In this article we discuss the role of resource allocation modeling in hospitals, as the scheduling of its staff become troublesome, because of several elements that it builds upon, and this end up depleting a lot of time and resources. Our objective is to propose a multi-agent approach through scenarios while taking into account the various constraints related to competence and medical emergency. The simulation of the planning process dynamics and allocation of these resources is developed under the Netlogo platform.

Finally, the approach is illustrated by examples inspired by actual cases experienced in a public health hospital.

Keywords: Modeling \cdot Multi-agent system \cdot Resource allocation \cdot Hospital environment \cdot Optimization

1 Introduction

In recent years, improvements in hospital performance have undergone significant changes in both the structural and the organizational aspects. Hospitals recognize that their staff is one of the most important assets of their organization and a key determinant of its success or failure. Besides, the increasing patient needs force hospitals to constantly improve their processes to optimize their resources and use them more efficiently. Indeed, their major concern (hospitals, care centers, etc.) is centered on finding the best managerial practices to ensure patients' satisfaction and on reaching the technical-economic optimum between the organization, planning and the allocation of these resources.

Through this work, we would like to provide managers and hospital planners with a tool that could be used to resolve problems related to the management of health care services, such as the resource allocation (material, human, financial, etc.), by considering

several constraints related to staff competences and the degree of urgency of medical interventions.

This article is divided into four parts. Firstly, it will present the problem and its context. Secondly, it will examine the literature review developed around the various modeling problems concerning resource management. Finally, it will illustrate the simulation developed under the Netlogo platform.

2 Motivation and Problematic

Nowadays, the hospital environment is changing and growing rapidly both in the organization and in the planning of resources. In addition, hospitals are continually subject to budgetary and legal constraints that require significant changes in managerial practices. Their main current requirement is to develop operational tools for managing and optimizing their resources.

In this work, we will focus on the resource allocation problem under the constraints of competences and urgency, using multi-agent modeling and simulating platform in order to allow the health sector to improve the quality health care services and ensure significant enhancement in terms of efficiency. This is a large-scale problem in a context requiring high performance and quality. This problem has been in fact the subject of much research in recent decades [1], it's has received substantial consideration and is recognized as a difficult optimization problem with practical relevance.

Furthermore, the need to adopt better management tools for efficient planning and assignment has become an absolute necessity. The approach presented in this paper is based on constraint optimization and decision support [2, 3] using the multi-agent model approach.

3 Literature Review

Scientific research in the hospital environment is continuously increasing. For this purpose, hospitals that are suffering from difficult socio-economic circumstances, must undergo new strategies for organizing, managing and planning their resources in order to minimize the costs incurred and insure patients' satisfaction. To this end, many researchers have thus studied this problem by trying to suggest new approaches suitable for this environment.

Moreover, it should be noted that, on the one hand, few authors have studied the problem of management and planning of resources in the hospital environment in its entirety, and haven't taken into account the constraints of resources; On the other, there are only a few who have validated their models by real cases.

Many authors were interested in studying in depth the planning and the allocation of the resources in health care organisms, especially for operating rooms [4–9]. Certainly the operating room has a vital and important role in the hospital process. However, we consider that the subject of resource allocation and its constraints should be examined at the general hospital level, not just at the operating room level. This is also confirmed by [10] in his thesis.

The work of [11] deals with the possibility of taking advantage of the field of production management to strategic planning in hospitals, with constraints in terms of limited material and human capacities. They have tried to optimize these resources by minimizing patient wait time through the building a heuristic that assigns patients on surgery days as well as operating rooms. Nevertheless, they have not taken into account in their algorithm the variability of surgery duration.

In the same vision of [12], an analogy between the hospital environment and industrial processes was established. According to them, a hospital requires a strong coordination among its sub-groups in order to plan successfully its resources and satisfy its patients. Right after, [13] is inspired by a technique used in the industrial world called AHP "Analytical Hierarchy Process" to manage the nursing service. [14] propose a framework for the classification of nurses rostering scheduling, in view of three tiers: personnel skills, work characteristics, and optimization objectives.

A meta-heuristic approaches and their hybrids have been successfully applied by [15], in order to resolve the nurse rostering problem to meet the daily operational requirements.

According to [16], the problems associated with the allocation of human resources in the care environment may be limited to sizing and staff rotation while respecting a set of constraints related to unexpected and unforeseen events.

Subsequently, in the work of [17], they developed a model for the strategic and tactical decisions of planning and estimating needs in hospital systems in order to improve the allocation of resources by eliminating periods of staff inactivity.

Later, [18] has added another level of decision-making in real time, which allows the healthcare system to better react to frequent and unexpected factors such as staff absenteeism and emergencies.

In [19], the authors used meta-heuristics that allow the automatic generation of schedules for hospital pharmacy staff. Their approach is based on ant colonies for scheduling tasks according to their skills. Several numerical tests were proposed to demonstrate the effectiveness of the method.

We have concluded through the literature review, that hospital systems are complex. Consequently, the use of modeling methods based on a heuristic approach or on a linear programming has proved generally insufficient to provide a model that combines the various constraints linked to the problem of resource allocation.

We must therefore adopt an agent-based modeling approach where the different problems and constraints (human resources, material, urgency, etc.) are grouped together in a relevant way. In the remainder of this article, we will focus mainly on combined modeling, as the literature analysis has shown that several studies have been conducted on the subject of the healthcare resources allocation, but without taking into account both the competence and the degree of urgency of care interventions.

4 In Hospital Environment

The management of the hospital is a complex system. This complexity is reinforced by the high level of demand and by the interdependence of the care processes [20], where a large number of constraints interact with usual scheduling practices [21].

Owing to the fact that hospitals know an exponential evolution, confronted with an increasingly demanding and challenging environment, they must comply with new resource management rules, while ensuring that they are optimally allocated. The use of industrial methods and tools is an excellent approach to achieve this goal.

In most of the studies presented above, the use of simple algorithm is a common practice. However, there are many methods of modeling basically dedicated to industrial systems, including also hospitals [22].

Modeling is a very important step in the study of the problem of resource allocation under constraints, because the theoretical model makes it possible to:

- Understand how the system works,
- Determine the adequate problem solution.

Moreover, modeling by multi-agent systems is one of the first approaches to adopt when studying social, geographical, phenomena. Thus, it is increasingly applied in the industrial field because it allows the possibility of modeling processes, individuals as well as their interactions and their impacts with high precision. In addition, it permits a decrease in the complexity of the systems [23].

The multi-agent modeling makes it possible to conceptualize and simulate an organized group of agents interacting with each other and reacting with their environment [24]. Thus it offers a formalism showing easily the dynamic representation of complex systems such as hospital field.

5 Description of the Proposed Approach

5.1 Studied Health Care Process

In the remainder of our work, we will perform an agent-based simulation on the Netlogo platform. The goal of this step is to sketch a medical and paramedical team during a normal working day. According to the considered implementation, we will consider a hospital with the following services:

- An emergency service,
- An operating room,
- A room for the surgery preparation,
- A treatment area,
- Two hospitalization rooms, equipped by four beds each.

Nevertheless, a reception area is located at the entrance, serving for patient's registration, and then to focus on the priority of their care.

Then the patient is oriented according to the degree of priority of his case, towards the other services: emergency, hospitalization, etc. (Fig. 1).


Fig. 1. Patient diagram flow

5.2 Simulations

The proposed simulation process is based on a group of agents including patients and a medical team (Tables 1 and 2):

Agents: I	Patients		
Color	Coc	le	Designation
	P1	L	Urgent patient
	P2		Patient in normal consultation
	P3		Patient awaiting surgery
	P4		Hospitalized patient

 Table 1. Patients code

Table 2. Physicians code

Agents:	physicians	
Color	Code	Designation
	M1	General practitioner
	M2	Surgeon
	M3	Anesthetist / Emergency physician

6 Hospital Performance Measurement

In order to optimize budgets and expenditures, performance monitoring in the hospitals is necessary in order to achieve the planned objectives. Assessing and improving the performance of such environment, depends mainly on the definition of relevant indicators which are capable to draw a clear picture of the hospital functioning as well as to identify the problematic points to be improved [25]. In our study, several key metrics were identified (refer to Table 3), in relation to competency and emergency constraints, and which measure the effectiveness of planning, patient satisfaction, etc.

Indicator	Explication	Formula
Hospital room occupancy rate	Provide an idea of the platform's capacity relative to the average number of patients	Total of occupied beds/total of available beds
Patient passage rates	Show availability of physicians	Total of treated patients/total of existing patients
Care quality rate	Evaluates the patient's satisfaction rate	\sum of patients submitted a complaint in time t/ \sum of patients treated in moment t
Operating room occupancy rate	Evaluate the use of the operative capacity of the operating room and give an idea of the flexibility margin of the block	Time planned for operating theater/theoretical time per week $(9 \text{ h} \times 6 \text{ days})$
Over-flow rate	Measures the volume of over- flow in relation to the capacity of hospitalization	\sum of patients needed to be hospitalized in moment t/ \sum of beds available in moment t

Table 3. Key metrics

6.1 Simulation Interface and Case Studies

The simulation interface is composed of three essential zones:

- 1st zone: includes a set of buttons on the left
- 2nd zone: shows the indicators plots
- 3rd zone: illustrates the work environment (Fig. 2).



Fig. 2. Simulation interface

This simulation is established on 2D space. The modeled hospital is considered as one stair. Spatial dimension was not included on this step. In the further research, the stairs will be additional variable to be taken in account.

Exits, rooms, offices and walls are considered as some static agents.

Nurses, patient and practitioner as considered as some dynamic and autonomous agents, who reacts with their virtual surround, by feeding variables and rules into the system, when creating the simulation platform.

6.2 Simulation Tool and Case Study

The simulation interface is divided into three areas:

1st area: includes the commands on the left side.

2nd area: shows the indicators plots,

3rd area: illustrates the work environment.

Figure 3 shows how patients are taken care of, according to their urgencies and illness severity.



Fig. 3. Simulations

As result of the simulation, we can see the staff movements, by considering for each individual his own characteristics, and showing visually the interactions among people in the different scenarios (urgency, surgery, normal medical check, etc.). Agent compute constantly his environment and proceeds as should be in the real world.

By the way, the modeled agents are capable to recognize the obstacles, walls and buildings.

The system has been able to simulate some professional behavior characteristics of medical staff and spontaneous acts of the ill persons. It shows also how the space was used by the occupants.

Performance of the staff allocation can be measured and monitored via lateral sliders.

7 Conclusion and Perspectives

In this paper, we have proposed a multi-agent approach for modeling and simulating the resource allocation problem in hospitals under constraints with the objective to optimally allocate the care demands in view of the urgency degree of the medical interventions, such as doctors, nurses, rooms, equipment ... etc.

The tool is upgradable and easy to be implemented in hospital environment. This solution can be applied to constrained scheduling problems in other fields.

The considered parameters are random and spatial, a fact which rather encourages a stochastic approach based on a 3D model. These aspects are under investigation. Future research should also take in account the shift working which conventionally for the hospital staff is (day, evening and night), in order to close the gap which currently exists between research and practice. Including social, psychological, physiological parameters should be a good opportunity to expand the simulation.

References

- Jebali, A.: Vers un outil d'aide à la planification et à l'ordonnancement des ressources dans les services de soins. Ph.D. thesis, Laboratoire d'Automatique de Grenoble (LAG) - Ecole Doctorale Organisation Industrielle et Systèmes de Production (2004)
- Sahraoui, S.A.: Un système d'aide à la décision pour une amélioration optimisée de la performance industrielle. Ph.D. thèse, Université de Savoie, France (2009)
- Camalot, J.P.: Aide à la décision et à la coopération en gestion du temps et des ressources. Ph.D. thèse, Institut National des Sciences Appliquées, Toulouse (2010)
- Dexter, F., Macario, A.: Decrease in case duration required to complete an additional case during regularly scheduled hours in an operating room suite: a computer simulation study. Anesth. Analg. 88, 72–76 (1999)
- Chaabane, S., Guinet, A., Smolskib, N., Guiraudc, M., Luquetb, B., Marcond, E., Viale, J.P.: La gestion industrielle et la gestion des blocs opératoires. Annales françaises d'anesthésie et de réanimation 22, 904–908 (2003)
- Belien, J., Demeulemeester, E.: A branch and price approach for integrating nurse and surgery scheduling. Eur. J. Oper. Res. 189, 652–668 (2007)
- Jebali, A., Bouchriha, H.: Evaluation de deux strategies de planification des interventions dans un bloc opératoire central. Logist. Manag. 15(1), 27–36 (2007)
- Hanset, A., Meskens, N., Roux, O., Duvivier, D.: Conception d'un modèle modulaire d'ordonnancement du bloc opératoire avec prise en compte des contraintes liées aux ressources humaines et matérielles. Logistique et Management (2011)
- 9. Van Oostrum, J., Bredenhoff, E., Hans, E.: Suitability and managerial implications of a master surgical scheduling approach. Ann. Oper. Res. **178**, 91 (2010)
- 10. Lamiri, M.: Planification des blocs opératoires avec prises en compte des aléas. Thèse de doctorat, Université de St Etienne (2007)
- Guinet, A., Chaabane, S.: Une approche de type MRP2 pour la gestion des blocs. Gestion et Ingénierie des Systèmes Hospitaliers, Luxembourg (2003)
- Andre, V., Fenies, P.: Modélisation et simulation des flux logistiques du Nouvel Hôpital d'Estaing. Revue Logist. Manag. 15(1), 49–59 (2007)
- Toploglu, S.: A shift scheduling model for employees with different seniority levels and an application in healthcare. Eur. J. Oper. Res. 198(3), 943–957 (2008)

- De Causmaecker, P., Vanden Berghe, G.: A categorisation of nurse rostering problems. J. Sched. 14, 3–16 (2011)
- 15. Burke, E.K., De Causmaecker, P., Berghe, G.V., Van Landeghem, H.: The state of the art of nurse rostering. J. Sched. 7, 441–499 (2004)
- 16. Trilling, L.: Aide à la décision pour le dimensionnement et le pilotage de ressources humaines mutualisées en milieu hospitalier. Ph.D. thesis, INSA de Lyon (2006)
- 17. Roth, A., Dierdonck, V.: Hospital resource planning: concepts, feasibility, and framework. Prod. Oper. Manag. 4(1), 2–29 (1995). Winter
- Jebali, A., Ladet, P., Hadj-Alouane, A.: Une méthode pour l'ordonnancement du bloc opératoire. Journal européen des systèmes automatisés 38, 154 (2004)
- 19. Guinet, A., Chaabane, S.: Operating theatre planning. Int. J. Prod. Econ. 85, 69–81 (2003)
- 20. Daknou, A.: Architecture distribuée à base d'agents pour optimiser la prise en charge des patients dans les services d'urgence en milieu hospitalier. Thèse de Doctorat, l'Ecole Centrale de Lille (2011)
- Trilling, L., Besombes, B., Chaabane, S., Guniet, A.: Investigation et Comparaison des Méthodes et Outils d'Analyses pour l'Etude des Systèmes Hospitaliers (2004)
- Augusto, V.: Modélisation, analyse et pilotage de flux en milieu hospitalier a l'aide d'UML et des réseaux de Petri, Vincent Augusto. Thèse de Docteur de l'Ecole Nationale Supérieure des Mines de Saint-Etienne (2008)
- 23. El Hankouri, M., Kharbach, M., Ouardouz, M., Bernoussi, A.: Modélisation multi-agents d'allocation des ressources sous contraintes de compétence et d'urgence dans le milieu hospitalier, lère édition du colloque international de logistique, Tétouan, Maroc (2017)
- 24. Amblard, F., Phan, D.: Modélisation et simulation multi-agents: applications pour les Sciences de l'Homme et de la Société. Lavoisier (2006)
- 25. Serrou, D., Abouabdellah, A.: Mesure de la performance de la chaîne logistique hospitalière en intégrant les dimensions: Coûts, Sécurité et Qualité: Application en cas du regroupement des pharmacies. Xème Conférence Internationale: Conception et Production Intégrées, CPI 2015, 2–4 Décembre 2015, Tanger – Maroc (2015)

Context Awareness-Based Ontology Using Internet of Things for Multimedia Documents Adaptation

Hajar Khallouki^(III) and Mohamed Bahaj

Mathematics and Computer Science Department, Faculty of Sciences and Techniques, Hassan I University, Settat, Morocco hajar.khallouki@gmail.com, mohamedbahaj@gmail.com

Abstract. The Internet of things (IoT) is a technical concretization of ubiquitous computing where technology is naturally integrated into the things of our daily life. The main objective of IoT is to make the world easier for human beings where things around us predict our preferences and context, and act autonomously without human interactions. Adapting multimedia documents to user context is a serious issue for user-aware development. Good context based-adaptive approach requires dynamic adaptation to the user, not only to its position, time of day, or to environment characteristics but also to its mood, preferences, or even disabilities. In this paper, we introduce a context-based approach which combines the IoT with semantic web services to predict the user current context and enable a dynamic adaptation of multimedia documents.

Keywords: Internet of things \cdot Ubiquitous computing \cdot Context \cdot Semantic web \cdot Dynamic adaptation \cdot Multimedia documents

1 Introduction

The term "Internet of things" was initially proposed by Kevin Ashton [1] in 1999 to describe things which are equipped with radio frequency identification chips (RFID chips). However, the concept has been developed during the last years and generalized towards an approach connecting a very large number of things to the internet, enabling them to provide services and collect context information independently.

The context awareness is a key element for creating smart applications. The collected contextual data is generally raw information from different distributed sources, which needs to be interpreted. Based on ontologies, it is possible to build semantic models which will be fed by this raw data and thus not only to increase their level of semantic representation but especially of being able to use them to make automatic decisions of adaptation of applications based on the context to the runtime.

The real-time adaptation, in particular, multimedia documents adaptation, raises complex scientific issues as well as new challenges for the execution and the development of applications. Several approaches [2, 3, 4] have been proposed in this area to model user context and enable a dynamic adaptation of multimedia documents. However, these approaches don't ensure the interoperability and communication of different devices and sensors.

The key contributions of this paper include:

- Context awareness ontology modeling based on IoT for multimedia documents adaptation.
- Evaluation of the proposed ontology using a set of SPARQL queries.
- Describing some examples of reasoning rules in order to predict and manage the adaptation services.
- Running a few experiments to calculate the response time of a SPARQL query on the proposed ontology and compare it with an existed one.

The rest of the paper is organized as follows. Section 2 introduces a description of related work on context awareness and sensor modeling within the IoT. Section 3 presents the context modeling and reasoning through ontology, and Sect. 4 outlines the experimentation process. Finally, Sect. 5 concludes the paper.

2 Related Work

Context-aware solutions have been widely used in various works. Adomavicius et al. [5] proposed the Context-Aware Recommender Systems (CARS) which generate a recommendation by adapting them to the specific contextual situation of the users. The contextual information helps to create an intelligent and useful recommendation system [6].

Bernardos et al. [7] identified two steps in representing context according to a model; the first step deals with the definition of new context information in terms of attributes, characteristics, relationships with existed specified context and the queries for synchronous context requests. The second step aims to validate the result of context modeling. Hence, the new context information becomes available to be exploited when required.

A context model has been proposed by Changbok Jang et al. [8] to provide users with suitable services and manage resources effectively by using context information in the Mobile Cloud environment. They defined context for modeling through diverse context definitions, classified ontology and represent hierarchically. The proposed context model by was expected to help have the optimized personalized service and effective IT resources management in the Mobile Cloud environment.

Furkh Zeshan et al. [9] presented a context-aware ontology and a framework along with an algorithm for web service discovery and ranking for Distributed Embedded Real-Time System (DERTS) to facilitate the discovery of device services in embedded and real-time system environments. The proposed service discovery framework also considers the associated priorities with the requirements posed by the requester during the service discovery process.

Since the last decade, a lot of semantic descriptions have been designed for sensor modeling. The SSN ontology [10] is a work provided by the W3C Semantic Sensor Network Incubator Group. It describes sensors, sensing, the measurement capabilities of sensors, the observations that result from sensing, and deployments in which sensors are used.

Maria Bermudez-Edo et al. [11] proposed a lightweight semantic IoT model, IoT-Lite and a set of ontology design guidelines for dynamic and responsive environments. The model is an extension of SSN appropriate for real-time sensor discovery.

We decided to use parts of SSN ontology for some reasons, the SSN ontology is currently used in a number of research projects which make it possible to increase the interoperability between different domains and applications. Another reason remains on the fact that SSN is the most up-to-date and developed sensor ontology [12].

The adaptation of a multimedia document is conditioned by the context. The context defines conditions that must be satisfied by the document to be played. The main objectives of our proposal is to build a complete and detailed ontology which extends previous works, to model all the components taken part in the user context prediction for multimedia documents adaptation and to design a hierarchy of classes, data and relation properties. To corroborate the correctness of the ontology, a validation against consistency issues must be performed.

3 Proposed Approach

3.1 Ontology Modeling

In this section, we introduce the details of our ontological knowledge model, which is proposed to predict the current context of the user and enable a pertinent adaptation of the displayed multimedia document.



Fig. 1. A snapshot of the proposed semantic model

Context modeling in context-awareness needs to model and represent the data collected from different sources according to a meaningful way. As shown in Fig. 1, suggested semantic model which is an instantiation of the Semantic Sensor Network (SSN) ontology.

The semantic model shown in Fig. 1 is well detailed in the following figure using Protégé platform for building the proposed ontology.

The ontology is composed of different classes such as the resource class which collects the descriptions of all user devices. For each device, it gives the user's hardware and software information and characteristics like screen resolution, CPU speed, memory capacity, battery level, OS name, a set of the applications that are running, etc. The schedule class which contains user related personal information, which can influence his current situation (Fig. 2).



Fig. 2. Class hierarchy

We developed the proposed ontology using OWL language (Fig. 3). OWL is one of the emerging Semantic Web technologies that are endorsed by the W3C for building ontologies [13, 14]. Figure 3 shows some classes, data and relation properties of the proposed ontology (CaMaOntology).

```
<owl:Class rdf:ID="Activity">
    <rdfs:subClassOf rdf:resource="#Schedule"/>
</owl:Class>
<owl:InverseFunctionalProperty rdf:ID="has">
    <rdf:type rdf:resource="&owl;FunctionalProperty",
    <rdf:type rdf:resource="&owl;ObjectProperty"/>
    <rdfs:domain rdf:resource="#User"/>
    <owl:inverseOf rdf:resource="#Belongs"/>
    <rdfs:range rdf:resource="#Profile"/>
</owl:InverseFunctionalProperty>
<Activity rdf:ID="Studying"/>
<owl:Class rdf:ID="Touch">
    <rdfs:subClassOf rdf:resource="#Interaction"/>
</owl:Class>
<Activity rdf:ID="Sleeping"/>
<owl:Class rdf:ID="Agenda">
    <rdfs:subClassOf rdf:resource="#Schedule"/>
</owl:Class>
<User rdf:ID="Anne">
   <at rdf:resource="#Hour1"/>
    <at rdf:resource="#Date1"/>
   <input rdf:resource="#Voice1"/>
    <locatedIn rdf:resource="#Hospital"/>
</User>
<owl:Class rdf:ID="TagDevice">
    <rdfs:subClassOf rdf:resource="#Device"/>
</owl:Class>
<User rdf:ID="Mohamed">
   <at rdf:resource="#Date1"/>
    <locatedIn rdf:resource="#University"/>
   <is rdf:resource="#Teaching"/>
</User>
<owl:ObjectProperty rdf:ID="attends">
   <rdfs:domain rdf:resource="#User"/>
    <rdfs:range>
```

Fig. 3. OWL description

To corroborate the correctness of the proposed ontology, a validation against consistency issues was performed in the next session.

3.2 SPARQL Queries

To retrieve the data stored in an ontology there are several query languages, the most common is called SPARQL [15]. In order to validate the correctness and functioning of the proposed ontology, we performed a set of SPARQL queries. The first test aimed to find out the location of the user. The following figure shows the detail and result of the query (Fig. 4).

74 H. Khallouki and M. Bahaj

Query	F	Results	
SELECT ?location		Concompanyation	location
shttp://CallaOatalagy.ou/#Hajaca.shttp://CallaOatalagy.ou/#lagatadk	> 2leastion	- Supermarket	
sintp.//cainaoinology.ov/#riajar> sintp.//cainaoinology.ov/#rocatedii	Rocation		
1			
Execute Query			
		<u></u>	
	-		
Query	Results		
SELECT ?location ?activity		location	activity
WHERE {	University		Teaching
<http: camaontology.ow#mohamed=""> <http: camaontology.ow#locatedin=""> ?location.</http:></http:>			
<http: camaontology.ow#mohamed=""> <http: camaontology.ow#is=""> ?activity</http:></http:>			
1			
Execute Query	J		
	F Describe		
Query 25	Results		
SELECT ?language ?multimediaDocument		language	multimediaDocument
WHERE { <nttp: camauntology.ow#jonn=""> <nttp: camauntology.ow#talks=""> ?language.</nttp:></nttp:>		h	Text1
<pre><nup. camauntology.ow#jonn=""> <nup. camauntology.ow#displays=""> /multimediaDocument</nup.></nup.></pre>			
1			
	_		
Execute Query			

Fig. 4. Queries performed on the proposed ontology

The first query returns the location of the user "Hajar", the second query is asking for the location and activity of the user "Mohamed" and the third query returns the language spoken by the user "John" and the multimedia content displayed by him.

To further evaluate the proposed ontology, other queries were tested. The queries results introduced in this section returned the expected information depending on the data provided in the proposed ontology.

3.3 Reasoning Rules

In this section, we define examples of reasoning rules which we intend to use in our prototype. These rules enable a dynamic discovery of adaptation services. In this paper, we define three kinds of multimedia content adaptation services:

- Transcoding: enables the conversion of format, e.g., JPEG to PNG.
- Transmoding: enables the conversion of types, e.g., text to sound.
- Transforming: allows the content change without changing the media type and format, e.g., text summarization, language translation, etc.

In order to improve the expressiveness of the proposed ontology and enable the definition of if-then statements, we used the rule based formalism *Semantic Web Rule Language* (SWRL) [16]. The following figure illustrates the first example of a reasoning rule (Fig. 5).

The first rule is an example of reasoning rule using SWRL formalism. It indicates that if the user is attending a meeting and receives an audio content, a transmoding service may be triggered.

```
      SWRL Rule

      displays(?x, Audio1) ∧ attends(?x, Meeting1) →

      AdaptationService(Transmoding)

      SWRL Rule

      talks(?x, Spanish) ∧ displays(?y, Text1) →

      AdaptationService(Transforming)

      SWRL Rule

      IsocatedIn(?x, Car1) ∧ is(?x, Driving) ∧ displays(?x, Text2) →

      AdaptationService(Transmoding)
```

Fig. 5. Reasoning rules

4 Experimentation

This section aims to measure the response time of a simple SPARQL query Q1(SELECT * WHERE {?x ?y ?z}) on two different ontologies. UPOMA ontology [3] which models the user context for multimedia documents adaptation, and, the proposed ontology (CaMaOntology) which extends UPOMA and SSN ontologies.

For this experimentation, we used a personal computer (PC) running Windows 7 (x32) operating system with a processor Intel(R) Core(TM) 2 DUO CPU T6570 @ 2.10 GHz 2 GB RAM. Using JENA library, we performed the query Q1 simultaneously on 10, 20, 30, 50, 100 triplets.

As shown in Fig. 6, the use of a parts of SSN ontology in the proposed ontology, doesn't have a big changes on the queries execution time.



Fig. 6. SPARQL query execution time comparison of the proposed ontology and UPOMA

5 Conclusion

This paper introduces context-aware approach through an ontology which extends the SSN ontology to predict the user current context and enable a dynamic adaptation of multimedia documents.

The proposed ontology was successfully evaluated and tested by querying it using Protocol and RDF Query Language (SPARQL). Although the proposed ontology is dedicated for multimedia documents adaptation, it can be used for other use cases, such as, smart health, smart tourism, etc.

References

- 1. Ashton, K.: That 'internet of things' thing. RFiD J. 22(7) (2011)
- Dromzée, C., Laborie, S., Roose, P.: A semantic generic profile for multi-media documents adaptation. In: Intelligent Multimedia Technologies for Net-Working Applications: Techniques and Tools. IGI Global (2012)
- Khallouki, H., Bahaj, M., Roose, P., Laborie, S.: SMPMA: semantic multimodal Profile for multimedia documents adaptation. In: Proceedings of the 5th IEEE International Workshop on Codes, Cryptography and Communication Systems (IWCCCS 2014), pp. 142–147. El Jadida, Morocco, 27–28 November 2014
- Khallouki, H., Bahaj, M.: Multimodal context based-adaptive architecture for multimedia documents. In: 2015 11th International Conference on Innovations in Information Technology (IIT), pp. 140–145. IEEE, November 2015
- Adomavicius, G., Mobasher, B., Ricci, F., Tuzhilin, A.: Context-aware recommender systems. AI Mag. 32, 67–80 (2011)

- Ham, N., Dirin, A., Laine, T.H.: Machine learning and dynamic user interfaces in a context aware nurse application environment. J. Ambient Intell. Humanized Comput., 1–13 (2016)
- Bernardos, A., Tarrio, P., Casar, J.: A data fusion framework for context-aware mobile services. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI 2008, pp. 606–613, August 2008. http://dx.doi.org/10.1109/MFI. 2008.4648011
- Jang, C., Choi, E.: Context model based on ontology in mobile cloud computing. In: Advanced Communication and Networking, pp. 146–151 (2011)
- Zeshan, F., Mohamad, R., Ahmad, M.N., Hussain, S.A., Ahmad, A., Raza, I., Mehmood, A., Ulhaq, I., Abdulgader, A., Babar, I.: Ontology-based service discovery framework for dynamic environments. IET Softw. 11(2), 64–74 (2017)
- Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Phuoc, D.L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The SSN ontology of the W3C semantic sensor network incubator group. Web Semant. Sci. Serv. Agents World Wide Web 17, 25–32 (2012)
- Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., Taylor, K.: IoT-Lite: a lightweight semantic model for the internet of things and its use with dynamic semantics. Pers. Ubiquit. Comput., 1–13 (2017)
- Smirnov, D., Stutz, P.: Use case driven approach for ontology-based modeling of reconnaissance resources on-board UAVs using OWL. In: 2017 IEEE Aerospace Conference, pp. 1–17. IEEE, March 2017
- 13. Smith, M., Welty, C., McGinness, D.: Web Ontology Language (OWL) Guide Version 1 (2003)
- Harmelen, F., Hendler, J., Horrocks, I., McGinness, D., Patel-Schneider, P., Stein, L.: Web Ontology Language (OWL) Reference Version 1.0 (2003)
- W3C: "SPARQL 1.1 Query Language," W3C (2013). https://www.w3.org/TR/2013/RECsparql11-query-20130321/. Accessed 19 Oct 2016
- Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: a semantic web rule language combining OWL and RuleML. W3C Member Submission, 21 May 2004, pp. 1–20 (2004)

Face Recognition Using Deep Features

Hamid Ouanan⁽⁾, Mohammed Ouanan, and Brahim Aksasse

Department of Computer Science, M2I Laboratory, ASIA Team, Faculty of Science and Techniques, Moulay Ismail University, BP 509 Boutalamine, 52000 Errachidia, Morocco ham.ouanan@gmail.com, ouanan_mohammed@yahoo.fr, baksasse@yahoo.com

Abstract. Recent studies discovered that the human brain has a deep face-processing network, where identity are processed by multiple different neurons. Consequently, we turn our attention for using deep architecture of neural networks to reach near human performance in the world of face recognition. In this paper, we make the following contributions: Firstly, we build a novel dataset with over four million faces labelled for identity by employing a smart synthesis augmented approach based on rendering pipeline to increase the pose and lighting variability. Secondly, a robust deep CNN model taking place. Finally, we set up a new real time application of this approach proposed. This application called PubFace, which allows users to identify anyone in public spaces. Experiments conducting on the well-known LFW dataset, demonstrating that the proposed approach achieved state-of-the-art results.

Keywords: Face recognition \cdot Artificial intelligence \cdot Deep learning \cdot Data augmentation \cdot Big data \cdot Smart digital

1 Introduction

Deep learning have been widely used in computer vision community, significantly improving the state-of-the-art. Thanks to Deep Learning, in particular Convolutional Neural Networks (CNNs), the past year (2016) has seen incredible breakthrough in artificial intelligence. In March 2016, Google DeepMind's AlphaGo computer program [1] won a Korean champion to go game by four wins at one, making it the first time a computer Go program had defeated an excellent human player. In June 2016, the Chinese team of search engine Baidu announced unmatched performances in machine translation: six points better than the state of the art. In September 2016, Google replied by a better point and integration of this technique in its famous translation tool [2]. In November 2016, the team of Oxford and Google described her lecture program on the lips [3]. These are just a few of the milestones artificial intelligence (AI) that has enabled in the past year (2016). The success of deep learning stems from the fact: the availability of very large amount of training datasets, which is the main key to build a great model, based on CNNs. However, in the area of face recognition, the new advancements remain limited to Internet giants like Facebook, Flickr and Google, which have the world's private largest databases. Besides, There are many challenges in dealing with this applications listed such as variation in illumination, variability in

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_8

scale, location, orientation and pose. Furthermore, facial expressions, facial decorations, partial occlusion, and lighting conditions change the overall appearance making it harder to recognize faces.

The remaining of the paper is organized as follows: Second section presents a review of recent advances in face recognition techniques. New approach of large-scale face recognition in the wild is given in the third section. In the fourth section, an extension of experimental results is present. Followed by presentation of the main features and some specific application areas of PubFace app and finally the last section concludes our paper.

2 **Related Works**

In this section, we draw up a state-of-the-art review of data augmentation algorithms and the face recognition methods giving good results.

Big data: Recently, the number of face images has been growing exponentially on social network such as Facebook and Twitter. As an example, the director of Facebook AI Research Yann LeCun has said, "almost 1 billion new photos were uploaded each day on Facebook in 2016" [4]. Large face datasets are important for deep CNNs [5]. However, to build large dataset by downloading the images from search engine is very difficult and most financially challenging. One way to get around a lack of large face datasets is to augment the data.

Data augmentation: To train face recognition systems based on deep convolutional neural networks, very large training sets are needed with millions of labeled images. However, large training datasets are not publicly available and very difficult to collect. In this work, we a method to generate very large training datasets of synthetic images by compositing real face images in a small dataset collected from social networks. There are many Smart approaches to augmenting the size of the training set 10-fold or more. The popular augmentation methods include simple, geometric transformations such as oversampling [6], mirroring [7, 8], rotating [9] the images, Hairstyles synthesis [10], Glasses synthesis [11], 3D Face Reconstruction [12], Illuminations synthesis [12].

State-of-the-art face recognition: Traditional feature extractors such as Gabor-Zernike features [13], HOG [14, 15], SIFT [16] produce a good results on controlled conditions (constrained environments) as represented in FERET benchmark [17]. However, the recognition performances of theses representations may decreased dramatically in the wild, which represented in LFW [18]. This is because these features cannot improve the robustness to visual challenges such as pose, illumination and expression ... etc. In light of these nuisance factors, deep CNN feature extractor obtained by concatenating several linear and non-linear operators replaced conventional features extractors. These features demonstrated their potential by produce promising face recognition rates in the wild. A popular approach of this class of methods is DeepFace [19], which using a CNN architecture trained on a dataset of four million images spanning 4000 subjects. This approach achieve excellent recognition

accuracy near human visual system in the LFW benchmark [18]. This work is extended by the DeepId series of papers [20, 21] by using multiple CNNs [22]. Other interesting approach are being proposed [23–25].

3 The Proposed Method

In this section, we present our contributions to improve face recognition performance in the wild, particularly in terms of pose and illumination invariant.

3.1 Data Augmentation

In this sub-section, we describe the process used to build a large synthetic face dataset. The different steps of this process are summarized below in Fig. 1:

Step 1: Select list of some names of public figures.
Step 2: Filtering a list of candidate identity names.
Step 3: Face images were detected using a robust face detector.
Step 4: Building a small dataset by collecting some photos for each identity.
Step 5: Increase the size of the small dataset building (in Step 4) by using The proposed method in [26].

Fig. 1. Main stages of the dataset building process

In the first time, we select list of some names of public figures (football players, actors and politics figures) for obtaining their faces images and informations via Facebook Graph Search. Next, we apply the filtering process in order to reduce the list of identities. Then, a robust face detection is applied [27–30]. Finally, we apply the smart augmented approach [12, 26] to increase the size of our dataset. In this manner, a final list of four mille public figures names is obtained. We call these images as Puball-dataset, which all the images have the size of 152×152 pixels. Table 1 gives some statistical information on the larger face datasets public and private.

Dataset	Identities	Images
Facebook	4,030	4.4 M
Google	8 M	200 M
MegaFace	690,572	1.02 M
CASIA	10,575	494,414
VGG Face	2,622	2.6 M
LFW	5,749	13,233
CelebFaces	10,177	202,599
Chen et al.	2,995	99,773
Puball-dataset (ours)	4000	5 M

Table 1. Dataset comparisons

In the next sub-section, we present the deep CNN architecture adopted and their training process used in our experiments.

3.2 Network Architecture and Training

We use the VGGNet, off-the-shelf deep models of [31], originally trained on the ImageNet, large-scale image recognition benchmark (ILSVRC) [32]. We fine-tuned this CNN model on our training dataset. The input to deep CNN architecture adopted is a RGB face image $(3 \times 96 \times 96)$: As shown in Fig. 2 the deep architecture that we have used for representing faces in images consists of many function compositions, or layers, followed by a loss function. The loss function measure how accurately the neural network classifie a face image. It comprises more than 40 layers, each linear operator followed by spatial batch normalization (SBN) and one or more non-linearities such as ReLU and max pooling. Input image of this architecture is passed through a series of convolution filters and non-linear projections to obtain the identity classification. This process is serially repeated several times giving them their popular "Deep" identity. A rectification layer (ReLU) follows all the convolution layers. The last three blocks are called Fully Connected (FC); they are the same as a convolutional layer. The resulting vector is passed to a classification layer to compute the class posterior probabilities.

image	:	:	:
conv {3, 64, 3,3, 1,1, 1,1}	conv (128, 256, 3,3, 1,1, 1,1)	conv (512, 512, 3,3, 1,1, 1,1)	conv (512, 512, 3,3, 1,1, 1,1)
sbn(64)	sbn(256)	sbn(512)	sbn(512)
relu(True)	relu(True)	relu(True)	relu(True)
conv (64, 64, 3,3, 1,1, 1,1)	conv (256, 256, 3,3, 1,1, 1,1)	conv (512, 512, 3,3, 1,1, 1,1)	maxpool(2,2, 2,2)
sbn(64)	sbn(256)	sbn(512)	Fc(4608)
relu(True)	relu(True)	relu(True)	Fc(4608)
maxpool(2,2, 2,2)	conv (256, 256, 3,3, 1,1, 1,1)	maxpool(2,2, 2,2)	Fc(1024)
conv (64, 128, 3,3, 1,1, 1,1)	sbn(256)	conv (512, 512, 3,3, 1,1, 1,1)	
sbn(128)	relu(True)	sbn(512)	
relu(True)	maxpool(2,2, 2,2)	relu(True)	
conv (128 128, 3,3, 1,1, 1,1)	conv (256, 512, 3,3, 1,1, 1,1)	conv (512, 512, 3,3, 1,1, 1,1)	
sbn(128)	sbn(512)	sbn(512)	
relu(True)	relu(True)	relu(True)	
maxpool(2,2, 2,2)	:	÷	

Fig. 2. CNN architecture adopted in our approach

4 The Experiments and Tests

The performance of the proposed approach is assessed by conducting experiments on the well-known LFW dataset, which is described briefly below. In addition, we compare our approach with competitive supervised methods and current best commercial system. The receiver operating characteristic curves (ROC) is used to evaluate the performance of our proposed approach.

4.1 Labeled Face in the Wild

The dataset contains 13,233 images of 5,749 people downloaded from the Web. This database, cover large variations including different subjects, poses, illumination, occlusion etc. Two views are provided to develop models and validate on one view and finally test on another. For evaluation, we have using the standard protocol which defines 3,000 positive pairs and 3,000 negative pairs in total and further splits them into 10 disjoint subsets for cross validation. Each subset contains 300 positive and 300 negative pairs, portraying different people.

4.2 Results and Discussion

We present the average ROC curves for them in Fig. 3. In addition, we compare the mean accuracy of the proposed approach with some methods which achieve state of the art and other commercial systems. The results are summarized in Table 2:



Fig. 3. ROC Curve of our proposed approach on LFW dataset

Method	Mean accuracy
DeepFace [19]	97.35%
DeepID2 [20]	95.43%
Yi et al. [33]	96.13%
Wang et al. [34]	96.95%
Human [35]	97.53%
Our Proposed approach	98.12%

Table 2. Accuracy of different methods on the LFW dataset.

It can be seen from the Table 2 that our approach performs well comparably to other methods and commercial systems. Our proposed method achieve a good results on LFW dataset, which contains faces with full pose, illumination, and other difficult conditions. It is robust, especially in the presence of large head pose variations.

5 PubFace Application

Having looked at the proposed approach of face recognition in the wild, we look at features and some of specific applications of PubFace app that we have developed: The number of facebook users in morocco is more than 12,000,000. PubFace is could scan billions of Facebook profile images in real time. Through a database, that we have building by followed the same process presented in the Sect. (3.1). This dataset called Pub-dataset, include approximately half of adult photos downloaded from Facebook, without their knowledge or consent, in the hunt for suspected criminals. Besides, PubFace is able to link most faces photos (even from a side view as well as when the person is directly facing the camera in the picture) with a profile on the social network. So, PubFace app will tell you who it is?

PubFace app maybe used in the context of:

City surveillance: Large cities require more resources to handle threats such as vehicle theft, pickpockets, assaults, gang violence and shootings. PubFace app can improve overall public safety by reducing response times and providing law enforcement agencies with the ability to handle emergencies in a more effective way.

Photo Organizing: With the rapid growth in the personal digital content thanks to the smart phones, there is a increase need for automatic organization to cluster picture collections based on person identities.

6 Conclusion

In this paper, we have presented a new approach of large-scale face recognition in the wild. Our new approach based on deep learning was trained on Puball-dataset (Sect. 3) and evaluated on LFW dataset. Experimental results demonstrate that the performance of the proposed approach is much better than some methods, which achieve state of the art and other commercial systems. Moreover, we have presented briefly the main features and some specific areas of application of our PubFace app, which mainly could help reducing crime by making everyone identifiable.

References

- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Den, Van, Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. Nature 529(7587), 484–489 (2016)
- Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al.: Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016)
- Assael, Y.M., Shillingford, B., Whiteson, S., de Freitas, N.: LipNet: End-To-End Sentence-Level Lipreading. https://arxiv.org/abs/1611.01599 (2016)
- 4. https://www.youtube.com/watch?v=vlQomVlaNFg&t=317s

- 5. Masi, H., Tran, I., Leksut, J.T., Hassner, T., Medioni, G.G.: Do we really need to collect millions of faces for effective face recognition? CoRR, abs/1603.07057 (2016)
- 6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Neural Information Processing Systems, pp. 1097–1105 (2012)
- Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: Proceedings British Machine Vision Conference (2014)
- 8. Yang, H., Patras, I.: Mirror, mirror on the wall, tell me, is the error small? In: Proceedings Conference on Computer Vision and Pattern Recognition (2015)
- 9. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings International Conference on Computer Vision (2015)
- Liu, J.X., Liu, S., Xu, H., Zhou, X., Yan, S.: Wow! you are so beautiful today! ACM Trans. Multimedia Comput. Commun. Appl. 11(1s) (2014). 20
- 11. Wen, Y., Liu, W., Yang, M., Fu, Y., Xiang, Y., Hu, R.: Structured occlusion coding for robust face recognition. Neurocomputing **178**, 11–24 (2016)
- Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., Gao, W.: Efficient 3d reconstruction for face recognition. Pattern Recogn. 38(6), 787–798 (2005)
- Ouanan, H., Ouanan, M., Aksasse, B.: Gabor-zernike features based face recognition scheme. Int. J. Imaging Robot. 16(2), 118–131 (2015)
- Dèniz, O., Bueno, G., Salido, J., De la Torre, F.: Face recognition using histograms of oriented gradients. Pattern Recognit. Lett. 32(12), 1598–1603 (2011)
- Ouanan, H., Ouanan, M., Aksasse, B.: Gabor-HOG features based face recognition scheme. TELKOMNIKA Indonesian J. Electr. Eng. 15(2), 331–335 (2015)
- Liu, C., Yuen, J., Torralba, A.: SIFT flow: dense correspondence across scenes and its applications. Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-23048-1_2
- Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1090–1104 (2000)
- Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, TR 07-49 (2007)
- 19. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deep-face: closing the gap to human-level performance in face verification. In: IEEE CVPR (2014)
- 20. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification verification. In: Advances in Neural Information Processing Systems (2014)
- Sun, Y., Ding, L., Wang, X., Tang, X.: Deepid3: Face recognition with very deep neural networks. CoRR, abs/1502.00873 (2015)
- Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)
- He, K., Zhang, X., Ren, S.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
- 24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
- 25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: CVPR (2015)
- Ouanan, H., Ouanan, M., Aksasse, B.: Novel approach to pose invariant face recognition. Procedia Comput. Sci. 110, 434–439 (2017)

- Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision 57(2), 137– 154 (2004)
- Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proceedings of Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp. 2879–2886 (2012)
- Ouanan, H., Ouanan, M., Aksasse, B.: Facial landmark localization: Past, present and future. In: 4th IEEE International Colloquium on Information Science and Technology (CiSt), pp. 487–493 (2016)
- Ouanan, H., Ouanan, M., Aksass, B.: Implementation and optimization of face detection framework based on OpenCV library on mobile platforms using Davinci's technology. Int. J. Imag. Robot.[™] 15(4) (2015)
- 31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
- 32. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, S., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Li, F.F.: Imagenet large scale visual recognition challenge. IJCV (2015)
- 33. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
- Wang, D., Otto, C., Jain, A.K.: Face search at scale: 80 million gallery. arXiv preprint arXiv: 1507.07242 (2015)
- Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: IEEE International Conference on Computer Vision (ICCV), October 2009

Myface: Unconstrained Face Recognition

Hamid Ouanan^(IM), Mohammed Ouanan, and Brahim Aksasse

Department of Computer Science, M2I Laboratory, ASIA Team, Faculty of Science and Techniques, Moulay Ismail University, BP 509 Boutalamine, 52000 Errachidia, Morocco ham.ouanan@gmail.com, ouanan_mohammed@yahoo.fr, baksasse@yahoo.com

Abstract. Face verification in unconstrained images, remains a challenging problem. Many works have been proposed to solve this problem. However, the performance gap existing between the human visual system and machines in face recognition remain important. This paper makes two contributions: firstly, for improving face recognition in the wild, at least in terms of pose variations, we propose a method for aligning faces by employing single-3D face model as reference produced by FaceGen Modeller. Secondly, we developed a novel face representation technique based on Gabor Filters. The proposed approach relies on combination of Gabor magnitude and Gabor phase informations into an unified framework, which capable to surpass standard representations in the well-known FERET dataset.

Keywords: Face recognition · Gabor filter · Support vector machine · FERET

1 Introduction

In face verification, images are presented in pairs and the task is to verify if they belong to the same or different persons. Face verification has recently gained lot of popularity owing to few public benchmark datasets being available. Applications of this task are in search and authentication domains such as entertainment, human machine interaction, homeland security, and video surveillance, access control to user authentication schemes in e-commerce, e-health, and e-government services. There are many challenges in dealing with this applications listed such as variation in illumination, variability in scale, location, orientation and pose. Furthermore, facial expression, facial decorations, partial occlusion and lighting conditions change the overall appearance making it harder to recognize faces.

The main objective of this work is to propose a reliable framework insensitive to challenges listed above, in particular capable to "identify faces from a side view" as well as when the person is directly facing the camera in the picture to approach human-level performance in this domain. In summary, The novelty of this paper comes from: (i) an effective 3D face alignment module; (ii) effective representation for describing faces using Gabor Filters. We also investigate various approaches to effectively reduce their dimension while improving their performance further; and (iii) extensive performance evaluation studies.

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_9

2 Related Works

The conventional pipeline of a typical face verification system requires these steps: face detection, facial landmark detection, alignment, representation and classification. However, several papers focus on a few of these aspects in order to improve the overall system performance. In this work, we have focused on both alignment and the representation steps.

In this section, we briefly review some recent related works on face alignment and face representation in the context of face verification.

State-of-the-art face alignment: Aligning faces in under in-the-wild conditions is still a most difficult problem that has to account for many factors like non-rigid face expressions and pose. Recently, some techniques bearing capable to compensate for these difficulties, which can be roughly divided into two main categories: (i) part-based methods, which represent the face by using a set of local image patches, extracted around of the predefined landmark points and (ii) Holistic methods, which use the whole texture of face as representation. The most-well known techniques and produced good results: In the first category methods like Active Shape Models (ASMs) [2, 3] and Constrained Local Models (CLMs) [4]. In the second category, methods like Active Appearance Models (AAMs) [5] and 3D Deformable Models (3DMs) [6]. However, no complete solution is currently present in the context of face recognition in the wild because the accuracy of those detection and localization landmarks algorithms degrades as the yaw or pitch angle of the face increases.

In this work, we have developed a method using analytical 3D modeling of the face based on face landmark.

Brief review of recent face verification approaches: Representing face images has been an important topic in computer vision and image processing. The diversity of feature extraction methods is surprising. In this section, we look at some methods, which produced better performance over large-scale database like LFW and FERET face databases. In [7], the authors proposed a facial image representation giving better results on FERET database [8], this method rely on Gabor filters (GFs) and Zernike moments (ZMs), where GFs is used for texture feature extraction and ZMs extracts shape features, in other hand, a simple Genetic Algorithm (GA) is applied to select the moment features that better discriminate human faces under several pose and illumination conditions. Next, the augmented extracted feature vectors are projected onto a low-dimensional subspace using Random Projection [9] (RP) method. In [10], the authors proposed a regularization framework to learn similarity metrics for face verification in the wild. This method achieves a good results on the (LFW) database [1]. In [11], the authors proposed a joint Bayesian approach based on the classical Bayesian face recognition approach proposed by Baback Moghaddam et al. [12]. This approach achieved 92.4% accuracy on the LFW dataset. Another interesting approach is Fisher vector encoding performs well on LFW. However, the accuracy of those algorithms degrades on extreme poses of face like profile. This show the need of techniques capable to compensate large pose variation. In [13], the authors show that humans achieve 97.53% accuracy on the LFW.

3 The Proposed Method

In this section, we present our contributions to improve face recognition performance in the wild, particularly in terms of pose and illumination invariant.

3.1 Unconstrained Face Alignment

Aligning faces under the wild conditions is still a challenging problem due to many factors such as pose. Pose variation is the most difficult problem in face recognition with respect to other variations. This is evidenced from the evaluation report of existing face recognition algorithms suffer the most from pose variations. The ability to estimate the head pose of another person automatically and effectively is a human skill that presents a challenge problem of computer vision systems, since it is considered to be a key question to achieve its advantage in terms of non intrusive in many face recognition applications. In this section we investigate the use of a generic 3D shape model and facial keypoint detectors to direct the alignment process of faces in images. To this end, this chapter makes two contributions: firstly, we use the same 3D model as reference for synthesizing frontal face view for all query faces. This 3D face model is produced take the average of the 3D scans from the USF Human-ID database. Secondly, we estimate the head pose of query image by computing the projection matrix which is then used to back-project query intensities to the reference coordinate system. Finally we apply kalman filter to reject the images that our procedure failed to align and fallback to deep funneled versions for this images.

We begin by rendering this reference 3D model into a frontal view. We refer to this as the reference frontal view I_R (see Eq. 1) which serves as our reference coordinate system, 68 facial landmarks $p_i = (x_i, y_i)^T$ are detected in this image using the method of [13], selected for its accuracy in real world face photos. For each point detected we associate the 3D coordinates $(P_i = (X_i, Y_i, Z_i)^T)$ given a query image I_Q , it is processed by first running the Viola-Jones detector [14, 15]. We again use [16, 17] to detect the same 68 landmarks in I_Q , giving us points $(p'_i = (x'_i, y'_i)^T)$. Using these, we form correspondences (p_i, P_i) from 2D pixels in the query photo to 3D points on the model. We then compute specific 3 × 4 camera matrix C_M by selecting suitable intrinsic and extrinsic camera parameters, using a standard calibration method.

$$p' \cong C_M P$$
 (1)

$$C_M = A_M[R_M t_M] \tag{2}$$

$$C_Q = A_Q[R_Q t_Q] \tag{3}$$

Where A_M : is the intrinsic matrix, R_M : is the rotation matrix, and t_M : is the translation vector (Fig. 1).

89

Input: Query image I_Q , textured 3D face model, rendered frontal view of this model (I_R). Output: Frontalized Face Step 1: Facial feature points ($p_i = (x_i, y_i)^T$) detected in the query image I_Q . Step 2: Same facial feature points ($p_i = (x_i, y_i)^T$) will be detected in I_R and their correspondence points ($P_i = (X_i, Y_i, Z_i)^T$) on the surface of the model. Step 3: Seek 2D - 3D correspondences between points (p_i, P_i). Step 4: Estimation the query 3x4 camera matrix C_Q used to capture the query image I_Q . Step 5: Back-projection query intensities to I_R (equation 3). Step 6: Estimation of visibility due to non-frontal poses by symmetry. Step 7:Apply kalman filters & Fallback to deep funneled versions for images that our procedure failed to align.

Step 8: Final frontalized crop canonical view.

Fig. 1. An overview of the face alignment proposed method

3.2 Face Representation

The principal motivation to use Gabor Filters is biological, since Gabor like receptive fields have been found in the visual cortex of primates. Gabor filters can exploit salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristics. Considering these overwhelming capacities and its great success in face recognition.

3.2.1 Gabor Filters Construction

The frequency and orientation representations of Gabor filters are similar to those of the human visual system and they have been found to be particularly appropriate for texture representation [18]. Gabor filters have been widely used in pattern analysis applications. The most important advantage of Gabor filters is their invariance to illumination, rotation, scale, and translation. Furthermore, they can with stand photometric disturbances, such as illumination changes and image noise.

A 2D Gabor function g(x, y) and its Fourier transform G(u, v) are as follows:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j\omega x\right]$$
(4)

$$G(u,v) = \exp[-\frac{1}{2}(\frac{(u-\omega)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2})]$$
(5)

Where: $\sigma_u = \frac{1}{2\pi\sigma_x}$ and $\sigma_v = \frac{1}{2\pi\sigma_y}$

$$g_{mn}(x,y) = a^{-m}G(x',y')$$
 (6)

where $a \ge 1$; $x' = a^{-m}(x \cos s\theta + y \sin \theta)$ and $y' = a^{-m}(y \cos \theta - x \sin \theta)$,

for m = 0, 1, ..., M - 1 and n = 0, 1, ..., N - 1, M is the number of resolutions and N is the number of orientations.

3.2.2 Feature Extraction Using Gabor Filters

The feature extraction procedure can then be defined as a convolution operation of the face image I(x, y) with the Gabor filter G(u, v). The result of this operation is a complex image defined by the amplitude and the phase for each pixel of the image:

$$\psi_{u,v}(x,y) = I(x,y) * G_{u,v}(x,y)$$
(7)

Based on this equation the magnitude and the phase responses of the convolution operation can be computed as follows:

$$A_{u,v}(x,y) = \sqrt{\operatorname{Re}(\psi_{u,v}(x,y))^2 + \operatorname{Im}(\psi_{u,v}(x,y))^2}$$
(8)

$$\phi_{u,v}(x,y) = \arctan(\frac{\operatorname{Im}(\psi_{u,v}(x,y))}{\operatorname{Re}(\psi_{u,v}(x,y))})$$
(9)

The great number of Gabor-based face recognition approaches found in the literature rely solely on the magnitude information when constructing the Gabor face representation and discard the phase information of the convolution output image (Figs. 2 and 3).

Output: The validity of the Identity claim

Step 1: Gabor filter construction with bank of 40 filters.

Step 2: Gabor features derived from the Gabor filter magnitude response (similarly from the Gabor filter phase) are computed for all frequencies (u = 4) and

orientations (v = 8) (GMFR).

Step 4: Downsampling by a factor ($\rho = 64$) the computed GMFRs (similarly GPFRs).

Step 5: The downsampled GMFRs are normalized using an appropriate normalization procedure.

Step 6: The downsampled and normalized GMFRs (similarly GPFRs) in vector form are concatenated to form the augmented GMFR vector.

Input: Face Image (128×128) pixels.



Fig. 3. Block scheme of the proposed approach

3.2.2.1 Gabor Magnitude Face Representation (GMFR)

The magnitude face representation vector is computed by taking the following steps:

Despite the downsampling procedure, the size of the descriptors presented still reside in a very high-dimensional space. In this paper, the KFA technique [19] is applied to the augmented Gabor magnitude face and augmented Gabor phase representation vectors to obtain a compact representation based Gabor-Magnitude and a compact representation based Gabor-Phase. Then, we use the SVM classifier [20, 21] based on RBF kernel to classify GM + KFA feature and GP + KFA feature extracted from face images. Finally, we combine both matching score at the matching score level using the fusion scheme shown in Fig. 6. The accuracy δ of the method proposed is computed using the following expression: $\delta = (1 - \gamma)\delta_{GM} + \gamma\delta_{GP}$.

Where δ_{GM} denotes the accuracy obtained from Gabor magnitude features, δ_{GP} denotes the accuracy obtained from Gabor phase features and $\gamma \in]0,1[$ denotes the fusion parameter.

4 The Experiments and Tests

The performance of the proposed approach is assessed by conducting experiments on the well-known FERET dataset, which is described briefly below.

4.1 FERET Dataset

The FERET face image database is a result of the FERET program, which was sponsored by the US Department of Defense. It has become a standard database for testing and evaluating state-of-the-art face recognition methods. Color FERET contains totally, 11338 face images. For our experiments, we adopt the standard FERET evaluation protocol, where four different probe sets are employed for determining the recognition rates of our proposed approach:

- The **Fb** probe set, contains 1195 images exhibiting different facial variations in comparison to the gallery images.
- The **Fc** probe set, contains 194 images exhibiting different illumination conditions in comparison to the gallery images.
- The **DupI** probe set contains 722 images acquired between one minute and 1031 days after the corresponding gallery images.
- The **DupII** probe set contains 234 images acquired at least 18 months after the corresponding gallery images.

4.2 Results

For the experiments, we provide results not in the form of error rates, recognition rate, and the cumulative match characteristic (CMC) curves.

All run-times are reported on an Intel Core i5 CPU, 2.50 GHz, 4 GB RAM. The robustness, flexibility and speed of this novel face recognition system have been demonstrated through this above results in the probe set Fb. Similar results are obtained in the remaining probe sets: Fa, DupI, and DupII (Figs. 4, 5 and 6).



Fig. 4. CMC curve obtained with Fb probe of the FERET database.



Fig. 5. DET curve obtained with Fb probe of the FERET database.



Fig. 6. EPC curve obtained with Fb probe of the FERET database.

5 Conclusion

In this paper, we have presented a new face identification approach in the wild. Our new approach was evaluated on FERET dataset. Experimental results demonstrate that the performance of the proposed approach is much better than some methods, which achieve state of the art. In the future, we will extend this approach to video processing. We believe that our method will demonstrate competitive performance.

References

- Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, TR 07-49 (2007)
- Zhou, H., Lam, K.M., He, X.: Shape-appearance-correlated active appearance model. Pattern Recogn. 56, 88–99 (2016)
- Jeni, L.A., Cohn, J.F., Kanade, T.: Dense 3D face alignment from 2D video for real-time use. Image Vis. Comput. 58, 13–24 (2016)
- Cootes, T., Edwards, G., Taylor, C.: Active appearance models. TPAMI 23(6), 681–685 (2001)
- Chen, H., Gao, M., Fang, B.: An improved active shape model method for facial landmarking based on relative position feature. Int. J. Wavelets Multiresolut. Inf. Process. 15 (1), 1–14 (2017)
- Kim, K., Baltruaitis, T., Zadeh, A., Morency, L.P., Medioni, G.: Holistically constrained local model: Going beyond frontal poses for facial landmark detection. In: Proceedings of the British Machine Vision Conference (2016)
- 7. Ouanan, H., Ouanan, M., Aksasse, B.: Gabor-zernike features based face recognition scheme. Int. J. Imaging Robot.[™] 16(2), 118–131 (2015)
- Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1090–1104 (2000)
- Menon, A.K.: Random projections and applications to dimensionality reduction. Ph.d. thesis, School of Information Technologies, The University of Sydney, Australia (2007)
- Cao, Q., Ying, Y., Li, P.: Similarity metric learning for face recognition. In: Proceedings of the International Conference on Computer Vision, pp. 2408–2415. IEEE (2013)

- Chen, D., Cao, X., Wang, L., Wen, F., Sun, J.: Bayesian face revisited: A joint formulation. In: Proceedings of the ECCV, pp. 566–579 (2012)
- Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recogn. 33, 1771–1782 (2000)
- Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: IEEE International Conference on Computer Vision (ICCV), October 2009
- 14. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. 57(2), 137-154 (2004)
- Ouanan, H., Ouanan, M., Aksass, B.: Implementation and optimization of face detection frameworkbased on OpenCV library on mobile platforms using Davinci's technology. Int. J. Imaging Robot.[™] 15(4), 81–88 (2015)
- Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 2879–2886 (2012)
- Ouanan, H., Ouanan, M., Aksasse, B.: Facial landmark localization: Past, present and future. In: 4th IEEE International Colloquium on Information Science and Technology (CiSt), pp. 487–493 (2016)
- Shen, L., Bai, L., Fairhurst, M.: Gabor wavelets and general discriminant analysis for face identification and verification. Image Vis. Comput. 25(5), 553–563 (2007)
- Khellat-Kihel, S., Abrishambaf, R., Monteiro, J.L., Benyettou, M.: Multimodal fusion of the finger vein, fingerprint and the finger-knuckle-print using Kernel Fisher analysis. Appl. Soft Comput. 2016(42), 439–447 (2016)
- 20. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20(3), 273-297 (1995)
- Ma, Y., Guo, G.: Support Vector Machines Applications. Springer, Switzerland (2014). doi:10.1007/978-3-319-02300-7

Content-Based Image Retrieval Using Gabor Filters and 2-D ESPRIT Method

Youness Chawki^(III), Khalid El Asnaoui, Mohammed Ouanan, and Brahim Aksasse

M2I Laboratory, ASIA Team, Faculty of Science and Techniques, Computer Sciences Department, Moulay Ismail University, Boutalamine, BP 509 Errachidia, Morocco youness.chawki@gmail.com, khalid.elasnaoui@gmail.com, ouanan_mohammed@yahoo.fr, baksasse@yahoo.com

Abstract. In this paper, we propose a novel descriptor for texture representation in Content-Based Image Retrieval (CBIR). Indeed, Gabor filters are the most used methods in this representation, but their major disadvantages are the choice of values and number of frequencies and orientations. To remedy this problem, our new method consists in extracting from the image itself with good precision the frequencies and the orientations; using the 2-D ESPRIT method (Estimation of Signal Parameters via Rotational Invariant Techniques); that will be injected in the Gabor filters. Our approach has been applied on the Coil_100 database, and the obtained results show the effectiveness and the rapidity of our technique compared to Gabor filters.

Keywords: CBIR \cdot Texture feature \cdot Gabor filter \cdot Frequency content \cdot Spectral analysis \cdot 2-D ESPRIT

1 Introduction

Content-Based Image Retrieval (CBIR) is searching of images from large image databases based on their visual contents. CBIR is a research area that is required these last few years to overcome the drawbacks of Text-Based Image Retrieval (TBIR). It consists in representing every image by a set of visual features such as color, shape and texture.

Texture is the second visual attribute widely used in image characterization. It can fill the gaps that the color is unable to do, especially when the color distributions are very close. More specifically, the texture can be viewed as a set of pixels (grayscale) spatially arranged in a number of spatial relationships, and creating a homogeneous region. Thus, several approaches and models [1, 2] have been proposed for modeling the texture, which we quote: statistical approaches, geometric approaches and the frequency approaches. For the latter, the Gabor filters are the most known and most used method.

Dennis Gabor was the first that introduced the Gabor filters in 1946 [3]. These later prove to be an interesting tool for texture analysis and they are widely adopted in the literature. The advantage of these filters is that their functioning is close to the human visual treatments, and they have the advantage of being programmable in frequency and

in orientation. Indeed, Gabor filters find their place in several areas such as: segmentation [4], pattern recognition [5, 6], classification [7, 8], content-based image retrieval [9, 10].

The 2-D ESPRIT method [11–13] is the most famous and the used method in bidimensional frequency estimation. It provides the pair frequency and its corresponding orientation automatically with good precision.

In this paper, we propose a new approach for CBIR systems based on a combination of the 2-D ESPRIT method and Gabor filters. Indeed, after extracting the frequencies and the orientations from images itself, these parameters will be injected in the Gabor filter.

The following sections are arranged as follows: Sect. 2 presents our new approach. Subsequently, in Sect. 3, the experimental results are presented. Finally we conclude our work with conclusions and perspectives.

2 The Proposed Method

We propose a new method for content-based image retrieval by combining the 2-D ESPRIT method and Gabor filters. In this case, we extract from the image itself the frequencies and the orientations that will be injected in the Gabor filter as input parameters.

The 2-D ESPRIT [12] is the most used spectral analysis method. The extracted frequency content is in the form of two components: one component is on the first frequency axis and the other component is on the second frequency axis. Figure 1 depicts the relationship between the frequency components f_{1i} , f_{2i} , module frequency F_i and the orientation θ_i :





Thus the frequency module is defined as follow:

$$F_i = \sqrt{f_{1i}^2 + f_{2i}^2} \tag{1}$$

The corresponding rotation angle θ_i is giving by:

$$\theta_i = \arctan(\frac{f_{2i}}{f_{1i}}) \tag{2}$$

Therefore the 2-D ESPRIT provides the frequencies and orientations that will be injected in the Gabor filters as input parameters. The number of frequencies and the orientations extracted are the same. In our case, we choice 6 frequencies and 6 orientations.

Gabor filters are multi-channel filtering techniques that allow the description of textures localized in frequency and orientation. In other words, the characteristic calculations are operated on each of the pixels. These contain the intensity variations on a smaller scale of frequency and orientation. The Gabor function can be written as follows:

$$G(x, y, \theta, F) = \exp(-\frac{1}{2}(\frac{x_{\theta}^2}{\sigma_x^2} + \frac{y_{\theta}^2}{\sigma_y^2}))\cos(2\pi F x_{\theta})$$
(3)

Where:

$$x_{\theta} = x \cos \theta + y \sin \theta$$

$$y_{\theta} = -x \sin \theta + y \cos \theta$$
(4)

 θ is the orientation, F the frequency and σ_x (respectively σ_y) the standard deviation of the Gaussian according the x axis (y axis respectively).

After constructing the Gabor filters, they are applied to a MxN image in order to extract the average and the variance. The average and variance have the following expressions respectively:

$$\mu_m = \frac{\sum_{x} \sum_{y} |G_m(x, y)|}{MxN}$$
(5)

$$\sigma_m = \sqrt{\frac{\sum\limits_{x} \sum\limits_{y} \left(\left| G_m(x, y) - \mu_m \right| \right)^2}{MxN}}$$
(6)

Thus the vector descriptor is as follow:

$$DV_{R} = \left[R_{\mu 1} \ R_{\sigma 1} \ R_{\mu 2} \ R_{\sigma 2} \ R_{\mu 3} \ R_{\sigma 3} \ R_{\mu 4} \ R_{\sigma 4} \ R_{\mu 5} \ R_{\sigma 5} \ R_{\mu 6} \ R_{\sigma 6} \right]$$
(7)

We use the RGB color space in order to extract the frequencies and the orientations of each layer. The color image is presented in Fig. 2:

97



Fig. 2. Color image in RGB space.

Therefore, the frequencies and the orientations extracted from the image in the R_layer will be injected in the Gabor filter for extracting the average and the variance. The same for the G_layer and the B_layer. The vector descriptor of the G_layer and the B_layer is given as follow respectively:

$$DV_G = \begin{bmatrix} G_{\mu 1} & G_{\sigma 1} & G_{\mu 2} & G_{\sigma 2} & G_{\mu 3} & G_{\sigma 3} & G_{\mu 4} & G_{\sigma 4} & G_{\mu 5} & G_{\sigma 5} & G_{\mu 6} & G_{\sigma 6} \end{bmatrix}$$
(8)

$$DV_B = \begin{bmatrix} B_{\mu 1} & B_{\sigma 1} & B_{\mu 2} & B_{\sigma 2} & B_{\mu 3} & B_{\sigma 3} & B_{\mu 4} & B_{\sigma 4} & B_{\mu 5} & B_{\sigma 5} & B_{\mu 6} & B_{\sigma 6} \end{bmatrix}$$
(9)

The final Descriptor Vector is as follow:

$$DV = \left[DV_R \ DV_G \ DV_B \right] \tag{10}$$

The similarity measure between a query image Q and a target image T is carried out by calculating, for each value of the vector, of the Swain similarity [14] D(Q, T) defined by:

$$D(Q,T) = \frac{\sum_{i=1}^{36} |DV_Q - DV_T|}{\sum_{i=1}^{36} DV_Q}$$
(11)

3 Experimental Results

To evaluate the method presented in this article, we have tested our approach on the Coil-100 database [15] including 100 classes, 72 samples each. All images are color and in PNG format, each is 128 by 128 pixels. Figure 3 shows the objects used in Coil-100 database.


Fig. 3. Objects used in Coil-100 [15].

In the implementation phase, we have used a computer, Processor: Intel® Core (TM) 2 CPU T5870 @ 2.00 GHz, 2 Go RAM, Windows 7.

We have developed a user graphic interface for displaying the results. We just display the 14 images similar to the query image.

Figure 4 shows the results of query image (obj2__0), our approach has returned 14 relevant similar images while Fig. 5 illustrates another example of query image (obj__5) with 14 images similar including two not relevant images (obj92__50 and obj92__55).



Fig. 4. Similar images for Obj2__0 query image.



Fig. 5. Similar images for Obj5_0 query image.

Our method is not only effective (the average precision is 90, 82%, but it is also very fast. Indeed, the Table 1 illustrates the elapsed time of query image of our approach with Gabor filter. For the latter, we chose the values of these input parameters as follows: $F_i = \{0.25, 0.5\}$ and $\theta_i = \{0, 45, 90\}$.

T 11 1	F1 1		c	•
Table I.	Elapsed	time c	of auerv	images.
			1	

	Gabor filter	Proposed method
Elapsed time (s)	192.96	70.63

To evaluate the efficiency of our system, we are interested to calculate the precision and the recall given as follow:

$$\operatorname{Recall} = \frac{R_{a}}{R}$$
(12)

$$Precision = \frac{R_a}{A}$$
(13)

Where

- R_a: Number of relevant images in the set of responses.
- R: Number of relevant images in the image database.
- A: Number of images in the set of responses.

Figures 6 and 7 show the precision average and the recall average respectively of our approach. We can see that our method is more accurate.



Fig. 6. Precision curve of proposed approach compared to Gabor filter.



Fig. 7. Recall curve of proposed approach compared to Gabor filter.

4 Conclusion and Perspectives

In this paper, we have proposed a new technique for Content-Based Image Retrieval systems. Our approach was able to overcome the problem of choice of values of the input parameters of Gabor filters. These later fuelled and powered by the frequencies and orientations extracted from the image itself and given by the 2-D ESPRIT method. The obtained results show that our method is very efficient ad very fast compared with classical Gabor filters.

In perspectives, we focus applying our approach to others images database and then to combine content frequency feature with color and shape feature.

References

- Haralick, R.M.: Statistical and structural approaches to texture analysis methods. Proc. IEEE 67, 786–804 (1979)
- Zhang, J., Tan, T.: Brief review of invariant texture analysis methods. Pattern Recogn. 35, 735–747 (2002)
- 3. Gabor, D.: Theory of communication. J. Inst. Electr. Eng. 93, 429-457 (1946)
- Zhang, J., Tan, T., Ma, L.: Invariant texture segmentation via circular Gabor filters. In: 16th International Conference on Pattern Recognition (ICPR), pp. 901–904. IEEE Computer society, Quebec (2002)
- Avila, C.S., Reillo, R.S.: Two different approaches for iris recognition using Gabor filters and multiscale zero-crossing representation. Pattern Recogn. 38, 231–240 (2005)
- Thangairulappan, K., Jeyasingh, J.B.V.: Face representation using combined method of Gabor filters, wavelet transformation and DCV and recognition using RBF. J. Intell. Learn. Syst. Appl. 4, 266–273 (2012)
- Arivazhagan, S., Ganesan, L., Priyal, S.P.: Texture classification using Gabor wavelets based rotation invariant features. Pattern Recogn. Lett. 27, 1976–1982 (2006)
- Rahmadwati, Naghdy, G., Ros, M., Todd, C., Norahmawati, E.: Cervical cancer classification using Gabor filters. In: Healthcare Informatics, Imaging and Systems Biology (HISB), First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology, pp. 48–52, IEEE Computer Society, Washington (2011)
- Bourkache, N., Sidhom, S., Laghrouche, M.: Apprentissage numérique pour la recherche d'informations en imagerie médicale: Modélisation des filtres de Gabor. In: International Symposium ISKO-Maghreb 13, Concepts and Tools for Knowledge Management (KM), November, Marrakech (2013)
- 10. Agarwal, M., Maheshwari, R.P.: Content based image retrieval based on log Gabor wavelet transform. Adv. Mater. Res. **403–408**, 871–878 (2012). Trans Tech Publications
- Zhou, Y., Feng, D.-Z., Liu, J.-Q.: A novel algorithm for two-dimensional frequency estimation. Sig. Process. 87, 1–12 (2007)
- 12. Rouquette, S., Najim, M.: Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods. IEEE Trans. Signal Process. **49**, 237–245 (2001)
- 13. Berthoumieu, Y., El Ansari, M., Aksasse, B., Donias, M., Najim, M.: A 2-D Robust high resolution frequency estimation approach. Sig. Process. **85**, 1165–1188 (2005)
- 14. Swain, M., Ballard, D.H.: Color indexing. Int. J. Comput. Vision 32, 11-32 (1991)
- 15. Deselaers, T.: Features for image retrieval. Diploma Thesis, RWTH Aachen University. Aachen, Germany, December 2003

A Retirement Pension from a Supply Chain Side: Case of the Moroccan Retirement Pension

Houda Mezouar^(™) and Abdellatif El Afia

Mohammed V University/ENSIAS, Rabat, Morocco houda.mezouar@gmail.com, abdellatif.elafia@gmail.com

Abstract. A retirement is on average 40 years of contributions and 25 years of pension: a lifetime legacy, a personal right that deserves to be given importance. The inability of current schemes to finance tomorrow's pensions, reforms, and the complexity of careers ... all these reasons motivated us to shed light on the processes of managing retirement. Therefore, we studied a specific case "The Moroccan pension scheme", and we approached it from a Supply chain point of view, with the objective of studying and simulating this supply chain, which has as major challenge "ensuring the continuity between the salary and the pension".

Keywords: SCOR · BPMN · Supply chain modeling · Retirement pension

1 Introduction

Supply chain management (SCM) is the activity of improving flow management within the company and its environment. In other words, it is the management of all the resources, means, methods, tools and techniques intended to control as effectively as possible the global supply chain of a product or service to the final consumer. A critical part of any organization's endeavor to compete in today's market is the design of their supply chain. This is true not only for product companies, but also for service companies [1]. In addition, the concept of SCM has been applied in manufacturing sector by various researchers and the results obtained are found to be very fruitful in terms of cost optimization and increased quality level. Despite of being the driving force of economic growth of every developed nation [2], the service operations have unique characteristics that are not found in manufacturing [3]. In fact, similar to the production of manufacturing goods, services production involves the collaboration of several actors; the service providers, the suppliers of other services or resources needed for the design and delivery of these services and the service clients, all working together to co-produce value in complex value chains. However, due to certain differences in manufacturing and services [1], there are inherent difficulties in developing standard models for services due to the peculiarities of service exchanges, which also contributed to the dearth of research in the area [4]. Services are considered as supply chain processes that are balanced around the capacity of the firm through the upstream sourcing processes [2]. A Service Supply Chain (SSC) has been researched and defined by various authors and subsequently

applied to various service sectors. Reference [5] defined the SSC management by the ability of the company/firm to get closer to the customer by improving its supply chain channels. The SSC will include responsiveness, efficiency, and controlling. While for [6], the SSC is the network of suppliers, service providers, consumers and other supporting units that performs the functions of transaction of resources required to produce services; transformation of these resources into supporting and core services; and the delivery of these services to customers. Reference [7], explained that the SSC is a service-network that reorganizes different service entities in order to satisfy customers' require. It is done by using modem management technology to break down and rebuild a system, which considers customers' demands as starting point and takes a complex service or an Integrated Service Package as a series of process in service, when the service-industries are developed to some extent. The concept can also be defined as follow: an integration of a series of entities (individual person, organization, enterprise) to provide personalized service directly or indirectly [8], in agreement with the authors of [9] how defined it by 'the supply and demand chain of service which integrates the service resources using new technologies and management models'. Moreover, we can say that the SSC management is an integrated management mode of service information, service processes, service capacity, service performance and service funds from the initial service supplier to the ultimate customer in the process of producer service outsourcing [10]. In Morocco as in most countries, the retirement is one of the services that attract the citizens' attention since the beginning of their careers. Much more than a matter of financial planning and actuarial study, managing the retirement sector's challenges requires a strategic analysis. Upon reaching retirement age, the citizen begins receiving a pension from the MPF instead of a salary from his recruiter. This pension is established following a file filing by the citizen administration, whose processing gives in output a payment. The whole challenge is to offer the right pension (correct calculation) to the good citizen (verification of rights) at the right time (the month following the suspension of his salary). Actually, any delay in a file processing represents months without income for a citizen. Therefore, to deal with this challenge, in this work, we analyze the management processes of the retirement pension offered by the Moroccan Pension Fund (MPF) by considering its service supply chain. For this, a modeling methodology has been respected, giving us a mapping of the studied supply chain, which will be the subject of a simulation work for optimization.

The remainder of this paper is organized as follows. Section 2 gives an overall overview about researches that deal with the pension schemes and the supply chain modelling, and then follows Sect. 3, which describes the Moroccan pension fund and gives general ideas about this organization. The modelling of the retirement pension supply chain using a methodology that combines SCOR and BPMN is presented in Sect. 4. This paper is concluded in Sect. 5.

2 Literature Review

2.1 Pension Schemes

Several authors, and from different point of views, have faced the retirement field. Reference [11] discusses the issue of retirement security of citizens in advanced age. It deals with the possibility of covering all inhabitants in the region. Due to the difficult changing of the pension scheme as a whole, the possibilities of the regional pension scheme are described and analyzed by the real conditions in the Czech Republic. The result is the defining of the conditions for the public pension fund to function, where the second pillar pension scheme seems optimal. A specific solution also contains an actuarial model of the functioning of the regional pension fund. Other researchers investigated the relationship between financial optimism and non-participation in pension schemes. within this framework, [12] show that financial optimism reduces the probability of employees joining employer run pension schemes and also the probability of the selfemployed subscribing to private pension plans. Their research suggests that both employed and self-employed individuals who are financially optimistic could face the very negative consequences of pension shortfall and low pensions income when they retire. Whereas the authors of [13] studied a generalized multi-period mean-variance portfolio selection problem within the game theoretic framework for a defined-contribution pension scheme member. While [14] examined the effect of the New Rural Pension Scheme (NRPS) on the labor supply behavior of the elderly in rural China. Using pooled data from two waves of the China Health and Retirement Longitudinal Study and an analytical framework of combination of regression discontinuity design and difference in difference method, we find no evidence that pension receipt from the NRPS program does significantly induce the elderly to withdraw from the labor market. Other researchers have even applied concepts from bounded rationality theory in this field, just as [15] how developed an integrative model to understand how pension scheme structure and pension scheme communication, impact pension participation and contribution rates at organizational level. Moreover, since the pension schemes have changed in both concept and detail with significant consequences for beneficiaries since the 1980s, [16] explored one of the major changes: the migration from defined benefit (DB) to defined contribution (DC) pension schemes focusing on this change's interface with accounting. In exploring this shift from DB to DC schemes, the authors used a critical perspective to reflect on this interface including how the change is accounted for in corporate reporting narrative. And in the framework of the redesign of defined benefit pension schemes, [17] presented the first study that quantified the redistributive effects of a rule change by a real world scheme (the Universities Superannuation Scheme, USS) where the sponsor underwrites the pension promise. They find that the pre-October 2011 scheme was not viable in the end, while the post-October 2011 scheme is probably viable in the end, but faces medium term problems. And in order to analysis the relationship between the level of a return guarantee in an equity-linked pension scheme, and the proportion of an investor's contribution needed to finance this guarantee, the authors of [18] have considered three types of schemes: investment guarantee, contribution guarantee and surplus participation. They find a negative (and for two contract specifications

concave) relationship between the participation in the surplus return of the investment strategy and the guarantee level in terms of a minimum rate of return. Furthermore, the introduction of the possibility of early termination of the contract (e.g. due to the death of the investor) has no qualitative and very little quantitative impact on this relationship.

We note that in these different works, the subject is modeled in an actuarial mathematical way, in this work we consider this service supply chain whose product is a retirement pension.

2.2 Supply Chain Modelling

The modelling is an activity of the brain [19], and it is for the scientific community a primary concern. In addition, according to [20], the lack of integration of supply chain systems with organization systems is seen as a cause of disruptions of the supply chains by several works that how the authors justified the importance of the modelling is when designing supply chain systems. Several supply chain models exist, [21] presents diverse approaches and models of the drug supply chain management, the same author chooses in [22] the Colored & Timed Petri Net to model the drug supply chain. while [23] shows a systematic review of the quantitative and analytical models for managing supply chain risks, and the authors used network analysis tools to analyze quantitative Supply Chain Risk Management. Reference [24], detailed the huge number of the supply chain modelling methods, the authors explained the two modelling paradigms, centralized and distributed. In other studies, the modelling approaches are divided into modelling type (mathematical, simulation, and multi-agent) and modelling settings (linear, integer, dynamic and stochastic problem settings) [25]. Reference [26], consider that the main role of the modelling techniques is to make the supply chain sufficiently understandable to ensure product quality; and that the modelling methods have a crucial role to play in pharmaceutical process development, by supplementing experimental studies in evaluation process sensitivity and operability. Reference [27], focused in the ability of the modelling methods to describe the real world phenomena, and divided the supply chain models into four types: stochastic model, deterministic model, hybrid model, IT-driven model. Among the most references used in this context, we find the Supply chain operation reference (SCOR) created in 1996; to understand, describe and evaluate supply chains [28]. It provides a common framework, standard terminology, common metrics, and best practices [28]. According to reference [29], the SCOR model consists of processes in three hierarchical levels. The first one is a process definitions level that consists of five different process types: Source, Make, Deliver, Return, and Plan. The second one is the process type level that divides the processes Make, Deliver and Source into three subcategories: make-to-stock products, make-to-order products, and engineerto-order products. The Plan process contains the overall process Plan supply chain (P1) and one planning process for each of the other level 1 processes. The Return process is divided into two processes; Source Return and Deliver Return. The last one is a process category level that describes the underlying processes of the second level. It is recommended to add a fourth level to represent tasks that describe the third level activities, which will be unique to each organization [30]. Moreover, among the most frequently used notations, we found the Business Process Model and Notation (BPMN), an Object

Management Group (OMG) standard. This notation was specially developed to facilitate human interpretation by both technical and non-technical personnel, and improve process execution [28].

3 Moroccan Pension Fund (MPF)

3.1 MPF Presentation

Created by the Dahir of March 2 1930, the MPF is a public institution with legal personality and financial autonomy. The MPF is subject to the dual technical and financial supervision of the Ministry of Economy and Finance. Its vocation is to work in the field of social protection and more particularly its component relating to pension benefits. The Fund also provides supplemental services for the population it covers [31].

3.2 MPF Missions

The MPF ensures the administrative management of pension schemes and benefits for third parties; the conducting actuarial and prospective studies of the managed schemes; and the financial management of pension schemes' reserves. In terms of administrative management, the MPF manages the civil pensions scheme for trainee and state officials, local government staff and trainees and the staff of certain institutions and public bodies; the military pensions scheme for the Royal Armed Forces personnel, and managerial the Auxiliary Forces personnel; and the supplementary and optional pension scheme "ATTAKMILI". The MPF also manages, regulated or contracted services on behalf of third parties, namely the non-contributory schemes (civil and military invalidity pensions, pensions and allowances of former Resistance members and former members of the Liberation Army...); the benefits or prepayments on behalf of the Ministry of Economy and Finance (withholding tax on income, exceptional pensions, emergency allowances Régie Tabacs); the prepayments on pensions under the Compulsory Health Insurance payable to the National Fund of Social Insurance Schemes; the prepayments in respect of mutual health transferred to the Mutual of the Royal Armed Forces, as well as the prepayments to the benefit of a certain number of bodies such as: the General Mutual of the National Education; Barid Al-Maghrib; National Security; and the conventional withholding tax for the benefit of financial institutions and associations of social welfare.

In terms of financial management, the MPF manages the various contingency funds made up of the financial surpluses generated by the schemes under management. They are invested in the capital markets according to the terms of the Dahir relating to the MPF and the decree of the Minister of the Economy and Finance that determines its use and prudential rules [31].

4 Modelling of the Retirement Pension Supply Chain

4.1 The Used Modelling Methodology

Apart from its capacity to guide actions as simulation or monitoring, a modelling approach has an explanation and description capacity (e.g. model for interpreting observed behaviors or share a universe of discourse) [32]. The ability to combine methods, tools and approaches of different sources and to arrange them in a specific purpose is the main role of a modeling methodology [33]. As part of our purpose to have an overall overview of retirement management processes in Morocco, we use, in this work, the suggested modeling methodology presented in [34] that combined SCOR and BPMN. This methodology proposes to model the supply chain in fourth levels organized in a descending order (that is, from the overall system, to decompose the latter into finer granularity subsystems), the three first levels are done with SCOR and the last one is done with the BPMN [34].

4.2 Strategic Level

At this level, SCOR defines the content and context of the supply chain. As shown in Fig. 1 the studied company is the MPF, its customer is the pensioner, and its supplier is the contributing affiliated. The affiliated manages the process Deliver (sD) that describes the management rules for its contribution to the pension plan. The MPF manages three processes, Source (sS) that describes the management of the affiliation and the contribution; Make (sM) that describes the management of the rights liquidation, Deliver (sD) that describes the management of the pensioner manages the Source (sS) process that describes its benefits of a pension. The process Plan (sP) is the one that balance aggregate demand and supply to develop a course of action, which best meets sourcing, production, and delivery requirements.



Fig. 1. The retirement pension supply chain at the first level of SCOR.

4.3 Tactical Level

The second level is a reflection of the strategy adopted by the organization to conduct its operations, in our system there is neither a stocked product nor an engineer-to-order product, our product (the retirement pension) is a make-to-order product, so the processes that we have at this level are: sS2 (Source Make-to-Order Product), sM2 (Make-to-Order), sD2 (Deliver Make-to-Order Product). We note that the processes Plan shown in Fig. 2 are P2 (Plan Source), sP3 (Plan Make), sP4 (Plan Deliver).



Fig. 2. The retirement pension supply chain at the second level of SCOR

4.4 Operational Level

At the third level, companies can specify the activities of sub-processes, best practices, the functionality of the software and existing tools. For our case, the MPF is the studied company, so we detail at this level the process sM2, sS2, sD2.

sS2 Source Make-to-Order product. The following figure (Fig. 3) shows the details of the process sS2 as described at the third level of SCOR; and which corresponds in our case to the affiliation and the contribution management processes. This level of the SCOR model details the process sS2 into five process elements. In our case study, the affiliation and contribution process consists of four elements. We therefore chose among the five elements the four ones that correspond functionally to our case, and which are sS2.1, sS2.2, sS2.3 and sS2.4. sS2.1 Schedule Product Deliveries corresponds to the reception and the allocation of the contributors' files, it receives as input information from the processes Sourcing Plans sP2.4 and Logistics Selection sES6. As output, sS2.1 executes the process sP2.2 Identify, Assess and Aggregate Product Resources. sS2.2 Process Element: Receive Product corresponds to the processing of the contributor's files. As output, sS2.2 requests the processing verification by executing the process sS2.3. sS2.3 Verify Product corresponds to the verification of the processed files, it receives as input the execution request from sS2.2. sS2.3 Transfer Product corresponds to the processed and verified files sending to the liquidation service. As output, sS2.3 executes the processes sED.4: Manage Finished Goods Inventories and sES.4 Manage Product Inventory the processes of establishing inventory information. For Services as this case - this may include tracking the number of service providers (in our case

contributors) and the financial resources committed (in our case contributions) at any given point in time.



Fig. 3. The sS2 process elements

sM2 Make-to-Order. The following figure (Fig. 4) shows the details of the process sM2 as described at the third level of SCOR; and which corresponds in our case to the liquidation process. This level of the SCOR model details the process sM2 into seven process elements. In our case study, the liquidation process consists of five elements. We therefore chose among the seven elements the five ones that correspond functionally to our case, and which are sM2.1, sM2.2, sM2.3, sM2.5 and sM2.6. sM2.1 Schedule Production Activities corresponds to the reception and the allocation of the liquidation files, it receives as input information from the processes sP3.4 Establish Production Plans and sEM5Manage Make



Fig. 4. The sM2 process elements

Equipment and Facilities. As output, sM2.1 executes the process sP3.2 Identify, Assess and Aggregate Production Resources. sM2.2 Issue Sourced/In-Process Product corresponds to the rights constitutions. As input, sM2.2 receives rules and calculation information from sM2.1. As output, sM2.2 gives feedback information to sM2.3. sM2.3 Produce and Test corresponds to the rights liquidation. sM2.5 Stage Finished Product corresponds to the rights concession, at this stage all the liquidated files are edited in a decision. As input, it receives data and files from the execution of sP3.4 Establish Production Plans. sM2.6 Release Finished Product to Deliver corresponds to the liquidated and conceded files sending to the Payment service. As output, it give the payment plan by executing the process sP4.4 Establish Delivery Plans.

sD2 Deliver Make-to-Order product. The following figure (Fig. 5) shows the details of the process sD2 as described at the third level of SCOR; and which corresponds in our case to the payment process. This level of the SCOR model details the process sD2 into fifteen process elements. In our case study, the payment process consists of six elements. We therefore chose among the seven elements the six ones that correspond functionally to our case, and which are sD2.2, sD2.4, sD2.10, sD2.11, sD2.13 and sD2.15. sD2.2 Receive, Configure, Enter and Validate Order corresponds to the reception, control and integration processes, it receives as input the necessary order rules and information (for our case it receives information such as account numbers) by executing the process sED1 Manage Deliver Business Rules. sD2.4 Consolidate Orders corresponds to the decree of the day and the closing of the deadline process this process execution prevents the data updates at the level of the liquidation to impact the payment



Fig. 5. The sD2 process elements

processing. Only data loaded before the deadline end is supported. sD2.10 Pack Product corresponds to the edition of decisions process. sD2.11 Load Product & Generate Shipping Docs corresponds to the validation and sending to accountants process. As input, sD2.11 receives the bank transfer parameters and documentation from the execution of the processes sED6 Manage Transportation and sED8 Manage Import/Export Requirements. As output, it gives the bank transfer history for the sED8 process. sD2.13 Receive and Verify Product by Customer corresponds to the commitments control process. sD2.15 Invoice corresponds to the accounting process.



Fig. 6. The fourth level modeling

The real-time level. At this level, it is the operational study of the processes of the circulation of retirement records within the CMR, so we will trace the itinerary of each file from its reception by the department concerned until its final archiving. This study also concerns the intermediate stages of the processing of each file as well as the interveners in each step, which will allow us to define the number of intervening in the processing of files and to stop all the documents generated following the liquidation of records. Among the two types of pension (civil and military), our study will be based on the treatment of the military scheme's retirement file. The operations performed when processing a retirement file, as shown in Fig. 6 are as follows:

- Once the files sent by the affiliate's administrations are received, the dispatcher registers them for a follow-up, and sends them to the head of the service.
- The head of the service assigns the files to the agents of his team, he also appoints a person who will take care of the follow-up.
- The liquidator starts by making an the file administrative study to check if the file contains all the documents necessary for this processing stage, it also carries out a study of law, thus verifying if the affiliate has the right to the pension. Once the file is ready for liquidation, the liquidator processes the file using the MPF information system. Once the file is liquidated, it generates the statement of the liquidation and sends the file to the auditor.
- The verifier in turn verifies the information on the liquidation before conceding the file, once the concession is made the system generates a pension number to the affiliate, it then generates the decision and draws up the dispatching slip Sending and forwarding the complete files to the head of service for signature.
- Once the files are signed by the head of service they are sent to the division secretariat for the head of division's validation and signature, and then sent to the payment service for validation and processing, copies of the files are after received from the payment service.
- The archiving agent then registers the files follow-up information and classifies them, and then it transmits copies to the affiliate's administrations for information. Finally, he sends the pension file to the archives center.

5 Conclusion

The work presented in this paper is the result of a literature study, an analysis and a modeling of the Moroccan supply chain of retirement pension that provides a comprehensive view of the various material and information flows in this supply chain. This work results are the mapping of the retirement pension supply chain, done according to an explained methodology that combines SCOR and BPMN and which spreads over four levels, the three first ones are done with SCOR model, and the modeling of the main execution processes of the fourth level is done in BPMN 2.0 through Bonita BPM 7. In order to meet its major challenge, and which is ensuring continuity between the salary and the pension, this work will be the basis for a future simulation work that implement and study the behavior of business processes to optimize the retirement pension supply chain.

References

- Sengupta, K., Heiser, D., Cook, L.: Manufacturing and service supply chain performance: a comparative analysis. J. Supply Chain Manage. 42, 4–15 (2006)
- 2. Giannakis, M.: Management of service supply chains with a service-oriented reference model: the case of management consulting. J. Supply Chain Manage. **16**, 346–361 (2011)
- 3. Akkermans, H., Vos, B.: Amplification in service supply chains: an exploratory case study from the telecom industry. Prod. Oper. Manage. **12**, 204–223 (2003)
- 4. Sampson, S., Froehle, C.: Foundations and implications of a proposed unified services theory. Prod. Oper. Manage. **15**, 329–343 (2006)
- 5. Kathawala, Y., Abdou, K.: Supply chain evaluation in the service industry: a framework development compared to manufacturing. Manag. Auditing J. **18**, 140–149 (2003)
- 6. Tuncdan, B., Erhan, A., Melike, D.K., Oznur, Y., Kaplan, Y.C.: A new framework for service supply chains. Serv. Ind. J. 27, 105–124 (2007)
- 7. Li, C., Liu, Y., Cheng, J.: The research on service supply chain. In: IEEE International Conference on Service Operations and Logistics and Informatics, Beijing (2008)
- 8. Wu, H., Yang, S.: Service supply chain: a conceptual framework compared with manufacturing supply chain. In: IEEE International Conference on Management and Service Science, Wuhan (2009)
- He, T., Ho, W., Xu, X.F.: A value-oriented model for managing service supply chains. In: IEEE International Conference on Industrial Engineering and Engineering Management, Macao (2010)
- Song, D., Xu, Y.: Integrated design of service supply chain in the perspective of producer service outsourcing. In: IEEE International Conference on Management and Service Science, Wuhan (2011)
- 11. Bendnar, J., Leitmanova, I.F.: Draft of the regional pension scheme functioning simulated in the Czech Republic. Kontakt **18**, 112–119 (2016)
- 12. Balasuriya, J., Gough, O., Vasileva, K.: Do optimists plan for retirement? A behavioural explanation for non-participation in pension schemes. Econ. Lett. **125**, 396–399 (2014)
- Wu, H., Zeng, Y.: Equilibrium investment strategy for defined-contribution pension schemes with generalized mean-variance criterion and mortality risk. Insur. Mathe. Econ. 64, 396– 408 (2015)
- 14. Manxiu, N., Jinquan, G., Xuhui, Z., Jun, Z.: Does new rural pension scheme decrease elderly labor supply? Evid. From CHARLS. China Econ. Rev. **41**, 315–330 (2016)
- Maloney, M., Carthy, A.M.: Understanding pension communications at the organizational level: insights from bounded rationality theory & implications for HRM. Hum. Resour. Manage. Rev. 27, 338–352 (2017)
- Josiah, J., Gough, O., Haslam, J., Shah, N.: Corporate reporting implication in migrating from defined benefit to defined contribution pension schemes: a focus on the UK. Account. Forum 38, 18–37 (2014)
- Platanakis, E., Sutcliffe, C.: Pension scheme redesign and wealth redistribution between the members and sponsor: The USS rule change in October 2011. Insur. Mathe. Econ. 69, 14– 28 (2016)
- Nielsen, J.A., Sandmann, K., Schlogl, E.: Equity-linked pension schemes with guarantees. Insur. Mathe. Econ. 49, 547–564 (2011)
- Avédissian, A., Valverde, R.: An extension proposition for the agent-based language modeling ontology for the representation of human-driven collaboration in supply chain systems. IFAC-PapersOnLine 48, 1857–1864 (2015)

- 20. Fahimnia, B., Tang, C.S., Davarzani, H., Sarkis, J.: Quantitative models for managing supply chain risks: a review. Eur. J. Oper. Res. **247**, 1–15 (2015)
- Jebbor, S., El Afia, A., Chiheb, R., Ouzayd, F.: Comparative analysis of drug supply and inventory management methods literature review. In: 4th IEEE International Colloquium on Information Science and Technology, Tangier (2016)
- Jebbor, S., El Afia, A., Chiheb, R., Ouzayd, F.: Management and control of stochastic drug supply chain by KANBAN and Petri Net. In: 3rd IEEE International Conference on Logistics Operations Management, Fez (2016)
- Gan, V.J., Cheng, J.C.: Formulation and analysis of dynamic supply chain of backfill in construction waste management using agent-based modeling. Adv. Eng. Inform. 29, 878–888 (2015)
- 24. Heckmann, I., Comes, T., Nickel, S.: A critical review on supply chain risk definition, measure and modeling. Omega **52**, 119–132 (2015)
- Long, Q., Zhang, W.: An integrated framework for agent based inventory-productiontransportation modeling and distributed simulation of supply chains. Inf. Sci. 277, 567–581 (2014)
- Sharma, B., Ingalls, R.G., Jones, C.L., Khanchi, A.: Biomass supply chain design and analysis: basis, overview, modeling, challenges, and future. Renew. Sustain. Energy Rev. 24, 608–627 (2013)
- Ouachi, Z., Allenet, B., Chouchane, N., Calop, J.: Le référencement des médicaments au niveau des établissements hospitaliers français. Le Pharmacien hospitalier 45, 57–65 (2010)
- Mezouar, H., El Afia, A., Chiheb, R.: A new concept of intelligence in the electric power management. In: 2nd IEEE International Conference on Electrical and Information Technologies, Tangier (2016)
- 29. Persson, F.: SCOR template: a simulation based dynamic supply chain analysis tool. Int. J. Prod. Econ. **131**, 288–294 (2010)
- Mezouar, H., El Afia, A., Chiheb, R., Ouzayd, F.: Toward a process model of Moroccan electric supply chain. In: IEEE International Conference on Electrical and Information Technologies, Marrakech (2015)
- 31. Portail caisse marocaine des retraites. https://www.cmr.gov.ma/
- 32. Saadi, A.: Quelle méthode adopter pour modéliser les processus métier de l'administration?. Doctoral thesis. Faculté des sciences et de génie Université LAVAL Québec (2006)
- 33. Féniès, P.: Une méthodologie de modélisation par processus multiples et incrémentiels: application pour l'évaluation des performances de la Supply Chain. Doctoral thesis. Computational Engineering, Finance, and Science. Université Blaise Pascal (2012)
- Mezouar, H., El Afia, A., Chiheb, R., Ouzayd, F.: Proposal of a modeling approach and a set of KPI to the drug supply chain within the hospital. In: 3rd IEEE International Conference on Logistics Operations Management, Fez (2016)

Advances in Web Technologies, Semantics and Future Internet

Creating Multidimensional Views from RDF Sources

Yassine Laadidi^(IM) and Mohamed Bahaj

University of Hassan I, Settat, Morocco yassine.laadidi@gmail.com, mohamedbahaj@gmail.com

Abstract. Business Intelligence (BI) systems have been adopted for decades to collect and analyze (periodically) a mass of relevant information from internal data sources. With the emergence of the Semantic Web (SW) technologies and vocabularies, no one could deny the necessity of including these external web sources in the decision-making process. However, the actual architecture of BI remains operational only in a well-controlled context where the sources are static and where the multidimensional scheme is defined in advance. Therefore, there is a strong need for new methods in order to extract information from dynamic data sources and enabling On-Line Analytical Processing (OLAP). In this paper, we propose a transposition method of multidimensional concepts over multiple ontologies sources in order to create the correspondent schema.

Keywords: OLAP · Data integration · RDF · Semantic Web

1 Introduction

Business Intelligence (BI) is known as the set of methods, practices and tools used by decision-makers to collect transform and summarize data from various data sources in order to get in-depth knowledge of the company and to define and sustain their business strategies.

The Data Warehouse (DW) is the physical incarnation of the multidimensional model that provides generalized and consolidated data in multidimensional view. As a database, the DW is designed to organize and store subject-oriented, integrated, time variant and non-volatile data [17]. On-line analytical processing (OLAP) tools are used to answer multi-dimensional requests over the DW and to create subject-oriented databases called OLAP cubes (or OLAP views). In fact, a DW represent a huge OLAP cube in which users can filter and access to a large amount of multidimensional data and therefore get the big view of company's activities. In general, the process of turning data into pertinent knowledge goes through three main stages:

- Data acquisition: consists of identifying and collecting potential data sources. These data are found in different format. They can be "flat" files (e.g., XML files) or/and database systems. These data sources are therefore heterogeneous so we will have to go through a so-called integration phase to be able to manipulate them.
- Data integration: consists of concentrating the collected data in a unified repository: the data warehouse (DW). It enables decision-making applications to have a

common, homogeneous and reliable source of information. This phase is based-on an Extract-Transform-Load (ETL) process.

• Data analysis: consists of consolidate, view, and analyze data according to multiple perspectives.

The data integration process is fully applied under a well-controlled context where data sources are well-known and where data is extracted and prepared in advance to be periodically loaded into a designed data warehouse. The data warehouse provides multidimensional data storage for OLAP and preserves historical data (see Fig. 1). There are two notable implementation of OLAP: Relational OLAP (ROLAP) and Multidimensional OLAP (MOLAP). In ROLAP, data is stored using a relational database management system (RDBMS) which allows a great storage capacity of data, the SQL language id used to build aggregation logics and views, which are in general too complex and very demanding in terms of resources and execution time. The MOLAP server use a dedicated multidimensional storage engine, compared to ROLAP, MOLAP cannot handle large amount of data but only summary-level information, yet, MOLAP allows very fast information retrieval and pre-generated calculations. Sometimes a hybrid OLAP (HOLAP) solution can be managed to merge the storage capacity of ROLAP with the fast-processing capacity of MOLAP.

The semantic web (SW) was created as an extension of actual Web through standards by the W3C in order to make machines able to 'understand' the semantics and the meaning of information and thereby enabling automated agents to access more intelligently to different sources of data. With the emergence of SW standards such as the Resource Description Framework (RDF) and ontologies, no one could deny the necessity of including semantic web data sources in the BI process in order to provide the necessary knowledge for companies to improve their services and increase their profits. However, the transition to an open BI system is hampered by many constraints: the heterogeneity of the data sources distributed on the web, the management of similarity conflicts during the fusion of ontologies, the transposition of multidimensional concepts to the RDF model to generate OLAP cubes and the problematic of detecting and applying unpredictable scheme modifications.



Fig. 1. The traditional BI architecture: data is extracted from source through Extract-Transform-Load processes (ETL) into a local data warehouse for On-Line Analytical Processing (OLAP).

Therefore, there is a need for a flexible system that, in hand, provides a homogeneous system for users and preserves historical traceability of information and, in the other hand, allowing an easy access of multidimensional views on the RDF graph. The method that we present in this paper is based on a novel BI architecture for open-world scenarios and aims to make transposition of multidimensional concepts on the RDF model in order to create a multidimensional ontology. The MD-ontology is created according to the requested OLAP view and through integration of several sources of external ontologies.

This paper is organized as follows: Sect. 2 presents a brief overview of the most significant related works. Section 3, introduces the architecture adopted and the proposed method. Finally Sect. 4 concludes the paper and presents the future steps.

2 Overview of Related Works

In this section we overview some significant concepts and related works regarding multidimensional (MD) modeling and SW formalisms.

The main purpose of the MD model is to provide much better query performance, especially against very large business data and has the great benefit of being easier to understand and simple [1] regarding the traditional entity-relationship model. One of the simplest forms of the MD model is the star scheme.

A MD model is composed of facts (or subject of analysis) processed according to different dimensions (i.e. perspectives). Numerous indicators (i.e. measures) are available for each fact instance in order to gain insight. Dimensions could be represented at different levels of aggregation. To define these levels, each dimension is provided with one or more hierarchies. As main part of OLAP, the MD model is designed to solve complex queries in real time.

In general, the definition of the MD data model requires, on one hand, a preliminary study/analyze of the business domain (defining key-business indicators) and, on the other hand, the identification of potential sources of data. Therefore, the MD model is oriented and restricted by both user's analytical needs (what he wants to view) and available data sources (what he has as possible views).

The emergence of semantic Web technologies (in particular ontologies) and capabilities has push companies to think of enriching the decision-making process from web sources while keeping the same easy and fast access to the right information. Hence, knowing the data semantics will certainly improve the quality of information retrieval.

RDF Schema (RDFS) provides a data modelling vocabulary for RDF data and allow describing taxonomies of classes and properties, for example, rdfs:subClassOf, rdfs: range and rdfs:domain. The Web Ontology Language (OWL) is a semantic web language designed to represent rich and complex knowledge about things, groups of things, and relations between things and gives a much richer vocabulary including RDFS. There are three species of OWL: OWL Lite, has limited expressiveness capabilities for classification hierarchy and has a limited notion of cardinality (1 or 0). OWL-DL is an extension of OWL adopting Description Logic (DL) to supports maximum expressiveness while retaining computational completeness and decidability [3]. OWL Full,

which is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. Description logic has already been applied in multidimensional modeling [4]. OWL-DL comes with automated reasoners (such as FaCT++, HermiT, RacerPro, etc.) for satisfiability checking over ontologies [5].

First methods aimed to design relational databases from ontologies using a mapping of OWL concepts to relational schemas (e.g. [6–9]) in order to get benefits of better mechanisms for storing, querying and manipulating data. In addition of the lack of MD design, the frequent changes applied to the ontologies make the adoption of these methods in an OLAP system very complex to maintain and does not allow a real-time access to the information.

The lack of MD design has been overcome in [4]; authors propose to transform a MD scheme to MD ontology using the RDF Data Cube vocabulary (QB) [13] to establish a relationship between the RDF definition of the data cube structure and the generated ontology via common concepts which enable OLAP operations and measures summarization. The QB4OLAP engine proposed in [14] extends QB and aims to transform MD data stored in relational DW into RDF triples (stored in a triple store) including dimension levels, measures, hierarchies within dimensions and the parent-child relationships among levels, finally, enabling OLAP operations via SPARQL queries. The main purpose of this approach is to publish and share statistical information from internal data warehouses.

There are three different approaches regarding ontology-based integration [10]: the single-ontology approach, which all data sources are related to one global ontology in order to provide the same view on a business domain; the multiple-ontologies approach, which each data source is described by its own ontology and managed independently; the hybrid approach, which each local data source is described by its own ontology using the common vocabulary of the global ontology. The hybrid approach is adopted mostly for reasoning over OWL-DL ontology sources (see [11, 12]). In fact, this approach requires re-definition of local ontologies according to the common application vocabulary and ontological mapping, which also cannot be automated since high expressiveness results the need of heavy computations [2].

In [15] authors present their method of transforming ontology structure into a star schema where the user begin by choosing an object property as the subject of analysis (fact), automatically creating the corresponding dependency graph [16] and finally rearranging the corresponding MD schema; the method support only OWL Lite vocabulary. In this paper, we introduced a discovery method of multidimensional concepts in order to design the correspondent star schema from OWL ontologies sources, the method is characterized by a simplification process of different ontologies sources in order to decrease the size of the targeted graph, keeping only basic expressiveness of dimensional concepts and hence providing less computations.

3 Architecture and Methodology

In this section we present an overview of our adopted OLAP system followed by the transposition method along with the definitions that sustain it.

Generally, there are two approaches followed in order to exploit semantic web sources: first approach consists of identifying and integrating data from data sources into a designed DW. In this case, the DW is considered as a global OLAP view that store all aggregated measures, dimensions and hierarchies, and which, users can filter and access to further specific details (i.e. granularities) according to their business analysis needs; the second approach would be more dynamic and consist of performing OLAP operations directly over the RDF graph with nearly no transformations. However, for the first static approach, the problem of unpredictable scheme changes of data sources disables the automation of the whole process and therefore could not be applied under an open-world scenario. The second approach is, in contrary, more dynamic but also more complex and provides a very lower level of data materialization (no historical data) and allows only lightweight transformations.

We assume that the second approach is the more appropriate for huge RDF graph where data and data structures are altered recurrently. This assumption came from the fact that semantic web technologies came with huge abilities to improve data discovery. However, the heterogeneity of RDF data sources makes too difficult to set a mediation system to define mapping between the local vocabulary used by users and source schemas due to the complexity of rewriting views and/or modifying the mapping for each data source alteration. As we focus in first place to enclose only a specific area of OWL concepts and relationships, the user must be able to build his requests and select needed elements directly from sources. Therefore, we choose to setting up an exploratory functionality built on top of a recommendation system (RS) for the user to explore and select requested OWL classes from all available sources based-on classes' names or/and URIs. The user can enrich the local vocabulary by defining new relationships (such as concepts similarities). Figure 2 illustrate the mechanism adopted and general phases involved during the building process:

- 1. Helped by the RS, user choose his dimensional concepts from all available data sources.
- 2. Based-on the selected OWL classes, a t-schema is created in order to enclose a precise area of data segments for time optimization and less computation, hence better performance.
- 3. For each measure and related nodes identified in the t-schema, a 'one-measure' cube is extracted.
- 4. Cube instances are stored and re-organized according to their dimensions.
- 5. DW is rebuilt with new data segments.

Selected concepts (OWL classes and properties) are needed as inputs to determine and construct a temporary OWL schema (t-schema) of related classes, object properties, data properties and literals. The purpose of a temporary schema is to surround a precise area of data segments and optimize as much as possible time processing by focusing only on pertinent sources of information. The construction the t-schema consists of three main steps:

• In the first step, the process is initialized by returning a set of triples for each OWL class from the selected list in which the class is a domain. All returned triples are added to the t-schema.



Fig. 2. The global architecture adopted in this paper. A cube schema is designed from RDF/OWL sources according to users requested view.

- The second step consist of linking related nodes with datatype objects, in other word, collecting all triples which the subject is a t-schema class and the predicate is a datatype property, which will enables to preserve literal values as potential measures.
- The third step, the whole schema is rebuild by adding all relationships existed between collected objects from ontology sources and which were not added during previous phases in order to complete the current schema by needed concepts and providing a powerful mechanism for enhanced reasoning about properties.

The resulted schema can be adjust by the user to add more nodes and relationships and/or remove others. At this point, the ontology structure (t-schema) is more specific and provide a temporary unified vocabulary of potential measures, related dimensions and eventual hierarchies (see Fig. 3).

In our designing method, a star schema is defined for each individual measure, in other words, each fact is composed of one measure and multiple related dimensions:

Definition 1. Let F be a single-measure fact, D_i a related dimension and S the star schema. We have:

$$\mathbf{F}^{\mathbf{m}} = \left(\mathbf{m}, \mathbf{I}^{\mathbf{F}}\right) \tag{1}$$

$$\mathbf{S} = (\{(F^m, D_i)...\}) \tag{2}$$

Algorithm 1 Creating the temporary schema (t-schema) from available OWL ontology sources using SPARQL batches



Fig. 3. Pseudo code for building the temporary RDF/OWL schema from RDF/OWL sources.

Where:

- m is a unique measure,
- $I^{F} = \{I_{1}^{F} \dots I_{i}^{F}\}$ is a set of fact instances.
- $\{(F, Dj)...\}$ associate the fact to all dimensions.

We consider a set of rules in order to outline potential measures, related dimensions and eventual hierarchies from the t-schema. The fist element to identify is the OWL class representing the measure, (e.g., order quantity, order amount, price, etc.) and detecting related dimensions and corresponding hierarchies:

Rule 1. A concept (i.e., OWL class) is a potential measure if it has at least one numerical value as literal range (even if it is a string type).

Rule 2. Every node in the RDF t-schema which is related directly to the measure concept is a potential dimension.

Rule 3. If the object property of a dimension class has cardinality greater than 1 then the range class is transformed into a hierarchy, otherwise, if cardinality is equal to 1 then the range class will be transformed into a dimension attribute.

Rule 4. subClass properties are used to categorize classes in the OWL language, and so, if the property describe a dimension class, the range object is considered as a hierarchy class which is related to the main dimension.

Rule 5. A class may have subclasses and these later also have their subclasses and so on, therefore to resolve this problem, only subclasses of the first level of class hierarchy are included in the t-schema structure to represent details about the super class.

Finally, cube instances (instances of the created schema) are extracted and re-organized in form of collections according to shared dimensions, for example, quantity and price are two measures which may have the same dimensions such as Product, Customer and/or Time, therefore, is too easy to re-build and reorganizes a global view according to those axis with a fact gathering both measures.

The user during this phase has possibility to reorganize cube-instances by ordering them and choosing desiring dimensions in purpose to build different views by changing measures and corresponding dimensions, filtering analysis result (slice and dice actions) and performing OLAP methods on pre-defined cube objects. Consequently, data materialization is managed by accumulating extracted cubes. Unfortunately, this last phase is not included in this process.

4 Conclusion

Exploiting SW capabilities for OLAP analysis still difficult due to the nature of RDF data sources. Therefore, In order to be able to explore, identify, extract and transform data immediately from heterogonous sources and/or local data warehouses and perform analytical tasks based-on users' specifications, a more supple BI architecture shall be applied.

We believe that SW data sources should be treated according to on-demand requests, and the global OLAP cube should be constructed from small cube-instances which will provide an accepted needed and available level of details. we introduced a discovery method of multidimensional concepts in order to design the correspondent star schema from OWL ontologies sources, the method is characterized by a simplification process of different ontologies sources in order to decrease the size of the targeted graph, keeping only basic expressiveness of dimensional concepts and hence providing less computations.

We also presented a different point of dealing with RDF data for building OLAP cubes from SW data sources. In our proposed approach, we consider single-measure cubes as the basics for building the 'general' OLAP cube, the main raison for that is because of the unpredictable changes of the RDF graph structure. However, we're still looking for a suitable solution for fast and coherent materialization process of all collected cubes in order to be able to analyze historical information and to perform OLAP aggregations. This issue is the main subject of future researches.

References

- 1. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley (2011)
- Abello, A., et al.: Using semantic web technologies for exploratory OLAP: a survey. IEEE Trans. Knowl. Data Eng. 27(2), 571–588 (2015)
- OBITKO: Web Ontology Language OWL. https://www.obitko.com/tutorials/ontologiessemantic-web/web-ontology-language-owl.html. Accessed 01 Aug 2017
- Prat, N., Megdiche, I., Akoka, J.: Multidimensional models meet the semantic web: defining and reasoning on OWL-DL ontologies for OLAP. In: Proceedings of the Fifteenth International Workshop on Data Warehousing and OLAP. ACM (2012)
- Neumayr, B., Schütz, C., Schrefl, M.: Semantic enrichment of OLAP cubes: multi-dimensional ontologies and their representation in SQL and OWL. In: OTM Confederated International Conferences "On the Move to Meaningful Internet Systems". Springer, Heidelberg (2013)
- Astrova, I., Korda, N., Kalja, A.: Storing OWL ontologies in SQL relational databases. Int. J. Electr. Comput. Syst. Eng. 1(4), 242–247 (2007)
- Vysniauskas, E., Nemuraite, L.: Mapping of OWL ontology concepts to RDB schemas. In: Information Technologies, pp. 317–327 (2009)
- 8. Liu, X.: Data warehousing technologies for large-scale and right-time data. Aalborg University. Defensed on June 2012
- Ho, L.T.T., Tran, C.P.T., Hoang, Q.: An approach of transforming ontologies into relational databases. In: Asian Conference on Intelligent Information and Database Systems. Springer, Cham (2015)
- 10. Wache, H., et al.: Ontology-based integration of information a survey of existing approaches. In: IJCAI-01 Workshop: Ontologies and Information Sharing, vol. 2001 (2001)
- Nebot, V., Berlanga, R.: Building data warehouses with semantic web data. Decis. Support Syst. 52(4), 853–868 (2012)
- Selma, K., et al.: Ontology-based structured web data warehouses for sustainable interoperability: requirement modeling, design methodology and tool. Comput. Ind. 63(8), 799–812 (2012)

- 13. The RDF Data Cube Vocabulary. W3C Recommendation, 16 January 2014. https://www. w3.org/TR/vocab-data-cube/. Accessed 1 Aug 2017
- 14. Bouza, M., et al.: Publishing and querying government multidimensional data using QB4OLAP. In: 2014 9th Latin American Web Congress (LA-WEB). IEEE (2014)
- Gulić, M.: Transformation of OWL ontology sources into data warehouse. In: 2013 36th International Convention on Information & Communication Technology Electronics & Microelectronics (MIPRO). IEEE (2013)
- Khouri, S., Ladjel, B.: A methodology and tool for conceptual designing a data warehouse from ontology-based sources. In: Proceedings of the ACM 13th International Workshop on Data Warehousing and OLAP. ACM (2010)
- 17. Inmon, W.H.: What is a data warehouse? Prism Tech Topic 1(1) (1995)

An Ontology Based Approach to Organize Supplier and Transportation Provider Selection Negotiation in Multi-agent System Model

Iman Achatbi^(IM), Khalid Amechnoue, and Saloua Aoulad Allouch

National School of Applied Sciences Tangier, Abdelmalek Essaadi University, Tangier, Morocco i.achatbi@gmail.com, kamechnoue@gmail.com, saloua.aoulad@gmail.com

Abstract. Since the advent of globalization and the evolution of organizations, the need for new and efficient processes for supply chain has become urgently important. One of supply chain management problems is the supplier selection which attracts the attention of many researches. Various efforts are made in this context, mainly the development of agent-based systems to automate the process of selecting suppliers. Negotiation is a critical approach to solve conflicting transaction between nodes and scheduling problems among supply chain members. In this paper, we propose a negotiation model based on agents to settle the problem of selecting supplier and transportation provider, and then alleviating the human interactions. The negotiation knowledge utilized by agent is organized by ontology in this paper, agents communicate via message exchange in the form of common ontology for agents participating in the negotiations.

Keywords: Agent \cdot Negotiation \cdot Ontology \cdot Supplier selection \cdot Transportation provider \cdot Model

1 Introduction

With the globalization and evolution of organizations, in a world increasingly competitive, to more responsiveness, agility and flexibility, new methods of management have seen the day. This brought about new challenges for the integration of legally separated firms and the coordination of materials, information and financial flows.

Raw materials and supplied components constitute 60% to 70% of finished goods cost [1]. Delivery times of suppliers and distribution reliability affecting more than production time on the stock level and the quality of service of each manufacturer [2]. Thus, ensuring timely and sufficient commodity saturation at a warehouse is important, in order to avoid surplus of goods in stock (overstock) and deficiency of goods in warehouses (out-of-stock).

Supplier selection becomes a considerable problem of supply chain management (SCM). Choosing the right suppliers involves much more than scanning a series of price list, and choices will depend on a wide range of factors which involve both quantitative and qualitative criteria [3], including price, quality, delivery time, service and so on. For

supply chain members with conflicting interests or viewpoints, negotiation is an essential approach for decision making and reaching the mutual agreement.

More and more, multi agent systems are seen as a new technology for improving or replacing technologies used in both transactional and analytical information technologies. The ability to negotiate is the unique feature of agents that distinguishes them from other software. In fact, agents are best suited for applications that are modular, decentralized, changeable and complex. Within an agent-based system for supplier selection negotiation, software agents are established to represent various parties and functions involving in the buyer–seller interaction process.

Ontologies not only provide a definition of the terms that can be used in communication; ontologies also provide the definition of the world in which an agent grounds its actions. Different agents of a system can reach a shared understanding by committing to the same ontology. Two important functions of ontologies are that they:

- (1) Enable agents to work cooperatively to communicate with each other,
- (2) Make the available information more accessible to automated agents.

The goal of this paper is to propose an agent based negotiation model to support supplier selection and transportation provider at the same time. Firstly, a multi-agent system (MAS) will be established to realize the proposed negotiation model. Then, the negotiation proposal, negotiation protocol, will be elaborated for the supplier selection environment.

The rest of the paper is organized as follows. Section 2 reviews literature on supplier selection, and agent-based negotiation models. The description of the problem is proposed in Sect. 3. Section 4 discusses the concept of agent-based negotiation. Section 5 provides the agent-based negotiation model for supplier selection and transportation provider selection, and shows the negotiation protocol between agents. Section 6 specifies the multi agent system construction. Finally, conclusion and future work follow in Sect. 7.

2 Literature Review

Multi-agent technology has been used in many areas but industry applications have taken the earliest advancement of agent technology when compared to others. The two major focusing points on automated negotiation are the negotiation protocols and agent decision-making models. As an example of decision-making models: model that automates the supplier selection process.

The authors of [4] proposed an agent-mediated coordination approach to automate the supply chain formation in dynamic and uncertain environments.

In [5] a MAS platform is established for individual companies to form an ecological virtual enterprise based on ontology theory and intelligent agents. An automated negotiation model for e-commerce decision making is proposed by [6]. The authors of [7] proposed an ontology-based approach to organize the multi agent-assisted supply chain negotiations.

The authors of [8] proposed a hybrid multi-agent negotiation protocol to regulate the mobile agent-assisted negotiation, and embedded an ontology operation protocol in it to govern negotiation knowledge.

The authors of [9, 10] tried to adopt agent-based negotiation model to solve multi product supplier selection problem but the negotiation protocol special for multi-product environment has not been focused on.

The authors of [11] developed a multi agent system to resolve the Problem of coordinating Orders in the manufacturer company, three ontologies allow software agents to share available knowledge and identify new knowledge.

Researchers have proposed different negotiation protocols to govern the interaction of agents involving in supplier selection process. Bilateral and multilateral agent negotiation protocols have been proposed in e-procurement and supply chain order fulfilment negotiations, based on the CNP (Contract Net Protocol) regulation [12, 13]. The buyer and seller bargain iteratively on multiple negotiation issues using the buyer-seller bilateral negotiation protocol [14]. Adopting this protocol, agents' decision making methods are studied using the genetic algorithms [15], the heuristic negotiation concession functions [16], the incomplete information inference [17], or the constraints-based fuzzy rules [18].

Ontology is able to integrate descriptive knowledge, procedural knowledge and reasoning knowledge [19]. In recent researches, ontology has been adopted to represent the negotiation protocol and make agent adapt to various negotiation mechanisms. Authors of [20] presented an ontological approach to automated negotiation, particularly suited to open environments. In their work, the negotiation protocol is defined in terms of shared negotiation ontology instead of being hard-coded within agents. In the work of [21], the authors extended the ontology presented by [20] to facilitate multi-item, multi-unit combinatorial reverse auctions.

3 Description of the Problem

Various optimization models are adopted for supply chain to organize process and maximize business objective of an organization, typically transportation, order selection, production and supplier selection. The use of intelligent agents to automate the process emerges as an important tool to facilitate this activity.

The process of selecting suppliers may have a number of disadvantages, due to the manual approach to procure raw material, related to the human factor.

Otherwise, supplier selection is very dependent on the experience of the manager, how analyses pre-determined price lists and takes the decision of which suppliers to choose. Thus, the desire to avoid mistakes in supplier selection process determines the need to automate the process.

Several studies were conducted to automate supplier selection process through the agent-based negotiation model, each model take into account a wide range of product details, including price, quality, delivery time, service and so on. However, there is absence of models how treat transport availability as selection criteria.

Actually, the process of selecting suppliers is strongly related to the process of selecting transportation providers, seen than suppliers may haven't available tracks to ensure transportation of merchandise. Therefore, the optimization model of supplier selection must integrate the process of selecting transportation providers. In this context, we propose a multi agent decision making model to settle the problem of selecting suppliers and transportation providers at the same time as depicted in the Fig. 1.



Fig. 1. Overview of supplier and transportation provider selection and negotiation process

4 Agent-Based Negotiation

Agents need to communicate and interact with other agents. This requires them to share a common understanding of terms used in communication. The specification of the used terms and exact meaning of those terms is commonly referred to as the agent's ontology [22]. Negotiation protocol which regulates the set of rules that govern the interaction of agents is an essential component that constitutes the agent-based negotiation model.

Basically, to build a practical agent negotiation model, three areas have to be considered [23]:

- Objects of negotiation, which include the range of points on which an agreement must be made;
- Negotiation protocols, including the rules that guide the interactions between the various agents;
- Decision-making models of the agents, which guides the agents towards obtaining the solution.

5 Agent-Based Negotiation Model for Supplier and Transportation Provider Selection

This section presents our solution for supplier selection and transportation provider selection. Multi agent system hands the decentralization in solving the problem, which is important in our case because agents are distributed in different node of supply chain.

5.1 Multi-agent System Architecture

The proposed solution will be able to model complex system for supply chain management to address the issues of communication among the parties involved in supply chain, it consists of: *Coordinator Agent, procurement agent, Seller Agent, TransportProduct Agent,* in purchasing company; *Supplier Agent, MaterialSeller Agent, TransportSupplier Agent* in supplier node; and finally in the transport service node, the *transport agent* is present (Table 1).

Proposed agents for the negotiation				
Label	Agent	Activities		
CA	Coordinator agent	 -Controls the interaction between other agents; -Creates PA for each SpA; -Creates TPA for each TA; -Takes the final decision about supplier selection and transportation provider selection. 		
SA	Seller agent	-Determine the set of required products.		
PA	Procurement agent	Receives command from CA about the amount of product must be bought,Agrees with SpA about price and delivery time.		
TPA	Transport product agent	-Receives command from CA to search transportation providers from supplier to purchasing company, -Agrees with the TA about price and delivery time.		
ТА	Transport agent	-Represents transportations providers in the bilateral bargaining process with the corresponding TPA.		
SPA	Supplier agent	-Represent supplier and conduct the bilateral bargaining with the corresponding PA.		
MsA	Materiel supplier agent	-Have knowledge about the stock of material		
TsA	Transport supplier agent	-have knowledge about transport planning.		

Table 1.	Proposed	agents for	the negotiation
----------	----------	------------	-----------------

To ensure process production of finished goods, timely coordination of the elements of orders with suppliers is required, namely lists of materials needed, their quantity, price, delivery time and quality.

Seller agent cheeks the database of finished goods and database of raw materials, it sends an order to the coordinator Agent to buy raw material if stock is insufficient. *Coordinator Agent* commands *Procurement* Agent to buy raw materials, the last one agrees with *Supplier* agent about price and delivery time (Fig. 2).



Request and response with database
 Order between two agents
 Interaction of agents with negotiation
 Information flow

Fig. 2. Multi agent system architecture

Once the supplier is selected and if transportation of raw material isn't assured by the supplier (transport not available), *TransportProduct* agent prompts the negotiation with *Transport* Agent to insure transportation of raw materials from supplier to the warehouse for raw material in production node. The outputs of *TransportProduct Agent* are supplier node, delivery node, product type, quantity, and delivery time.

5.2 Negotiation Protocol

The negotiation protocol governing the interaction of agents involving in the negotiation model for supplier selection and transportation provider selection is a hybrid protocol of combinatorial procurement auction and multi-bilateral bargaining [9].

Regarding the one-buyer-many-seller negotiation model, the buyer is represented by multiple PA instances initialized by the CA to negotiate with respective sellers represented by SPAs for suppliers' selection and TAs for transportation providers' selection. Through iterative negotiation concessions, each pair of BA instance and SPA will generate a negotiation result [8]. The CA then collects all the negotiation results and selects the preferable suppliers of raw material and transportation providers (Fig. 3).



Fig. 3. Negotiation protocol for selecting suppliers & transportation providers

Firstly, the seller agent checks databases of raw materials and finished products in order to determine the type and amount of raw materials to buy.

Then, the coordinator agent receives the order from seller agent to buy raw material, it create instances of the procurement agent PAs for all suppliers, and transport product agent TPAs for all transportation providers.

Regarding the Procurement agent, it conducts the bilateral bargaining with the corresponding SPA. The negotiation is about price, quantity, quality, delivery time, and transport availability.

Meanwhile, the transport product agent conducts the bilateral bargaining with the corresponding TA, The negotiation is about price, transport quality, and delay.

Finally, the CA select cooperative suppliers for products based on the negotiation results between the PA and the SPAs and TPAs and TAs.

The message exchange for the negotiation takes the form of alternate and iterative exchanges of "call for proposal (cfp)" and "propose" messages [24].

6 Multi Agent System Construction

As a promising technology JADE [25] will be used to automate procurement management process and Eclipse Luna as integrated development environment to construct the multi agent system. Despite of other multi-agent technologies JADE is available as free software component hence development cost is marginal.

Ability to perform under limited resource environment, installation and access through the mobile devises increased the rapid growth of multi-agent technology. Instead of the standalone environment the agent is accessible through the web interface with minimum bandwidth. Protégé and Ontology Bean Generator to create domain ontology and to transform it into JADE classes, MySQL to support the database, ACL (agent communication message) messages to transfer information, share knowledge and negotiate with each other using FIPA negotiation protocols.

7 Conclusion

In this paper, an agent-based negotiation model for supplier selection and transportation provider selection is proposed. In the proposed MAS, the interaction involved in the buyer-seller negotiation is represented as multilateral buyer-seller negotiation.

Regarding the negotiation knowledge representation, three ontologies are established to structure concepts and agent action during the exchange of messages between agents.

Our purpose in the future work is to implement the proposed system in the JADE platform, more effort will be made on the agent decision making methods to improve the robustness of the negotiation protocol.

References

- 1. Ouzizi, L.: Planification de la production par co-décision et négociation de l'entreprise virtuelle, Metz (2005)
- 2. Harmon, R.L., Peterson, L.D.: Reinventing the factory II: managing the world class factory, vol. 2. Simon and Schuster (1992)
- Ho, W., Xu, X., Dey, P.K.: Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. Eur. J. Oper. Res. 202(1), 16–24 (2010)
- Wang, M., Wang, H., Vogel, D., Kumar, K., Chiu, D.K.W.: Agent-based negotiation and decision making for dynamic supply chain formation. Eng. Appl. Artif. Intell. 22(7), 1046– 1055 (2009)
- Wang, X., Wong, T.N., Wang, G.: An ontological intelligent agent platform to establish an ecological virtual enterprise. Expert Syst. Appl. 39(8), 7050–7061 (2012)
- Cao, M., Luo, X., Luo, X., Dai, X.: Automated negotiation for e-commerce decision making: a goal deliberated agent architecture for multi-strategy selection. Decis. Support Syst. 73, 1– 14 (2015)
- Wang, G., Wong, T.N., Wang, X.: An ontology based approach to organize multi-agent assisted supply chain negotiations. Comput. Ind. Eng. 65(1), 2–15 (2013)
- Wang, G., Wong, T.N., Wang, X.: A hybrid multi-agent negotiation protocol supporting agent mobility in virtual enterprises. Inf. Sci. 282, 1–14 (2014)
- Yu, C., Wong, T.N.: An agent-based negotiation model for supplier selection of multiple products with synergy effect. Expert Syst. Appl. 42(1), 223–237 (2015)
- 10. Yu, C., Wong, T.N.: A multi-agent architecture for multi-product supplier selection in consideration of the synergy between products. Int. J. Prod. Res. **53**(20), 6059–6082 (2015)
- 11. Plinere, D.S., Borisov, A.N., Aleksejeva, L.Y.: Interaction of software agents in the problem of coordinating orders. Autom. Control Comput. Sci. **49**(5), 268–276 (2015)
- 12. Lin, F.R., Lin, Y.Y.: Integrating multi-agent negotiation to resolve constraints in fulfilling supply chain orders. Electron. Commer. Res. Appl. 5, 313–322 (2006)
- Renna, P., Argoneto, P.: Production planning and automated negotiation for SMEs: an agent based e-procurement application. Int. J. Prod. Econ. 127, 73–84 (2010)
- Talluri, S.: A buyer-seller game model for selection and negotiation of purchasing bids. Eur. J. Oper. Res. 143, 171–180 (2002)
- Choi, S.P.M., Liu, J.M., Chan, S.P.: A genetic agent-based negotiation system. Comput. Netw. Int. J. Comput. Telecommun. Netw. 37, 195–204 (2001)
- Narayanan, V., Jennings, N.R.: An adaptive bilateral negotiation model for e-commerce settings. In: 7th International IEEE Conference on E-Commerce Technology, Munich, Germany, pp. 34–39 (2005)
- Jonker, C.M., Robu, V., Treur, J.: An agent architecture for multi-attribute negotiation using incomplete preference information. Auton. Agent. MultiAgent Syst. 15, 221–252 (2007)
- Chen, Y.M., Huang, P.N.: Agent-based bilateral multi-issue negotiation scheme for e-market transactions. Appl. Soft Comput. 9, 1057–1067 (2009)
- Chau, K.W.: An ontology-based knowledge management system for flow and water quality modeling. Adv. Eng. Softw. 38, 172–181 (2007)
- Tamma, V., Phelps, S., Dickinson, I., Wooldridge, M.: Ontologies for supporting negotiation in e-commerce. Eng. Appl. Artif. Intell. 18, 223–236 (2005)
- Giovannucci, A., Rodriguez-Aguilar, J.A., Reyes, A., Noria, F.X., Cerquides, J.: Enacting agent-based services for automated procurement. Eng. Appl. Artif. Intell. 21, 183–199 (2008)
- Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. 5(2), 199–220 (1993)
- Faratyn, P., Sierra, C., Jennings, N.R.: Negotiation decision functions for autonomous agents. Multi-Agent Racionality 24, 159–182 (1996)
- 24. FIPA: FIPA Communicative Act Library Specification (2002). http://www.fipa.org/specs/ fipa00037/SC00037J.html
- 25. Um, W., Lu, H., Hall, T.J.: A study of Multi-agent based supply chain modeling and management. iBusiness 2(4), 333 (2010)

Deep Neural Networks Features for Arabic Handwriting Recognition

Mustapha Amrouch^(☉) and Mouhcine Rabi

Laboratory IRF-SIC Faculty of Sciences, Ibn Zohr University, Agadir, Morocco m.amrouch@uiz.ac.ma, mouhcineh@gmail.com

Abstract. This work aims to compare the learning features with Convolutional Neural Networks (CNN) and the handcrafted features. In order to determine which the best between these two type of features. We consider our previous baseline HMM system [1] for Arabic handwritten word recognition. Experiments have been conducted on the well-known IFN/ENIT database. Achieved results using CNN features are better than those obtained by the hand-crafted features. This demonstrates the high efficiency of CNN results from the strong capability for hierarchical feature learning given a large amount of data. However, Hand-engineered features are not generated from an optimization process to be compatible with the specific problem, and insufficient to be encoded with supervision.

Keywords: Handwriting recognition \cdot Extraction features \cdot Convolutional neural networks \cdot Hidden markov models

1 Introduction

Since the work of Krizhevsky et al. [2] deep neural networks becomes the most prominent trend in virtually all fields of automatic learning statistical applications [3–7]. With their introduction in this area, major improvements have been observed. As a result, powerful intelligent systems have been developed. In addition, the most widespread neural networks known as Convolutional Neural networks (CNN), developed initially by Lecun et al. [8], for the task of reading handwritten digits.

Currently, CNN and deep CNN reputed by their noteworthy successes in solving challenging tasks with unmatched performances outperforming other approaches. Such as object classification [9, 10], automatic speech recognition [11, 12], word spotting and document analysis [13, 14], challenge ImageNet [2, 3, 9, 10], natural language processing [15, 16], image segmentation [17, 18] or road sign classification [19].

Due to their remarkable performance. CNN have found their way into handwriting text recognition [20, 21, 22] as well. Indeed, a significant usage of their should also be noted, especially for both Latin characters recognition [23] and handwritten digits [7, 24, 25, 26], where the approaches based mainly on CNN progressively dominate the competitions and beating benchmark performances by wide gaps.

For example, the system based on CNN (convNet) proposed by Ciresan et al. [27], for recognizing of handwritten Chinese characters has won the first place in ICDAR-2011 competition. The same system demonstrated extraordinary ability and

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems,

Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_14

reaches of near-human performance on MNIST database [25]. Furthermore, another system developed by author Graham based on an alternative of CNN called DeepCNet [28], which won first place in the ICDAR 2013 Chinese Handwriting Recognition Competition [29].

In addition, CNNs are also useful for other computer vision tasks as generic feature extractors. This high efficiency of CNN results come from the strong capability for hierarchical feature learning given a large amount of data. It seems that CNN is replacing hand-crafted features for a wide variety of problems. Because the hand-crafted features are often generated without an optimization process compatible to the specific problem. However, CNN is typically used as a static model whose input feature is fixed-dimensional.

In this work, Our goal aims to compare the efficiency of handcrafted vs CNN features, for the task of Arabic handwriting recognition. Both obtained from raw pixels with two different ways. In addition, both appear to be an arduous and difficult task particularly in the case of the unconstrained Arabic handwriting which has high variability. An Explicit hand crafted features frequently extracted without taking into account the specificity of the studied problem. In contrary, we assume that CNN could be suited and beneficial for the Arabic script. it can extract subtle features which allow to separate confusing words and handling shape variations, which will likely be key in handling the excessive cursiveness of the Arabic handwritten [30]. To confirm our hypothesis, we use our baseline HMM system for Arabic word handwritten Rabi et al. [1], that takes into consideration the context of character using a relevant technique of cross learning. We consider two strategies-on the one hand, handcrafted features-HMM, on the other hand, CNN-features-HMM. We evaluate the performance of each strategy on the publicly available IFN/ENIT database. Experimental studies reveal that the proposed CNN-features-HMM based method shows satisfactory classification accuracy and outperformed handcrafted-features-HMM and some other exiting methods.

The rest of this paper is organized as follows: Sect. 2 presents the CNN model, HMM modeling and a brief overview over two proposed strategies. Experimental results are given and analyzed in Sect. 3. Finally, Conclusions and perspectives are drawn in Sect. 4.

2 Method

We present in this section the CNN principle, the used Markov modeling and the architecture of the CNN-HMM combination.

2.1 CNN Model

Since the work of Krizhevsky et al. [2] deep neural networks becomes the most prominent trend in virtually all fields of automatic learning statistical applications [3, 34-37]. With their introduction in this area, major improvements have been observed. As a result, impressive performance systems. In addition, the most widespread neural networks known as Convolutional Neural networks (CNN), developed initially by Lecun et al. [38], for the task of reading handwritten digits. Instead of using the more complicated architecture AlexNet [2], OverFeat [31], GoogLeNet [32], VGGNet [33], ResNet [34]. Our CNN architecture (illustrated in "Fig. 1", see below) is similar to LeNet-5 [5] with some modifications (without the second fully connected layer). The adopted structure comprises two convolutional layers with 5×5 receptive fields (i.e., kernel) and two sub-sampling layers over non-overlapping regions of size 2×2 with fully connected and output layers. Our CNN with each one input image of size 28×28 (passed through the stack of its layers) is represented by $1 \times 28 \times 28-6C2S-12C2S$. In following, convolutional layers are labelled C_x , sub-sampling layers are labeled S_x , where x is the layer index.



Fig. 1. A typical CNN architecture for proposed CNN for offline Arabic handwriting recognition

The first convolution layer (C_l) has 6 feature maps with 784 nodes/neurons each (28 × 28 pixels image), each is obtained by applying a distinct kernel of 5 × 5, that contains 25 weights, and a bias. So that it can extract different types of local features. The use of a kernel such a size converts 28 spatial dimension to 24 (i.e., 28 - 5 + 1) spatial dimension. Therefore, each 1st level feature map size is 24×24 . Each feature map has different set of weights. For example, C_l contains $25^*6 + 6 = 156$ trainable parameters, and (24*24*25*6) + (24*24*1*6) = 89656 connections.

All the nodes in a feature map share the same set of weights and so they are activated by the same features at different locations. This weight sharing not only provides invariance to local shift in feature position but also reduces the true number of trainable parameters at each layer. This local receptive field can extract the visual features such as oriented edges, end-points, corners of the images. Obtained results using (C_1) are illustrated in "Fig. 2".



Fig. 2. Visualization of convolutional layer C1 (6 feature maps produce by using 6 distinct kernels).

In 1st sub-sampling/pooling layer (S_I), the 1st level feature maps are down-sampled from 24 × 24 into 12 × 12 feature maps by applying the max pooling method that checks for the maximum value on its local receptive field, multiplies it by a trainable coefficient, adds a trainable bias and passing through an activation function for generating the output. More formally it can be shown as follows in (1):

$$x_j^l = f\left(w_j^l sub\left(x_j^{l-1}\right) + b_j^l\right) \tag{1}$$

Where sub(.) represents a sub-sampling function through local region; *w* and *b* are multiplicative coefficient and additive bias, respectively. In this study, we use sub(x) = Max(x) and a none overlapping scheme (i.e., stride = 2) and a 2 × 2 region, so the output image becomes 2-times smaller of the convolution layer. In addition, this sub-sampling operation reduces both the spatial resolution of the feature map and sensitivity to shift and distortions. "Figure 3", shows the results obtained by (S_1).



Fig. 3. Visualization of Sub-sampling layer S1(6 feature maps 2-times smaller of the convolution layer).

In the same way, the following layers (C_2 and S_2) have the same utility as previous layers (C_1 and S_1). Indeed, second convolution and 2nd sub-sampling operations are similar to 1st convolution and 1st sub-sampling operations, respectively.

The second convolutional operation (C_2) provides 12 different feature maps; a kernel size of 5 × 5 generates a feature map with size of 8 × 8. In total this layer uses 12 distinct filters and produces 12 different feature maps. Then 2nd sub-sampling (S_2) operation resizes each feature map to size of 4 × 4. When training this architecture, the feature maps of (S_2) are merged into a feature vector feeds into the fully connected layers. it means that these 12 feature maps values are considered as 192 (= 12 × 4 × 4) distinct nodes those are fully connected to 946 units (the output nodes) represents the size of vocabulary of the IFN/ENIT dataset (946 town/village names). The following Table 1 summarizes the architecture of our proposed CNN.

Layer	Layer type	Size	Output shape
1	Convolution	28 5×5 filters	(6@24)
2	Max Pooling	2×2 , stride 2	(6@12)
3	Convolution	24 5×5 filters	(12@8)
4	Max Pooling	2×2 , stride 2	(12@4)
5	Fully Connected	192 units	-
6	Softmax	946 units	-

Table 1. Architecture of Cnn

As in classical feed-forward neural networks, in our CNN, we introduce the non linearity by applying the non-linear function ReLU as in (2):

$$f(x) = \max(0; x) \tag{2}$$

The choice of ReLU instead other non-linearities functions is justified by the work of the Nair and Hinton [35].

2.2 HMM Modeling

The problem of recognizing the Arabic words can be viewed as characters sequence recognition. Let *I* is an Arabic word image which is composed of a set of characters. The modeling of the whole word image is obtained by the concatenation of the sequence of characters arranged horizontally. Each word can be segmented implicitly on units (characters or graphemes). We deal these units as being observed sequentially from a Markov model that pass through states $S = s_1, s_2,...,s_k$. That justifies the use of HMM, a sequence of length *T* is denoted as $O = o_1, o_2,...,o_T$. in which o_i corresponds to the *i*th units. Define $Y = y_1, y_2,...,y_L$ as the label of the image. *L* is the number of units in the image, y_i is the *i*th unit's label.

In this study, the used approach is analytical and based on character modelling by HMM. In total, 167 character models HMMs are built [36] including:

- The 26 basic characters of the Arabic alphabet.
- Certain characters whose shapes are different according to their position in the Word (at the beginning, middle or end).
- The characters with the presence of additional marks.
- The spaces inter pseudo word.

As shown in "Fig. 4", the model architecture $\lambda = \langle \Pi | A | B \rangle$ of a character is right-left topology, where λ represents the HMM. The key parameters of λ are the initial state probability distribution $\pi = p(q_0 = s_i)$, the transition probabilities $a_{ij} = p(q_t = s_j | q_{t1} = s_i)$, and a model to estimate the observation probabilities $p(o_t | s_i)$.



Fig. 4. Character HMM topology

There is no specific theory to set the number of hidden states in character model, often the solution is empirical. Word model is built by concatenating the appropriate character models "Fig. 5".



زنوش :Fig. 5. HMM model for Arabic word

The words models HMMs λ_w training is exactly the arduous task of a recognition system. The CNN features obtained from each image words using the pre-trained CNN or the extracted handcrafted features are considered as sequences of observations. We seek to deduce the model that generated them.

Once the topologies of the models λ_w were chosen, details of the procedure are explained above, training allows to re-estimate the parameters of each word model HMM λ_w (the probabilities of input, transitions and emissions), which allows to model the samples of the dataset. To do this, technically, we determine the parameters of $\lambda_w = (\Pi_w | A_w | B_w)$ that maximize the likelihood of the observations sequence $O = \{o_i\}$ o_2, \dots, o_n . The training is performed with Baum-Welch algorithm [37] under maximumlikelihood (ML) criterion until the likelihood converges. The best found HMM of each word is saved. Then, all resulting models consisting are the reference models of our system. After the learning phase, Recognition of a word image is performed by maximum a posteriori (MAP) estimation. Given an observation sequence O, we want to find the label sequence S that satisfies $S = \arg \max_{x} \log p(S/O)$ We use Viterbi algorithm [38] to get the most probable state sequence. It allows to decode the best state sequence candidates based on a criterion of maximum likelihood. Practically, it takes the word to be recognized as a sequence of observations $O = o_1, o_2, ..., o_n$ extracted from the image and determines the sequence of states $S = s_1, s_2, ..., s_n$ that has the maximum probability of generating O.

2.3 The Used Strategies

In this section, we present the overview of each used strategy to compare CNNs features and handcrafted features for offline Arabic handwriting recognition. The CNN-features-HMM model is shown in "Fig. 6", the system was developed to integrate the CNN and the HMM classifiers. We use HMM to model the dynamics of Arabic handwriting and CNN is employed to extract salient features. As illustrated in the diagram, the normalized input images are provided to the first convolutional layer and the designed CNN is trained by stochastic gradient descent (SGD) with momentum [39]. Our HMM baseline is trained by a new features vector obtained from the outputs of the hidden layer (FCL). Once the HMM classifier has been well trained, it performs the recognition task and makes new decisions on testing images with such automatically extracted features.



Fig. 6. Structure of the strategy CNN-features-HMM.

On another hand, the handcrafted features-HMM architecture is shown in "Fig. 7". Extraction features is preceded by baseline estimation; and the extracted features are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image. The sliding windows are shifted in the direction of writing (right to left). In each window we extract a set of 28 features represent the distribution features based on foreground pixels densities and concavity features. Each window is divided into a fixed number n of cells. Some of these features are extracted from specific areas of the image delimited by the word baselines.



Fig. 7. Structure of the strategy Handcrafted-features-HMM.

3 Experiments and Results

This section describes the details of our experiments. We present the used dataset and we describe the chosen initial hyper-parameters of our CNN and we demonstrate that the strategy CNN-features-HMM outperforms that of hand-engineered features.

In this experiment, on the one hand we used the KERAS [40] tool with TensorFlow backend, which is an open source of deep learning written in python, for implementing our CNN. On the other hand, we have used the toolbox HTK (Hidden Markov Model Toolkit [41]) to realize our baseline HMM system. All experiments are conducted on a regular PC (2.7 GHz 4-core CPU, 4G RAM and Windows 64-bit OS).

To evaluate which are the relevant features between those two kinds, we use the IFN/ ENIT database that consists of 946 handwritten Tunisian city names and their corresponding postcodes. The old version (v1.0p2 version) of the database contains 26,459 Arabic names handwritten by 411 different people.

Our CNN was trained on this dataset. We split 10% of the training set as validation set. the feed-forward net is trained under cross entropy objective by stochastic gradient

descent (SGD) with momentum until the training process converges (Stability of the error). We use this optimization method with a momentum set to 0.9, a mini-batch size of 50 and The base learning rate was initialized for all trainable parameters at 0.01, and we adjust it manually during the training process, by dividing it by 10 when the validation set performance stops improving. We decrease the learning rates 3 times before stopping the training process, which is terminated at epochs 20.

Several experiments were performed to evaluate the recognition rate of our system according to the kind of features. All tests that have been done were on test scenario "abc-d", three subsets (abc) are used for training and validate our approach and another one (d) for testing. The first one uses the handcrafted features for training and test. However, the second based on CNN features. Table 2 shows the obtained results of these tests on scenario "abc-d" using the both kind of features.

able 2. Recognition Rate on sechario deb				
Models	RR* %			
HMM	87.93			
HMM	88.95			
	Models HMM HMM			

Table 2. Recognition Rate on scenario acb-d

*Recognition Rate

The experimental results shown in Table 2, demonstrate that the strategy CNN-features- HMM outperforms the second strategy handcrafted-features-HMM. Using a CNN codes, we achieve a rate of 88.95%, involving an increase in the accuracy by 1.02% compared to using hand engineered features. This shows that the use of CNN codes reliably improves the recognition rate.

A comparative study of the performance of our suggested learning CNN features was also performed with other results of different approaches published on the same database. In this context, our results are statistically important compared to accuracies achieved by the various offline systems recognition of cursive Arabic handwritten of the state-of-the-art on the scenario abc-d of IFN/ENIT database. (see Table 3).

System	Features/Models	RR*%
Irfan et al. [42]	Handcrafted features/HMM	85.12
Alkhateeb et al. [43]	Handcrafted features/HMM + re-ranking	83.55
Maqqor et al. [44]	Handcrafted features/Multiple classifier	76.54
ElMoubtahij et al. [45]	Handcrafted features/HMM	78.95
Khaoula et al. [46]	Handcrafted features/DBN	82.00
Rabi et al. [1]	Handcrafted features/HMM	87.93
Present work	CNN features/HMM	88.95

Table 3. Gives the comparative results on scenario abc-d

As it can be noted from Table 3, most of the previous systems are based on HMM and hand-crafted features-based approach. However, our suggested model CNN based-HMM instead to use hand engineered features, it extract automatically and directly the relevant features from the image of word. In addition, as shown in the table III our system

outperforms the results obtained with other current methods a significant achievement was made with the recognition rate of 88.95% on the scenario "abc-a". This prove the effectiveness of CNN model, specially its ability to generate a salient features directly from word. In fact, CNN, with automatic feature extractor stage, deduces features that differentiate between words, and then HMM classifier insists on predicting the correct class of word. These learned features, being more robust than computed hand-crafted features, establish an adequate representation for words.

4 Conclusion and Perspectives

In this work, a comparison has been conducted between the learning CNN features and the handcrafted features. The study was made on the basis of our baseline HMM system. In order to determine the salient features among those two kinds. We have considered two strategies 'CNN-features-HMM' and 'Handcrafted-features-HMM'. The first combination took the CNN as an automatic feature extractor and HMM as recognizer. That allows to operate directly on the images and extracting relevant characteristics without much emphasis on feature extraction and pre-processing stages. We showed that this strategy gives a promising results on IFN/ENIT which significantly outperforms the second strategy, which based on hand-engineered features and our HMM baseline system. Contrary to this strategy that use hand-crafted features which is a laborious and time consuming task, the most important advantage of the fusion of CNN and HMM is the ability to extract automatically salient features directly from raw pixels.

Despite to these promising and encouraging results. It seems that our CNN-features-HMM is not optimal. A set of Adjustment can be envisioned either for the CNN or for the HMM. In the case of CNN, a number of parameters might be tuned empirically: the structure of CNN, the size of batch, the number and nature of used kernels, the used training technique, the kind of non linear functions utilized in model, etc. Furthermore, additional gain in performance was obtained by adding dropout technique in the fully connected hidden layers. As to HMM, for example improvements might be made based on the type of character topology and the fine tuning of GMM employed in each state. As future work, extracted CNN features will be processed by an enhancing HMM using statistical language models that are incorporated as a post-processing into the process of recognition.

References

- Rabi, M., Amrouch, M., Mahani, Z., Mammass, D.: Recognition of cursive Arabic handwritten text using embedded training based on HMMs. In: Engineering & MIS (ICEMIS, INSPEC Accession Number: 16467172. IEEE (2016). doi:10.1109/ICEMIS.2016.7745330
- 2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25 (2012)
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, S., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Li, F.F.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision 115(3), 211–252 (2015)

- Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. 2(1), 1–127 (2009)
- LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: International Symposium on Circuits and Systems, pp. 253–256 (2010)
- 6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436-444 (2015)
- Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning. In: NIPS Workshop on Deep Learning and Unsupervised Feature Learning, vol. 2011, p. 4 (2011)
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Comput. 1(4), 541–551 (1989)
- Albeahdili, H.M., Alwzwazy, H.A., Islam, N.E.: Robust convolutional neural networks for image recognition. (IJACSA) Int. J. Adv. Comput. Sci. Appl. 6(11) (2015)
- Kaiming, H., Xiangyu, Z., Shaoqing, R., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition European. In: Conference on Computer Vision. arXiv:1406.4729v4 [cs.CV] (2015)
- Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: The International Joint Conference on In Neural Networks (IJCNN), pp. 2809–2813. IEEE (2011)
- Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Proces. Mag. 29(6), 82–97 (2012). IEEE
- Sharma, A., PramodSankar, K.: Adapting off-the-shelf CNNs for word spotting & recognition. In: International Conference on Document Analysis and Recognition, pp. 986– 990 (2015)
- Simard, P.Y., Steinkraus, D., Platt, J.C. Best practices for convolutional neural networks applied to visual document analysis. In: International Conference Document Analysis and Recognition, pp. 958–962 (2003)
- Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A neural probabilistic language model. J. Mach. Learn. Res. 3, 1137–1155 (2003)
- Collobert, R., Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning, pp. 160–167. ACM (2008)
- 17. Couprie, C., Farabet, C., Najman, L., LeCun, Y.: Indoor semantic segmentation using depth information. In: International Conference on Learning Representation (2013)
- Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR, abs/1311.2524 (2013)
- Ciresan, D., Meier, U., Masci, J., Schmidhuber, J.: A committee of neural networks for traffic sign classification. In: The 2011 International Joint Conference on in Neural Networks (IJCNN), pp. 1918–1921. IEEE (2011)
- 20. LeCun, Y., Bottou, L., Bengio, Y.: Reading checks with multilayer graph transformer networks. In: International Conference on Acoustics, Speech, and Signal Processing (1997)
- Bluche, T., Ney, H., Kermorvant, C.: Tandem HMM with convolutional neural network for handwritten word recognition. In: 38th International Conference on Acoustics Speech and Signal Processing (ICASSP2013), pp. 2390–2394 (2013)
- 22. Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A.: Synthetic data and artificial neural networks for natural scene text recognition. arXiv preprint arXiv:1406.2227 (2014)

- Yuan, G.B., Jiao, L., Liu, Y.: Offline handwritten English character recognition based on convolutional neural network. In: 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 125–129 (2012). doi:10.1109/DAS.2012.61
- 24. Goodfellow, I.J., Bulatov, Y., Ibarz, J., Arnoud, S., Shet, V.: Multi-digit number recognition from street view imagery using deep convolutional neural networks. In: ICLR (2014)
- 25. Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Deep big simple neural nets excel on handwritten digit recognition, CoRR, abs/1003.0358 (2010)
- Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Convolutional neural network committees for handwritten character classification. In: International Conference of Document Analysis and Recognition, vol. 10, pp. 1135–1139 (2011)
- 27. Cireşan, D., Schmidhuber, J.: Multi-column deep neural networks for offline handwritten Chinese character classification. arXiv preprint arXiv: 1309.0261 (2013)
- Graham, B.: Sparse arrays of signatures for online character recognition. arXiv:1308.0371 (2013)
- Yin, F., Wang, Q.F., Zhang, X.Y., et al.: ICDAR 2013 chinese handwriting recognition competition. In: Proceedings 12th International Conference Document Analysis and Recognition, pp. 1464–1470 (2013)
- Parvez, M.T., Mahmoud, S.A.: Offline Arabic handwritten text recognition: A survey. ACM Comput. Surv. 45(2), 23–35 (2013)
- 31. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., Le-Cun, Y.: Overfeat: integrated recognition, localization and detection using convolutional networks. CoRR (2013)
- 32. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. CoRR, vol. abs/1409.4842 (2014)
- 33. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv Technical report (2014)
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. https:// arxiv.org/abs/1512.03385 (2015)
- Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807– 814 (2010)
- El-Hajj, R., Likforman-Sulem, L., Mokbel, C.: Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition. IEEE PAMI 31(7), 1165–1177 (2009)
- Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Stat. 41(1), 164– 171 (1970). http://dx.doi.org/10.1214/aoms/1177697196
- Forney Jr., G.D.: The Viterbi algorithm. Proc. IEEE 61(3), 268–278 (1973). http://dx.doi.org/ 10.1109/PROC.1973.9030
- Sutskever, I., Martens, J., Dahl, G., Hinton, G.: On the importance of initialization and momentum in deep learning. In: JMLR W & CP, vol. 28(3), pp. 1139–1147 (2013)
- 40. Keras (2016). https://github.com/fchollet/keras
- 41. Young, S., et al.: The HTK Book V3.4. Cambridge University Press, Cambridge (2006)
- Irfane, A., Fink, G., Mahmoud, S., et al.: Improvements in sub-character hmm model based arabic text recognition. In: 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 537–542. IEEE (2014)
- Alkhateeb, J.H., Ren, J., Jiang, J., Al-Muhtaseb, H.: Offline handwritten arabic cursive text recognition using hidden markov models and re-ranking. Pattern Recogn. Lett. 32, 1081– 1088 (2011)

- Maqqor, A., Halli, A., Satori, K., Tairi, H.: Off-line recognition Handwriting combination of multiple classifiers. In: 3rd International IEEE Colloquium on Information Science and Technology, IEEE CIST 2014 (2014)
- 45. El Moubtahij, H., Akram, H., Satori, K.: Using features of local densities, statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition (2016)
- 46. Jayech, K., Mahjoub, M.A., Amara, N.B.: Arabic handwritten word recognition based on dynamic bayesian network (2016)

SCH-WSD: A Semantic-Conceptual Hybrid Approach for Web Services Discovery

Hicham Laabira^(⊠), Khalid El Fazazy, and Redouane Ezzahir

Laboratory of Systems Engineering and Information Technology (LiSTi), ENSA, Ibn Zohr University, Agadir, Morocco hicham.laabira@edu.uiz.ac.ma, {k.elfazazy,r.ezzahir}@uiz.ac.ma

Abstract. Web services discovery occupies a crucial part in the semantic web, as it aims to return the most relevant web services that better meet the user's needs. In this paper, we propose a new client-side web services architecture, designed to improve the performance of web services discovery. It is based on a hybrid approach that includes both the semantic and conceptual approach, which we called SCH-WSD (Semantic-Conceptual Hybrid approach for Web Services Discovery). SCH-WSD measures the degree of similarity between the queries and the web services using the inputs, outputs and category as matched elements. Ontology Web Language for Services (OWL-S) is used as semantic web services description language. A theoretical analysis and an experimental evaluation on some benchmarks illustrate the practical effectiveness of our approach and its ability to provide users with web services that perfectly meet their requirements.

Keywords: Web service \cdot Web service discovery \cdot Services architecture \cdot Hybrid approach \cdot OWL-S

1 Introduction

Web services description is an important step in web services consumption cycle adopted by service providers. Indeed, to make their services visible in web services directories, the providers are called to write and publish specific interfaces that define elements and access points of their web services, using standardized technologies like Web Services Description Language (WSDL) [5]. Several recommendations have been proposed to facilitate semantic description of web services such as Semantic Annotations for WSDL and XML Schema (SAWSDL) [11], Web Service Modeling Ontology (WSMO) [12] and Ontology Web Language for Services (OWL-S) [6, 7]. SAWSDL is a technique for the web services description based on annotations designed to improve and overcome some limitations of WSDL [11]. This means that SAWSDL extends and imports semantic annotation to non-functional features like Quality of Service (QoS). WSMO is an ontological approach that essentially aims at giving a semantic description of web services on a Web Service Modeling Language (WSML) that describes all elements of WSMO. OWL-S is an extensible and

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_15 expressive high level ontology formalized by Web Ontology Language (OWL) [2]. The purpose of this ontology is to facilitate the discovery and composition of web service by specifying three conceptual parts in any service description: *ServiceProfile*, *ServiceModel* and *ServiceGrounding*.

Web services discovery is the process of finding the appropriate web services that reflect the user's desired requirements. It occupies an essential step in semantic web. Several methods and techniques have been proposed for improving the performance of web services discovery, a good taxonomy of them is presented in [25]. From [16, 17], the approaches for web services discovery can be classified into three broad categories:

- 1. Non-logical approaches: Encompass various technologies such as Universal Description, Discovery, and Integration (UDDI) [1] which supports only keyword based matching [13]. Among the research methods that exploit syntactic similarity metrics, we can find iMatcher1 [18] and DSD Matchmaker [23]. The iMatcher1 is a syntactic matching mechanism founded on four basic metrics, while DSD-Matchmaker is a graph-based discovery technique [23]. All of these methods are focused on syntactic matching techniques between the user's request and the advertised services. Consequently, unsuitable services can be returned or ignore certain services even if they have a good semantic similarity. To address this problem, some studies have investigated the concept of semantic in logical approaches.
- 2. Logical approaches: Based on ontologies and logical formalisms to describe the semantic web services. They utilize the same type of matching that considers the semantic of functional features. However, in terms of matched elements, each one of them has its own properties and features. References [4, 24] use a matching focused only on "Inputs and Outputs (IO)" of web service, but with higher levels of matching to those used in the non-logical approaches. In addition, [20] exploits not only "Inputs and Outputs" but also "Category (IOC)", whereas some works like [8, 10] adopt a semantic matching process based on "Inputs, Outputs, Preconditions and Effects (IOPE)", to estimate the degree of similarity between the user's requirements and web service.
- 3. Hybrid approaches: Merge both previous categories through syntactic similarity metrics. Several works in this category try to provide powerful approaches for web services discovery such as OWLS-MX [9], OWLS-iMatcher2 [19] and FC-Match [21]. However, all these approaches rely on syntactic similarity metrics often lead to problems related to the performance and accuracy of delivered results.

In order to improve the performance and accuracy of web services discovery results, we propose a Semantic-Conceptual Hybrid approach for Web Services Discovery (SCH-WSD), which combines the best features of existing approaches and further enhance them by mixing their matching levels and similarity functions.

The remainder of the paper is structured as follows. Section 2 describes our Semantic-Conceptual Hybrid approach for Web Services Discovery. We start by proposing a new client-side web services architecture on which we incorporate our hybrid approach. Then, we present similarity approaches. In Sect. 3, the experimental results are discussed and Sect. 4 concludes the paper.

2 Semantic-Conceptual Hybrid Approach for WSD

In this section, we propose SCH-WSD, a hybrid process for semantic web services discovery, which returns to users the web services adapted to their expectations. But, before going to technical details, we report in the following subsection a client-side architecture where we have integrated our hybrid approach.

2.1 Architecture

Web services architecture provides a conceptual model and a context for understanding the web services and the relationships between components of this model. Figure 1 presents a new client-side architecture for web services discovery. This architecture is focused on a matching process, and it allows us to measure with an effective manner the degree of similarity between the user's request and the published service. The main components involved in this architecture are the following:



Fig. 1. Overview of our Architecture.

1. Web Services Registry: This register is considered as a database of web services profiles, which will be necessarily used in the web services discovery. The web services registry stores and standardizes the descriptions published by service providers. In our work, all these descriptions are described by using OWL-S ontological approach.

- 2. **Ontologies Registry:** Includes a set of domain ontologies formalized by OWL for knowledge representation and modelling. These ontologies have a hierarchical structure in which vertices present the concepts and edges reflect the links between the basic concepts. This register has a fundamental role in web services discovery manager, which will be used later by our approach to match "IOC" or "IO" declared in web service description with those expressed in a query.
- 3. Web Services Discovery Manager: The component on which we incorporate the main idea of our contribution. The principal objective of this manager is to provide best web services that satisfy the user's needs. It measures the degree of matching between the user's request and advertised web service, and provides the scores of semantic and conceptual matching. Both matching techniques are mutually triggered when this manager receives a given request.
- 4. **Control Manager:** The two main functionalities performed by this manager are: (i) *Results Aggregation* that aggregates the scores provided by the web services discovery manager, and (ii) *Results Thresholding* that uses a determined threshold in order to control the previous aggregated results. The purpose of this control is to exclude web services that have a similarity scores less than a chosen threshold. As a result, user receives a relevant list of the most appropriate web services.

The following subsection describes different similarity approaches managed by the web services discovery manager.

2.2 Similarity Approaches

SCH-WSD relies on a crucial phase in its functioning, this phase includes two types of pairings for determining the degree of matching between a given request and a web service.

Semantic Matching. Reference [15] defines five levels of similarity, inspired from [24]. SCH-WSD uses the same levels but with a new adaptation of scores (i.e. degrees) so as to match inputs, outputs and category, as illustrated in Table 1. In addition, the degree of two levels *Plugin* and *Subsume* varies according to the distance D (i.e. number of links) between two matched concepts. This distance is used to avoid giving the same score of similarity to two concepts in the domain ontology, which are related to a parent concept but with two different distances. Note that we modified the

Matching level	Description	Degree
Exact	Proposed and Required concept are the same	3
Plugin	Proposed concept is more general than Required concept	$2 + \frac{1}{1 + D}$
Subsume	Proposed concept is more specific than Required concept	$1 + \frac{1}{1 + D}$
Sibling	Proposed and Required concept are two direct sons of a	0.5
	parent concept	
Fail	Otherwise	0

Table 1. Description of matching levels with their degrees.

description of *Plugin* and *Subsume* only to match inputs. In this case, *Plugin* means that required input is more general than proposed input, and *Subsume* means that required input is more specific than proposed input.

Conceptual Matching. Conceptual matching defines the measures of conceptual similarity between a consumer's request and a published service. Several measures of conceptual matching have been proposed to determine the similarity value between two given concepts [3, 14, 22]. In WuP [14], the authors calculate the degree of similarity between two concepts C_1 and C_2 based on a conceptual similarity metric. This metric is focused on the distance from Lowest Common Subsumer (LCS), as illustrated in Fig. 2 and defined according to (1).



Fig. 2. Relationships used in WuP.

$$SIM_{wup}(C_1, C_2) = \frac{2 \times N3}{N1 + N2 + 2 \times N3}$$

=
$$\frac{2 \times depth(Lcs(C_1, C_2))}{depth(C_1) + depth(C_2)}$$
(1)

Where:

 $Lcs(C_i,C_j)$: denotes the Lowest common subsumer between two given concepts C_i and C_j .

depth(C_i): denotes the length of the path from C_i to the root in an ontology.

We investigated the measure WuP and we found that WuP can fail in an interesting special case. Indeed, using WuP, the degree of similarity between two sibling concepts C_1 and C_2 that have the same parent is greater than or equal to two concepts C_1 ' and C_2 ' that are located in the same ontological path (i.e. C_1 ' is ancestor of C_2 ' or the reverse). However, we can easily notice that C_1 ' and C_2 ' are more similar than C_1 and C_2 .

To address the aforementioned problem, we define in (4) a refinement of WuP named SIM_{depth} , in order to evaluate exactly the conceptual similarity between two

concepts C_1 and C_2 . This metric relies on the use of hierarchical structure of concepts, and more precisely on arcs of domain ontologies that define the matched concepts. The value of similarity between two concepts by this metric is always between 0 and 1, and it can be computed as follows:

• If C₁ is ancestor of C₂ or the reverse

$$SIM_{depth}(C_1, C_2) = SIM_{wup}(C_1, C_2)$$
⁽²⁾

• Otherwise

$$SIM_{depth}(C_1, C_2) = \frac{2 \times depth(Lcs(C_1, C_2))}{depth(C_1) + depth(C_2) + f_{depth}(C_1, C_2)}$$

$$= \frac{2 \times depth(Lcs(C_1, C_2))}{depth(C_1) + depth(C_2) + depth(C_1) + depth(C_2)}$$

$$= \frac{2 \times depth(Lcs(C_1, C_2))}{2 \times (depth(C_1) + depth(C_2))}$$

$$= \frac{1}{2} \times SIM_{wup}(C_1, C_2)$$
(3)

From (2) and (3), SIM_{depth} is summarized according to (4):

$$SIM_{depth}(C_1, C_2) = \begin{cases} SIM_{WUP}(C_1, C_2), C_1 \text{ is ancestor of } C_2 \text{ or the reverse} \\ \frac{1}{2} \times SIM_{WUP}(C_1, C_2), \text{ Otherwise} \end{cases}$$
(4)

To illustrate the efficiency of our refined conceptual metric, we calculate and compare its values with those of WuP by using an extract from travel ontology¹, as highlighted in Fig. 3. In this example, we can clearly notice that *Destination* and *Capital* are more similar than *UrbanArea* and *RuralArea*. Indeed, the concept *Capital* is an indirect subclass of *Destination* in this ontology (i.e. *Capital* is a subclass of *City*, *City* is a subclass of *UrbanArea* and *UrbanArea* is a subclass of *Destination*), while *UrbanArea* and *RuralArea* are two opposing concepts. Now, we do the calculation and comparison:

- SIM_{wup}(Destination, Capital) = $\frac{2 \times 1}{1 + 4} = 0.4$
- SIM_{wup}(UrbanArea, RuralArea) = $\frac{2 \times 1}{2 + 2} = 0.5$
- SIM_{depth}(Destination, Capital) $= \frac{2 \times 1}{1+4} = 0.4$
- SIM_{depth}(UrbanArea, RuralArea) = $\frac{1}{2} \times \frac{2 \times 1}{2+2} = 0.25$

¹ http://projects.semwebcentral.org/projects/owls-tc/.



Fig. 3. An extract from travel ontology

The values obtained by our refinement of WuP are agreed with the expectation and then solves the problem observed in WuP.

3 Experiments and Discussion

In this section, we present the experiments that we performed to assess the performance of our SCH-WSD compared to some existing approaches. We appraise our experiments by using Java as programming language, OWLS-TC (See footnote 1) is an artificial collection of web services with their domain ontologies, Pellet² a reasoner integrated with Jena³ ontology API and OWLS-MX v2.0 to calculate all hybrid variants of OWLS-MX. In order to evaluate the effectiveness of SCH-WSD, we used a weighted average of *Precision* and *Recall* called *F-measure*. This metric is calculated by (7), where it reaches its best score at 1 and worst at 0. In our work, the *Precision* represents the division of relevant services provided by SCH-WSD on the total number of retrieved services, it is calculated by (5). The *Recall* represents the division of relevant services provided by SCH-WSD on the total number of relevant services provided by SCH-WSD on the total number of relevant services, it is calculated by (6).

$$Precision = \frac{|Relevant Services \cap Retrieved Services|}{|Retrieved Services|}$$
(5)

² https://github.com/Complexible/pellet.

³ http://jena.apache.org/.

$$Recall = \frac{|Relevant Services \cap Retrieved Services|}{|Relevant Services|}$$
(6)

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$
(7)

In order to choose the two thresholds values used in SCH-WSD *threshold_{ioc}* and *threshold_{io}* for matching Inputs, Outputs and Category (IOC-Matching), and for matching Inputs and Outputs (IO-Matching) respectively, we conducted a series of tests on web services specified in OWLS-TC. Based on provided results, we have chosen *threshold_{ioc}* = 0.75 for IOC-Matching and *threshold_{io}* = 0.7 for IO-Matching. After that, so as to evaluate the effectiveness of our approach SCH-WSD on some benchmarks, we carried out experiments on two categories of samples, in which each sample is composed of a query and a list of web services. The results below show this evaluation for IOC-Matching and IO-Matching.

3.1 IOC-Matching

Figure 4 presents the performance of SCH-WSD compared to two approaches syntactic and hybrid (i.e. syntactic and semantic) approach by using IOC-Matching. As it is seen from this figure, the maximum values of *F-measure* obtained by these two approaches cannot reach to that obtained by SCH-WSD (*F-measure* = 1 in Fig. 4a and b when *threshold_{ioc}* = 0.75). Indeed, despite the hybrid approach is designed to improve the performance of results returned by the syntactic approach, it also suffers from precision. Because it uses a measure of syntactic similarity of cosine type in its functioning, which diminishes the overall effectiveness of this approach.

3.2 IO-Matching

Figure 5 displays the performance of SCH-SD compared to hybrid variants of OWLS-MX (OWLS-M1, OWLS-M2, OWLS-M3 and OWLS-M4) by using IO-Matching. Focusing on this figure, we notice that SCH-WSD has a better *F-measure* compared to all hybrid variants of OWLS-MX. Indeed, whatever the maximum value of *F-measure* of each hybrid variant cannot achieve to that reached by SCH-WSD (*F-measure* takes 1 in Fig. 5c and 0.9 in Fig. 5d when *threshold*_{io} = 0.7). The major drawback of OWLS-MX is that it requires user intervention in its functioning. As a result, it decreases the precision of results provided to users in the case where this intervention is poorly controlled. Further, all hybrid variants of OWLS-MX are based on syntactic measures for IO-Matching, which often provoke some problems related to the reliability of retrieved results.



Fig. 4. F-measure of *SCH-WSD* vs. *Syntactic Approach* vs. *Hybrid Approach*: (a) first sample and (b) second sample.



Fig. 5. F-measure of *SCH-WSD* vs. hybrid variants of *OWLS-MX*: (c) first sample and (d) second sample.

4 Conclusion

In this paper, we have proposed a client-side web services architecture to improve the web services discovery. This architecture is founded on a hybrid approach called Semantic-Conceptual Hybrid approach for Web Services Discovery (SCH-WSD). Afterwards, we have also highlighted the different aspects involved in SCH-WSD. Finally, the performed experiments illustrate promising results for providing users with a set of suitable web services.

The heterogeneity of web services presents a challenge to apply SCH-WSD on different descriptions of web services, because each provider can publish his/her services through various descriptive models. For this, we need to create a consistent meta-model to transform any representations utilized by service providers under OWL-S model.

Acknowledgments. The authors wish to thank the National Center for Scientific and Technical Research (CNRST) in Morocco for funding this research under grant number 19UIZ2015.

References

- 1. Patil, N., Gopal, A.: Comparative study of mechanisms for web service discovery based on centralized approach focusing on UDDI. Int. J. Comput. Appl. 14, 28–31 (2011)
- Antoniou, G., Van Harmelen, F.: Web ontology language: OWL. In: Handbook on Ontologies, pp. 91–110. Springer, Berlin, Heidelberg (2009)
- 3. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE. Trans. Syst. Man. Cybern. **19**, 17–30 (1989)
- 4. Fethallah, H., Chikh, A.: Automated retrieval of semantic web services: a matching based on conceptual indexation. Int. Arab J. Inf. Technol. **10**, 61–66 (2013)
- Chinnici. R., Moreau. J.J., Ryman, A., Weerawarana, S.: Web services description language (WSDL) version 2.0 part 1: core language.W3C recommendation (2007)
- Martin, D., Burstein, M., Mcdermott, D., Mcilraith, S., Paolucci, M., Sycara, K., Srinivasan, N.: Bringing semantics to web services with OWL-S. World Wide Web 10, 243–277 (2007)
- Martin, D., Paolucci, M., McIlraith, S., Burnstein, M., McDermott, D., McGuinness, D., Srinivasan, N.: Bringing semantics to web services: the OWL-S approach. In: International Workshop on Semantic Web Services and Web Process Composition, pp. 26–42. Springer, Berlin, Heidelberg (2004)
- Keller, U., Lara, R., Lausen, H., Polleres, A., Fensel, D.: Automatic location of services. In: Proceedings of the 2nd European Semantic Web Conference (ESWC), pp. 1–16. Springer (2005)
- Klusch, M., Fries, B., Sycara, K.: Automated semantic web service discovery with OWLSMX. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 915–922 (2006)
- Stollberg, M., Keller, U., Lausen, H., Heymans, S.: Two-phase web service discovery based on rich functional descriptions. In: Proceedings of the 4th European conference on The Semantic Web, pp. 99–113. Springer (2007)
- 11. Kopecký, J., Vitvar, T., Bournez, C., Farrell, J.: SAWSDL: semantic annotations for WSDL and XML schema. IEEE Internet Comput. **11**, 60–67 (2007)

- Roman, D., Keller, U., Lausen, H., De Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., Fensel, D.: Web service modeling ontology. Appl. Ontol. 1, 77–106 (2005)
- 13. Khanam, S.A., Youn, H.Y.: A web service discovery scheme based on structural and semantic similarity. J. Inf. Sci. Eng. **32**, 153–176 (2016)
- Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting on Associations for Computational Linguistics, pp. 133–138 (1994)
- Najar, S., Pinheiro, M.K., Souveyet, C.: A new approach for service discovery and prediction on pervasive information system. In: Procedia Computer Science, pp. 421–428 (2014)
- 16. Bitar, I.E., Belouadha, F.Z., Roudies, O.: Semantic web service discovery approaches: overview and limitations. arXiv preprint arXiv (2014)
- Chabeb, Y., Tata, S., Ozanne, A.: YASA-M: a semantic web service matchmaker. In: Proceedings of the International Conference on Advanced Information Networking and Applications, pp. 966–973 (2010)
- Schumacher, M., Helin, H., Schuldt, H.: CASCOM: Intelligent Service Coordination in the Semantic Web. Springer, Birkhäuser Basel, Basel (2008)
- 19. Kiefer, C., Bernstein, A., Lee, H.J., Klein, M., Stocker, M.: Semantic process retrieval with iSPARQL. In: European Semantic Web Conference. Springer, Berlin, Heidelberg (2007)
- Jaeger, M., Rojec-Goldmann, G., Liebetruth, C., Mühl, G., Geihs, K.: Ranked matching for service descriptions using OWL-S. In: Proceedings of KiVS, pp. 91–102. Springer (2005)
- Bianchini, D., De Antonellis, V., Melchiori, M., Salvi, D.: Semantic-enriched service discovery. In: Proceedings 22nd International Conference on Data Engineering Workshops, pp. 38–38 (2006)
- Ehrig, M., Haase, P., Stojanovic, N., Hefke, M.: Similarity for ontologies a comprehensive framework. In: Proceedings of the 13th European Conference on Information Systems (2005)
- Klein, M., König-Ries, B.: Coupled signature and specification matching for automatic service binding. In: Proceedings of the European Conference on Web Services, pp. 183–197 (2004)
- Paolucci, M., Kawamura, T., Payne, T.R., Sycara, K.: Semantic matching of web services capabilities. In: Proceedings of the 1st International Semantic Web Conference, pp. 333– 347. Springer, Berlin, Heidelberg (2002)
- 25. Pakari, S., Kheirkhah, E., Jalali, M.: Web service discovery methods and techniques: a review. Int. J. Comput. Sci. Eng. Inf. Technol. 4, 1–14 (2014)

Optimal Regulation of Energy Delivery for Community Microgrids Based on Constraint Satisfaction and Multi-agent System

Mostafa Ezziyyani¹⁽⁾ and Loubna Cherrat²

¹ Faculty of Sciences et Techniques of Tangier, Abdelmalek Essaâdi University, Tangier, Morocco ezziyyani@fstt.ac.ma ² Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco cherratloubna2@gmail.com

Abstract. With the existence of several energetic resources and local production site by consumers a new strategy for managing the distribution of energy is indispensable. This paper aims to develop a simulation platform for energy resources management of a Micro Grids Network to optimize the electricity consumption. Using the remote control systems and data integration from distributed databases the system regulates automatically the distribution following the need of each customer and need of Micro Grid. The solution use an incremental search algorithm based on the total satisfaction of the constraints by priority order. In this paper, as software platform solution, we use the multi-agent system (MAS) technology. This choice is motivated by the functional ability of agents, and their selfadaptation to the environment (i.e. change the feature). The ability of the interaction between the agents and their mobility will define and specify the real-time needs of each Micro Grids according to its production and consumption capacity and the need of its neighbors. The functional architecture of the operating system is based on a graph, where each node can be a customer or producer of energy or both of them associated with list of requirement constraints. We used the principle of Distributed Databases to facilitate communication inter-agents and to optimize the time of data transfer between agents of different Micro Grids and simplified access "on demand" to the data with high availability. Thanks to the distributed databases solution, we can easily integrate the critical data on a data center and improve the response time of readjustment and equilibration of the electricity distribution and consumption.

Keywords: Optimization \cdot Integration \cdot Micro Grids \cdot Data base \cdot Big data \cdot Smart Grid \cdot Energy \cdot Agent \cdot Real time \cdot Constraints \cdot PCC \cdot Distributed databases

1 Introduction

Nowadays, the cost of energy resources becomes very high. Several countries spend the billions of dollars to find solutions to reduce this cost. This is usually done by the different ways: (1) the identification and development of new types of energy resources based for

instance on solar and (2) the exploitation of the usage of consumers (housing, electrical vehicle, etc.) of energy to adjust their consummation. This solution is less costly than the previous one, since it can be implemented by educating the consumers and making her/him more sensitive on energy cost. To achieve the reduction of energy a solution combining these both solutions is more feasible [1, 2].

In the meantime, a spectacular development of devices such as Smart Grids, sensors, tablets, smart phones, etc. may contribute in implementing solutions belonging to the previous categories (1) and (2). A Smart Grids contains advanced technology that enables enhanced, two-way communication between a utility and its customers. The resulting information provides customers with:

- Tools to help manage their energy
- Improved energy efficiency
- Improved reliability (fewer outages)

To manage perfectively the grid power generated through renewable generation sources, our project, entitled, SGiRE: "Smart Grld based Platform for Managing Renewable Energy" will serve as a blueprint for future Smart Grid implementation and will accelerate a realization of the "future Tool" safely delivers reliable electricity with greater efficiency and improved environmental performance. SGiRE will gain knowledge about customer (their profiles, consummation habitudes, etc.) needs and usage patterns. In addition, the Users of this tool will be able to gather information about Smart Grid's storage capabilities, supply and delivery.

The Smart Grid demonstration improvements will enhance service for the entire of the Micro Grids through improved service reliability, reduced energy delivery costs, more efficient energy consumption, and better information flow. Two basics feature:

- Smart Generation: The production cannot be carried out except if there is a need (Cost-effective solutions). The resources will be used to add renewable energy while reducing and shortening system outages.
- Smart Distribution: Develop the smart applications like automated meter reading, smart switches and smart capacitors. Improved customer delivery and service quality will result from future advanced technologies. This also provides the ability to communicate with customers on prices and conditions of the system.

Smart Grid provides real-time information that increases awareness of electricity use and identifies opportunities to reduce consumption and save money. And, Can automatically set temperatures based on season, resulting in up to 20% savings in heating and cooling bills. And, Helps customers understand the impact of electricity use and encourages them to conserve energy, help the environment and save money.

One of the major attribute of the smart grid is to integrate renewable and storage energy resources at the consumption premises. This project seeks design, implementation and testing of a system that integrates renewable and storage energy resources to a smart home. The proposed system provides and manages a smart home energy requirement by installing renewable energy. So implement wireless sensor network (WSN) technology to communicate control information among different components (solar energy generator, inverter, energy storage and conditioning, appliances, smart meter, etc.).

Interfacing renewable energy generator with the smart grid is an optimization problem that seeks to find the optimum solution among three conflicting parameters which are local generation (solar energy), local utilization, and external grid state (stability). The wireless sensor network needs to implement novel algorithms that will find the optimum state where the system is stable. A research question that will see to answer is how different components will communicate and does the health of the solar energy generator allow the energy to be flowed to the grid? If there is a surplus in the grid power, what is the optimum operational mode of the renewable generator given the local energy consumption and the state of the grid?

The main thrust of this project is that design and development a distributed system based on SMA technology with real interaction with WSN. That will enable the communication among different components as well as providing the processing capabilities to implement the algorithms necessary to coordinate and optimize the operation of the system as a whole [4–6].

2 Motivation and Impact on Socio-Economic

Today in Morocco, and like all other countries, energy is more expensive. The big concern is, the search for sustainable sources of energy, cheaper and less polluting. The solution converges to renewable energy, not centralized in nature, challenging the old logic. The kilowatt cleanest and cheapest being the cause. In addition, the development of digital technologies is an opportunity to rethink the energy networks in complex ecosystem.

The Morocco experienced a remarkable evolution in exploitation of renewable energy (Sun, Beach, Wind). Indeed, Morocco is implementing large scale solar energy project through the Moroccan Agency for Solar Energy (MASE) and the Industrial consortium and many separate project for wing energy. Implementing such project in Morocco will benefit its initiatives and will enable the development of engineers who will be able to address these challenging issues with global perspective. In many prospects to exploit natural resources for production of energy and a better management of distribution and consumption, Morocco will become a leader energy producers. Therefore, Smart Grid technology and applications, becomes an absolute necessity for success in this challenge.

3 Sensors Control and Prevention System of Energy Production and Consumption Figures

Energy demand in Morocco is expected to increase steadily between 5–7% per year until 2020. Is highly dependent on energy imports, Morocco will face considerable energy costs and the growth of electricity consumption. Thus, emphasis should be placed on the transmission and distribution which refers to the process of delivering electric energy

from the high voltage grid to consumers and it includes electric lines and transformers (substations) that take power from the high voltage grid and progressively step down the voltage, As well as line management systems that improve efficiency, such as Smart MicroGrid. These are responsible for delivering electric power—no matter the generation source, be it solar, gas, oil, wind or otherwise—using digital technology that allows for a two-way communication between the utility and its customers [7, 8].

As previously described, the distributed database required to operate the multi-agent system, on which is entirely based the design of our project for the management of energy flow between consumers and the distribution network at which they are connected requires a continuous and real-time data on consumption and production at each network node represented by one or several consumers whose behavior is identical energy needs, and delivery points that represent the primary electric energy sources.

In the case of a Smart Micro Grid, it's of course about a specific type of grids whose main characteristics are the voltage level: 20 kV and 22 kV for medium-voltage MV, 380 V for low-voltage network LV, typology and network structure. We distinguish then loop networks or antenna ones, aerial or underground, and other mixed cases. Nowadays, this type of grid is the necessary distribution network to ensure the power supply of a neighborhood or even of a big city or a province.

To transmit electric energy efficiently, medium voltages are used. It is then converted to low-voltage electricity (MV/LV) through substations (transformers), that are characterized by its nominal power expressed in kVA and the number of departures LV. Consumers are connected either directly to the MV grid for power demand greater than 80 KVA or on the BT network for a power demand less than or equal to 80 KVA, according to regulations imposed by the electric energy distributors (ONE is The only company responsible and working as transmission system operator TSO).

The electric energy consumed on LV grids, often by households is accounted for using electricity meter. An electricity meter is thus a device that measures the amount of electric energy consumed. it is installed at the consumer and it is made on the basis of electromechanical technology (counter disc). Electricity meters are typically calibrated in billing units, the most common one being the kilowatt hour [kWh]. Furthermore the production costs can be compared with end consumer prices in the market which are fixed by decree of the Moroccan Prime Minister. Periodic readings of electric meters establish billing cycles and energy used during a cycle (The price per KWh is more expensive when going from one slice to another lower top). But in rural areas, an electronic prepaid card is inserted in the PERG for consumers who are geographically dispersed where the record of index counters presents a very expensive cost for the distributor. The consumer then buys its energy even before consumption. Here, a single tariff is applied.

For customers connected to the MV network, their consumption is recorded using a Smart meters which provided much more information of the electric energy consumption. Smart meters are the next generation of electricity meters and the difference compared to the old meters is that they are able to transmit and receive data. Smart metering is one way to help customers understand their electricity consumption and help them to save energy.

Through a quick feedback and monthly bills, with statistics over the electricity used, the customers will get a better understanding of their electricity consumption. Customers should be able to analyze and optimize the electric energy consumed per time: peak hour, peak hour and peak hour. It also gives the maximum power, the average power factor and minimum, the cumulative active energy in kWh import/export, the accumulated reactive energy in kvarh import/export. These Smart meters usually have a network interface which allows for the remote index reading. This option is not yet operational due to insufficiency of telecommunications infrastructure in distributors. The supplier or distribution system operators will profit from smart metering since they do not need to dispose so much expensive peak power. Through load variable tariff customers can profit as well, since they can optimize their electricity consumption against the given prices. costs are not the same for all customers MT. This depends on the nature of the activity performed by the consumer, whether a small farmer or industry [10–12].

From the above, the smart Micro Grid is identified as two interlocking grids: LV network where consumption is linked to household habits that represent the majority of consumers connected to the grid, and a major source of MV network supply the LV network, where consumers directly connected network, drivers often have a Generating Station with high powers.

The necessary data acquisition for the operation of **la**our automated system for managing production and demand will be carried out in three steps to a better optimization of investment required for the installation of smart meters that communicates the real-time information through communication and data transmission infrastructure the control center hosting the application object of our project:

- In terms of MV/LV stations (customer or distributor) whose number is less important than LV customers. The distribution Stations MV/LV for to supply LV grid that is mainly supplied for households. Because they have a uniform behavior towards energy needs, energy absorbed by the electric MV/LV distribution, reflects clearly this behavior. For MV/LV customer, the consumption depends on the nature of the economic activity. Several scientific models are exploited for the prediction of consumption.
- 2. In terms of LV customers who have their own means of energy production (solar, wind, etc.), a way to manage the flow of energy flowing of this node.
- 3. For each new subscription and renewal.
- Generalization following a timetable which takes into accounts the available financial resources.

This intelligence, desired at the level of such meter has processing capabilities and control to be defined on the basis on discussing focusing on the introduction of the pacer distributor in consumer privacy to influence lifestyle which has a direct impact on energy consumption and maximum power consumption. Knowing that today most retailers are in the process of finalizing their regional centers of remote telecontrol that ensures continuous monitoring of the distribution and control of bodies cuts (breaker, switch) to ensure a better quality of service by:

- A development charge between different arteries forming the network to avoid overload drivers and reduce the voltage drop in the end.
- A quick recharge customer following incidents insulation faulty line section.

The system to perform must then be grafted onto the existing platform in order to exploit first the infrastructure already in place, namely the breaking components remotely controlled and communication means, and secondly the Human resources qualified for the management and distribution networks that can provide their expertise in the field [13].

4 General Applicative Context and Proposed Solution

A smart MicroGrid refers to a distribution network for electrical energy, starting from electricity generation to its transmission and storage with the ability to respond to dynamic changes in energy supply through co-generation and demand adjustments. At the scale of a small town, a MicroGrid is connected to the wide-area electrical grid that may be used for 'baseline' energy supply; or in the extreme case only as a storage system in a completely self-sufficient MicroGrid. Distributed generation, storage and intelligence are key components of a smart MicroGrid [14, 15].

A typical scenario to consider is a customer implementing a renewable energy generator system that will be used to deliver electricity to its needs. Such a customer can draw more energy from the grid, store its energy, or send energy back to the smart grid. In other words, customers can buy, store, or sell energy. To make this scenario happen, the customer's system needs to communicate with the grid (by interacting through the meters) in order to get all necessary control information. The research question that we plan to answer in this proposal is how to design a wireless sensor network that will enable efficient and cost effective network communication between the renewable energy system and the meter. What is the efficiency software that response to requirement of the system? What is the technology that most suitable for such project. This project is shared into three parts:

To achieve our goal, first we need to develop a simulation platform for energy resources management of a MicroGrid Network to optimize and manage the distribution on demand. With the existence of several energetic resources and local production site by consumers, the system will automatically establish an efficient electricity distribution. In this project, we propose the use of the multi-agent technology. This choice is motivated by the functional ability of agents, and their self-adaptation to the environment (i.e. change the feature), and their ability to communicate with other agents and intersite mobility will define and specify the real-time needs of each site according to its production capacity and consumption and that of its neighbors. The architecture of the operating system is based on a graph, where each node represents a site energy, which can be a customer support or a producer of energy (Fig. 1) [15–17].



Fig. 1. Smart grid features

In the second part, and to facilitate communication inter-agents and to optimize the time of data transfer between agents from different sites, and the ability to store the data on other sites, gives more flexibility in terms of Security, availability and overload the network. In this perspective, the use of the principle of Distributed databases in the context of this project report a very suitable solution for our situation. Distributed databases offer many benefits: reduced costs, increased flexibility, and simplified access "on demand" to data with greater agility. Indeed, Distributed databases offers space and computing power to store the several centers, but also the possibility of analyzing the data, processing them and distribute databases effectively split the computational capabilities available, enabling more people easier access to more and more data. On the other hand, allows the prediction of risk in an electrical network model for information. Thanks to the distributed databases solution, we can easily migrate critical data on a data center and improve the response time of readjustment and equilibration of the distribution and consumption [18].

In the last part, this project aims at controlling and adjusting automatically the consumption following the need of each customer at real time using remote control systems. The project starts by defining all parameters influence on the energy consumption for all components one related to others. Each parameter will be represented by a coordinator sensor to give local consummation information. The calculation of real-time needs of delivery or energy demand of each network element, based on the ratio of balancing consumption/production on the entire network [19]. To achieve this scientific need, we are planning to use data mining technology. This is, a procedure to follow to use the data, whatever their forms, in order to extract knowledge. There are the following steps:

- Access and preparation of data, for processing at each site, stored in a structured (database, tabular files) or unstructured (text, image, etc.);
- Using data mining techniques derived from statistical or machine learning;
- Evaluate and validate the extracted knowledge on the report of balancing consumption/production.
- Deployment of knowledge to use and effective decision-making readjustment of the distribution.

5 General Overview of Related Research

This section presents the details of our research axis. We provide an overview of WSN, SMA and Data processing. We highlight what makes them different from traditional data integration methods and why they have enabled the solution of previously unsolved problems. We discuss four research problems that help solving data integration for smart grid system. We list problems in what we consider an increasing level of difficulty. For each research problem, we give a formal description, the main research issues we have identified as of now and a proposed research plan. After research issues and solutions are presented, we explain how the proposed research will be developed for correctness, how architecture will be evaluated for accuracy/validity and how our method will be evaluated for efficiency. This section explains the application of WSN, MAS, Data integration and Data processing on the Smart Grid, where data sets are large, high dimensional or have rich information content (numbers, strings, data).

5.1 Wireless Sensors Network

First, we will adequate the information gathered by the transducers of the physical sensors to a useful format for the developed device. Generally, wireless devices have quite high average power consumption. Thus, in our development we will take into account several issues to save energy and reduce the power consumption as much as possible. The first issue taken into account is to disable the unused parts of the device while they are not being used, which reduces the power consumption significantly. The other issue is the protocol used for communication. We have to design an energy efficient protocol in order to save energy when the information must be transmitted between the electronic low power device and the tablet PC or the mobile phone. Finally, we will design the appropriate algorithm for the behavior of the nodes, that is, the algorithm will decide when de node will be in sleep mode, where there is very low power consumption, and when the node must be active in order to transmit or to receive the information. Moreover, we will add fault tolerance and security to the network. It will allow us to optimize the implementation of a modular sensor node that will be used for different type of applications, with high bitrates. After the device deployment we will deploy new Linux-based operative system that allows gathering data from the physical sensor interfaces and forward them to the Bluetooth interface to be received by the tablet PC and the mobile phone.

5.2 Multi Agents System and Simulation

Our project concerns the design and development of a system for Intelligent Decision Support to discover the need for balancing the energy ratio of the two factors consumption and production. And to make decisions looking distribution of energy based on the cooperation of a sensor network simulated by autonomous agents. It involves developing a new system with intelligent autonomous components to exploit the capabilities (or services) of sensors. In general, the expected features are, among others,

- Permit an entity to describe his skills.
- Automatic detection and classification, possibly with multiple new skills to exploit them.
- Authority to give optimal decision and forecasts for the pretreatment of critical situations.
- Treatment transparent and cooperative delegated services.
- Implementation and achievement of the testes in different application domains.

To do this, the system can cooperate within the framework of a federation, and provide appropriate responses. For example, before or during the search of suitable energy sources to meet a critical need, the other remote systems with available resources and services can be combined with those of the local system to optimize the distribution in a manner transparent. This feature relies on the classification, localization and deployment of powers established by several heterogeneous distributed systems.

5.3 Data Integration

As like as Smart Grid environment, in the distributed environment where a query involves across several heterogeneous sources, communication cost must be taken into consideration. In his paper we describe two query optimization approaches using dynamic programming technique for a given set of integrated heterogeneous sources. The primary objective of the optimization is to minimize the total processing time including load processing, request rewriting and communication costs, to facilitate communication inter-sites and to optimize the time of data transfer from different sites. Moreover, the ability to store the data on center site gives more flexibility in terms of Security/Safety and overloading the network. In contrast to optimizers which consider a restricted search space, the proposed optimizer searches the subsets of sources and independency relationship which may be deep laniary or bushy trees. Especially the execution de query can be started traversal anywhere over any subset and not only from a specific one.

The main problem is to maintain a distributed data warehouse, consisting of multiple local data warehouses (sites) adjacent to the collection points, together with a coordinator. This coordinator uses ontologies to explicit the semantic of sources. The heterogeneity is caused by the diversity of smart grids that may have various constructors, models, technologies, etc. Once the distributed data warehouse constructed, we need to classify the data to extract the profiles of consumers. The development of adapted algorithms represents a crucial issue that has to be described. The basic idea of such algorithms is to translates a set of sources into distributed distinct subsets and generates distributed warehouses, with the following concept: (i) each generated data warehouse performing some computation and communicating the query result to the coordinator, and (ii) the coordinator synchronizing the results and (possibly) communicating with the warehouses. The semantics of the sub queries generated by system ensure that the amount of data that has to be shipped between warehouses is independent of the size of the underlying data at the sites.

The solution allows for a wide variety of optimizations that are easily expressed in the interrogation and thus readily integrated into the query optimizer. The optimization algorithm included in our prototype contributes both to the minimization of synchronization traffic and the optimization of the data processing at the local sites. Significant features of this approach are the ability to perform both distribution-dependent and distribution in dependent optimizations that reduce the data transferred and the number of evaluation rounds.

5.4 Data Integration and Data Processing

The basic idea of this algorithm is: data in the network is transmitted as the entire relationship or a fragment from source to others, which is obviously a redundant way. When a relationship transferred to another venue, not every data is involved in connection operation or useful. Therefore, the data is not involved in the connection or useless data needs not to be transmitted circularly in the network. The basic principle of this optimization strategy is to use semi-connection operation to only transmit the data involved in the connection in the network as far as possible.

In the perspectives of this project we address the following research areas:

- Hybrid integration of data sources:
- · Recommendation of energy resources based on consumer profiles
- Inter-agent negotiation
- Invoice generation based on the usage of energy

The challenge is to integrate the large amount of data from customer demand with the data on power grid performance. The data from customers is noisy and rich in information. Data-analysis applications have to filter the data and find trends in close to real time. Utilities plan to use such trend information to make decisions on power grid operation. The goal will be to have software operate the power grid in accordance with such trends.

While the needs of large utilities are driving development, second tier power grid operators also have to implement smart grid technologies. The solutions they choose are of particular interest to midsize businesses because these smaller utilities are often of comparable size. Such solutions may offer these businesses expanded data processing and analyzing at reasonable cost. Access to large amounts of data: Customer files, market analyses, and internal sources could yield more useful information with expanded and more powerful analytics. As smart grid development pushes the upper limits of data management, the middle capabilities expand as well. That could give midsize businesses cost-effective access to the technology they need to squeeze more information out of their data.

6 Functional Architecture

In this part and first, we start with the description of the functional architecture of the system. This architecture is based on the concept of directed graph defined as follows: Each node represents a geographical site brings together a set of resources in the production of energy and consumer group. The arcs among nodes (sites) represents the amount of energy lost during transfer from one site to another. In his considered at the same site there is no transfer of energy between production stations.

For the organization of the functional architecture we define the two possibilities of the following links:

Site-to-Site. This link means that energy production plant can power another via a temporary transfer (Fig. 2).



Fig. 2. Site-to-Site

Site-to-Consumer. This link means that a customer uses energy generated by a production site (Fig. 3).



Fig. 3. Site-to-Consumer

To simplify the system we define the various components, into five categories:

- 1. Base stations of energy production that can generate the energy: Distribution substations, wind turbines or/and solar panels.
- 2. Local generation/consumption of energy, which can be: House, Laboratory, Research Centre, etc.
- 3. Mobile Generation/consumption, which can be: vehicle, mobile emergency units.
- 4. The static Consumer of the energy.
- 5. The mobile Consumer of the energy.

Since the amount produced and consumed by each site depends on producers and consumers in a given site and exchanges with other sites, and with the consideration of mobile generators and consumers we must highlights the reorganization phase sites in dynamically moving a consumer or a generator from one site to another.
To ensure the balancing management and distribution of energy between the sites, each site is associated a database for storage of the quantities produced by the generator and the quantities consumed by the customers of this site. Data from this these databases will enable us to make decisions on prevention and global production and consumption in the networks to effectively manage energy transfers between sites (Fig. 4)



Fig. 4. Proposed Solution based on collection of local data in data centre where the coordinator communicates with others via distributed agents

Thereafter we will define the data model to use to perform the algorithms presented above. The system we propose is divided into two Levels: *inter-Nœud* and *intra-Noeuds*.

6.1 Inter-node Architecture

Each node Ni is characterized by:

- Production capacity of each energy resource at a given Sj of this node: Prod(Sj,t).
- Actual energy consumption of a client at a given time: $ConsReel_i^l(C_i)$.
- The amount of energy transmitted by a source Sk to a customer of a node Ni at time

t: $Quantite_i^t(S_k, C_j)$.

- Geo- localization Lj(x,y).
- The nature of the Nat sources Nat(Pi).

Table 1.

S 1	Prod(Sj,t)		C1	$ConsReel_i^t (C_j)$
S2	Prod(Sj.t)	\geq		$ConsReel_{i}^{t}\left(C_{j} ight)$
S3	Protest,t	\sim	C3	$ConsReel_{i}^{t}\left(C_{j} ight)$
		\searrow		
		\sim	*	
Sn	Prod(Sj,t)		C m	$ConsReel_{i}^{t}\left(C_{j}\right)$

We define the following variables:

Whose total amount transmitted to customers Cj at time t represented by a node Ni is:

$$TotaQuantity_i^t(C_j) = \sum_k Quantity_i^t(S_k, C_j)$$
(1)

- The total amount of production of a site represented by a node Ni:

$$Capacity(N_i) = \sum_{j} Prod_i^t(S_j)$$
⁽²⁾

- Total consumption of a site represented by a node Ni:

$$ConsTot(N_i) = \sum_{j} Cons_i^t(C_j)$$
(3)

- The quantity of energy lost at time t by a customer Cj is:

$$LostQuantity_i^t(C_j) = TotalQuantity_i^t(C_j) - ConsReel_i^t(C_j)$$
(4)

with the condition:

$$0 < LostQuantity_i^t(C_j) < \beta$$
⁽⁵⁾

with $\boldsymbol{\beta}$ is the tolerance allowed for the loss of energy when demand for end customers satisfaction.

6.2 Intra-node Architecture

In this section we present the graph representing the electricity network in which each node Ni (Pi, Ci, Ri) is a region that includes the energy Pi production resources, consumer list that use the resources of this area Ci lists and constraints to meet Ri (Fig. 5).



Fig. 5. Inter-node architecture

Whether:

$$Arc(i, j) = transmittedQuantity(N_i) - ReceivedQantity(N_i)$$
(6)

With:

transmittedQuantity (N_i, N_j) represents the amount of energy transmitted from the site Si to the site Sj and *ReceivedQantity* (N_i, N_j) represents the quantity received by the site Sj transmitted by Si.

We must therefore ensure balance between production and consumption and energy transfer between the site S_i and S_i as follows

Balancing (Si) = Prodi
$$(S_i, t)$$
 + TotaQuantity(Si)
- \sum transmittedQuantity(Si, Sj) (7)

We seek to develop an algorithm that minimizes the values Arct(i,j) and maximize Balancing (Si) for all i and j. This means to minimize the energy lost during transfer or balancing process among the sites represented by the graph and takes a right distribution of energy between the sites.

7 Sources Classification and Constraints Satisfaction

Thereafter we propose a solution for the distribution and classification of sources and end customers at all sites (Micro-Grid). That means, the construction of nodes of the graphs to consolidate energy resources and consumers of the same category according to two criteria: Geographic and satisfaction rate. Then developing a distribution algorithm and balancing of energy between the sites represented by the graph shown in the previous section.

Algorithm.

Energy sources classification algorithm and consumers is mainly based on consumers who share the same energy resources with the objective of minimizing the energy transfer between the sites (inter-site).

For the definition of all sites (Micro-Grid) we follow the following steps:

1. Selection of centers for gathering geographically close energy resources or when the distance less than Dis (Si) for each center Ci. The choice of the threshold depends on the region.

$$Consomateurs(S_i) = \{C_j / Quantite_i^t(S_i, C_j) \neq 0\}$$

2. Appropriation of the consumers fed by the same source of energy resources even micro-Grid.

$$Consomateur(S_i) = \{C_j / Quantite_i^t(S_i, C_j) \neq 0\}$$

- For consumers fed by one or more Micro Grid, they are affected each site whose difference between the total consumption and the total cross-site production and the maximum possible.
- 4. Consomateur $(S_i) = \{C_i / Quantite_i^t(S_i, C_i) \neq 0\}$

8 Conclusion

In this paper we proposed a solution to ensure the balancing of distribution and consumption of energy between micro-grid groups. This solution will minimize the loss of energy during the transfer process and to consumption of energy. The solution is based on the real-time study of consumer needs and distributed generation on all energy site.

References

- 1. Pepermans, G., Driesen, J., Haeseldonckx, D., et al.: Distributed generation: definition, benefits and issues. Energy Policy **33**, 787–798 (2005)
- 2. Molderink, A., Bakker, V., Bosman, M., et al.: Management and control of domestic smart grid technology. IEEE Trans. Smart Grid **1**(2), 109–119 (2010)
- Carrasco, J., Franquelo, L., Bialasiewicz, J., et al.: Power electronic systems for the grid integration of renewable energy sources: a survey. IEEE Trans. Ind. Electron. 53(4), 1002– 1016 (2006)
- Justo, J.J., Mwasilu, F., Lee, J., Jung, J.-W.: AC-microgrids versus DC-microgrids with distributed energy resources: a review. Renew. Sustain. Energy Rev. 24, 387–405 (2013)
- Guerrero, J., Chandorkar, M., Lee, T., Loh, P.: Advanced control architectures for intelligent microgrids-part I: decentralized and hierarchical control. IEEE Trans. Ind. Electron. 60(4), 1254–1262 (2013)
- Guerrero, J., Loh, P.C., Lee, T.-L., Chandorkar, M.: Advanced control architectures for intelligent microgrids – part II: power quality, energy storage, and AC/DC microgrids. IEEE Trans. Ind. Electron. 60(4), 1263–1270 (2012)
- 7. SolarRay, Grid-tie package systems without batteries. http://www.solarray.com/ CompletePackages/Grid-Tie-No-Batteries T.php, posted on (2012)
- Westermann, D., Kratz, M.: A real-time development platform for the next generation of power system control functions. IEEE Trans. Ind. Electron. 57(4), 1159–1166 (2010)
- Ekneligoda, N.C., Weaver, W.W.: A game theoretic bus selection method for loads in multibus DC power systems. IEEE Trans. Ind. Electron. 61(4), 1669–1678 (2014)
- 10. Başar, T., Olsder, G.L.: Dynamic Noncooperative Game Theory. Series in Classics in Applied Mathematics. SIAM, Philadelphia (1999)
- Vásquez, J., Guerrero, J., Miret, J., Castilla, M., de Vicuña, L.: Hierarchical control of intelligent microgrids. IEEE Ind. Electron. Mag. 4(4), 23–29 (2010)
- Guerrero, J., Vasquez, J., Matas, J., de Vicuña, L., Castilla, M.: Hierarchical control of droopcontrolled AC and DC microgrids – a general approach toward standardization. IEEE Trans. Ind. Electron. 58(1), 158–172 (2011)
- Guerrero, J., Vásquez, J., Matas, J., Castilla, M., de Vicuña, L.: Control strategy for flexible microgrid based on parallel line-interactive ups systems. IEEE Trans. Ind. Electron. 56(3), 726–736 (2009)
- Hill, C., Such, M., Chen, D., Gonzalez, J., Grady, W.: Battery energy storage for enabling integration of distributed solar power generation. IEEE Trans. Smart Grid 3(2), 850–857 (2012)
- Liu, Y., Yuen, C., Huang, S., Hassan, N.U., Wang, X., Xie, S.: Peakto-average ratio constrained demand-side management with consumer's preference in residential smart grid. IEEE J. Sel. Topics Signal Process. PP(99), 1–14 (2014)
- Hassan, N.U., Pasha, M.A., Yuen, C., Huang, S., Wang, X.: Impact of scheduling flexibility on demand profile flatness and user inconvenience in residential smart grid system. Energies 6(12), 6608–6635 (2013)
- Balaguer, I., Lei, Q., Yang, S., Supatti, U., Peng, F.Z.: Control for grid-connected and intentional islanding operations of distributed power generation. IEEE Trans. Ind. Electron. 58(1), 147–157 (2011)

- Liu, Y., Hassan, N., Huang, S., Yuen, C.: Electricity cost minimization for a residential smart grid with distributed generation and bidirectional power transactions. In: IEEE PES Innovative Smart Grid Technologies (ISGT), Washington, DC, pp. 1–6, February 2013
- 19. Zhang, D., Shah, N., Papageorgiou, L.G.: Efficient energy consumption and operation management in a smart building with microgrid. Energy Convers. Manag. 74, 209–222 (2013)

Using Image Segmentation in Content Based Image Retrieval Method

Mohamed Ouhda^(⊠), Khalid El Asnaoui, Mohammed Ouanan, and Brahim Aksasse

Faculty of Sciences and Techniques, Department of Computer Science, ASIA Team M2I Laboratory, Moulay Ismail University, BP 509 Boutalamine, 52000 Errachidia, Morocco ouhda.med@gmail.com, khalid.elasnaoui@gmail.com, ouanan_mohammed@yahoo.fr, baksasse@yahoo.com

Abstract. Today's world is digital with the appearance of many devices that are used in image acquisition. Nowadays, it becomes easy to store huge amount of images by using image processing techniques. The rapid access to these masses collections of images and retrieve similar images of a given image (Query) from this huge collection of images presents major challenges and requires efficient algorithms. The main goal of the proposed system is to provide an accurate result with lower computational time. For our purpose, we introduce in the content based image retrieval (CBIR) system the classification step, and we apply k-means clustering technique to match image's descriptors. This work provides a detailed view of the solution we have adopted, and that perfectly meets our needs. For validation, we apply all of these techniques on two image databases in order to evaluate the performance of our system.

Keywords: K-means \cdot Segmentation \cdot Indexing \cdot Similarity measure \cdot CBIR \cdot Classification

1 Introduction

An image is worth more than a lot of words, an image can indeed describe a sunset, the smile of an Olympic winner, a family meal... much better than many words. For this reason, the media encyclopedias or private individuals use images to describe events, people, and objects. For users, images have often more meaning than long sentences. However, due to the development of many devices (digital cameras, scanners...), the increase in network transfer performance and storage systems, there has been a real explosion in the number of images in recent years which a user wishes to access. Therefore, users need tools in order to retrieve quickly and efficiently the large mass of available information. These tools are regrouped in the information retrieval system (IRS). Nowadays, content-based image retrieval (CBIR) systems have made great strides in the retrieval of specific objects in images. However, current image retrieval systems are still poorly performing in the semantic search of images in general databases. One reason of this is the way in which images are described in computer systems. The content of an image can be described in two different levels: At the digital

© Springer International Publishing AG 2018 M. Ezziyyani et al. (eds.), *Advanced Information Technology, Services and Systems*, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_17 level, an image contains colored pixels from which color descriptors, textures, and shapes can be extracted. At the semantic level, an image can be interpreted and can have at least one meaning. Unfortunately, in today's information systems, images are described digitally while users are interested in their semantic content and it is currently difficult to find correspondences between the digital and semantic level. This is what is called semantic deviation.

Recently, the research focuses on CBIR systems that are fetching the exact cluster of relevant images and reducing the elapsed time of the system. For this purpose, various techniques have been developed to improve the performance of CBIR system. Clustering is one of them. Moreover, clustering is the technique that used to partition the data into groups of similar objects. Note that CBIR systems use usually three low level primitive to extract inherent information based on color, texture and shape descriptor [1, 2].

The intent of the classification process is to categorize all objects in a database color image into one of the several classes, or "clusters". This categorized data is based on their similar features present in a database image. Image classification is a labeling process, in which image pixels are categorized into different classes in case of a single image, but in the case of database image instead of pixels, we take an image of an object and categorized into different classes. Note that image classification algorithms have been successively applied to a range of problems, including image segmentation and color quantization, change detection for land cover monitoring, and data mining among others [3, 5, 6].

The proposed system uses classification based on k-means clustering algorithms which are perfectly suited to the main difficulty residing in the optimal features selection that is able to produce clusters with spatial homogeneity.

Our main goal in this work is to set up an algorithm that can retrieve similar images of a given image (query). Toward this end, we have developed and tested our algorithm for image retrieval using three image databases. We used Euclidean distance to match images and to measure accurate similarity for image retrieval. We have successfully applied our proposed method to image retrieval using only visual image content.

The rest of the paper is organized as follows: Sect. 2 deals with related works. In Sect. 3, we develop the mean steps of the proposed method. Section 4 presents the experimental result. Finally, Sect. 5 concludes the paper.

2 Related Works

In this section, we will briefly outline some important contributions from the existing literature.

The method reported by [1] showed that the color histogram of the image and its bin values are analyzed to understand and extract the color information in the image. The histogram dimension is reduced by deleting trivial bins and only those bins that represent color data significantly are considered. Based on the dimensions of the histogram, it is clustered and indexed.

The descriptor presented by [2] is based on 2-D histogram method with statistical moments by integrating the texture using Gabor filters and by applying distributed

computation to retrieve an image. The most problem using the histogram is that three different images can have the same histogram. That is why [3, 4] combined the histogram with k-means to increase the image relevance of the image. Indeed [3] suggested a refinement method further refines the histogram by splitting the pixels in a given bucket into several classes based on color coherence vectors. Various features are calculated for each of the clusters and these features are further classified using the k-means. While [4] compared two main techniques frequently used in CBIR which are: The normal color histogram using the gray level co-occurrence matrix (GLCM) and the color histogram using the k-means method.

In order to reduce the response time, [5–7] classified images in the images database after clustering by k-means. In fact, [5] combined two algorithm of regrouping with the integration of the algorithm k-means which are useful for the classification of the images as well as the extraction of the objects. The approach reported by [6] is Modified K-Nearest Neighbor (MKNN) that can be considered a kind of weighted KNN (K-Nearest Neighbor) in order that the query label is approximated by weighting the neighbors of the query. The procedure computes the fraction of the same labeled neighbors to the total number of neighbors. MKNN classification is based on validated neighbors that have more information in comparison with simple class labels. In [7], the authors introduced a hybrid solution for content-based image retrieval system with a combination of k-means and hierarchical image clustering algorithms. Their main goal is to make image retrieval faster as possible.

Moreover, the approaches [8–10] presented algorithms for increasing the efficiency of clustering k-means: in [8], a new partitioned clustering algorithm based on the notion of a contribution of a data point is given. They applied the algorithm to content-based image retrieval and they compared its performance with the k-means clustering algorithm. Unlike the k-means algorithm, the algorithm optimizes on both intra-cluster and inter-cluster similarity measures. It has three passes and each pass has the same complexity as iteration in the k-means algorithm. [9] proposed a k-means clustering algorithm aiming to develop clusters from each image database records; it can later be used for optimizing image searching access period. The stored images in the image database are only limited for the JPEG-type images. In this algorithm, cluster construction is based on maximum and minimum PSRN's (Peak Signal to Noise Ratio) calculation values from individual records on basic images and it will be treated as key images in every search for records with such cluster utilization.

In addition, [10] has given a method that consisted for image clustering based on combining the particle swarm optimization (PSO) with k-means clustering algorithms. It is presented as CBIR method that uses the color and texture images as visual features to represent the images.

3 The Proposed Method

According to several techniques developed in recent years, each one has many disadvantages. We present in this work a new descriptor more relevant in terms of accuracy and robust to geometric transformations (translation and rotation). Therefore, to minimize the response time, we have introduced a classification step in our system which classifies each image into a class, the launch of a search query is done in the class closest to the image and not in the entire database. This technique makes our system more rapid and more relevant.

Moreover, a very important step is the automatic verification of the results which makes it possible to correct the search if the cluster does not appropriate to the query image.

We can generally present each image by three characteristics: color, shape, and texture [2, 15]. For a fast and an improved image extraction performance, we use the extraction of the color characteristic that is an essential image feature and often used in Content-Based Image Retrieval systems. The proposed system CBIR is shown in the Fig. 1, images are stored in a database called image database, After the pre-processing, the images are segmented using the k-means image segmentation method. We use the color space RGB. The result of segmentation is NR regions. For each region and for each component, the mean and the standard deviation are given by:



Fig. 1. Architecture of CBIR system used.

$$\mu_i = \left(\mu_R^{r_i}, \mu_G^{r_i}, \mu_B^{r_i}\right) \tag{1}$$

$$\sigma_i = (\sigma_R^{ri}, \sigma_G^{ri}, \sigma_B^{ri}) \tag{2}$$

where:

$$\mu_R^{ri} = \frac{1}{RP} \sum_k^{RP} M_R(i,j) \tag{3}$$

And

$$\sigma_R^{ri} = \sqrt{\frac{1}{Rp} \sum_{k}^{RP} \left(M_R(i,j) - \mu_R^{ri} \right)^2} \tag{4}$$

and *RP* is the number region pixels r_i of the red component.

Then, a unique feature vector is constructed and saved in the feature database after the step of classification. This vector is illustrated as:

$$V = (\mu_1, \sigma_1, \mu_2, \sigma_2, \dots, \mu_i, \sigma_i, \dots, \mu_{NR}, \sigma_{NR})$$
(5)

Now, when a query image is submitted by the user, the same procedure is applied as explained above to obtain its feature vector.

- 1. Offline phase: (Classification)
- Segmenting the image into *NR* regions (k-means):
- Building the descriptor vector V using Eq. (5)
- Retrieve the nearest class.
- Add the image to the nearest class.
- Update the descriptor vector of the class by the following formula:

$$Vc = \frac{\sum_{i=1}^{N} V_i}{N}$$
(6)

where V_i is the descriptor vector of the *i* image and *N* is the number of images in the class.

- 2. Online phase:
- Compute the vector descriptor of the query image V_O
- Determinate the appropriate class by comparing V_Q with all $\{V_C\}$

$$d(V_{\mathcal{Q}}, V_{\mathcal{C}}) = \max\left(\left|\mu_{\mathcal{Q}_i} - \mu_{\mathcal{C}_i}\right|\right) + \max\left(\left|\sigma_{\mathcal{Q}_i} - \sigma_{\mathcal{C}_i}\right|\right) \tag{7}$$

- Sort the classes according the nearest to farthest.
- Compute the distances between features descriptor by the formula:

$$D(V_{Q}, V_{M}) = \sqrt{\sum_{i}^{NR} \left(\mu_{Q_{i}} - \mu_{M_{i}}\right)^{2} + (\sigma_{Q_{i}} - \sigma_{M_{i}})^{2}}$$
(8)

where Q is the query image, NR number of region, M an image in database, V_Q descriptor of Q, V_M descriptor of M and $D(V_Q, V_M)$ is the distance between the images Q and M

- Retrieved similar images in the appropriate class.
- Determinate the class for each retrieved image.
- If the class of the query image is at least equal to the first 13 images retrieved, then the images retrieved are similar, otherwise, we repeat the same process in the second nearest class, until we find the appropriate class.

3.1 K-Means Segmentation

Image segmentation is essential for CBIR because it extracts necessary and significant features from images. In a regular CBIR system, the retrieval performance is related to the efficient image segmentation result. In general, features are extracted from the entire image like histogram method, which means that trivial background information can bias the feature and adversely affect the retrieval performance. To deal with database image segmentation, the CBIR system needs of regular and efficient pre-processing and segmentation algorithms. The choice of descriptors for image search system content is most important, in the sense that this choice affects the expected results. K-means clustering is a very effective method to extract the vector descriptor. This algorithm splits the given image into different clusters.

• Adaptative k-means

Since the random selection of the initial cluster centers from image data is not an appropriate task, we use the peaks of histograms. In this regard, we have computed the histogram for each color component. This solution can be described by the following algorithm:

- a. Compute the histograms for each color component
- b. Split each histogram into *R* sections
- c. Compute the peaks in each section and sort the peaks, $P_1, P_2, ..., P_k$ where $P_i, i \in [1, k]$ has the highest number of elements
- d. Start to construct the color seeds for highest peak P_i

if $(P_i \rightarrow red)$ mark the pixels in the red component and calculate the g_{mean} and b_{mean} for marked pixels from green and blue components

if $(P_i \rightarrow green)$ mark the pixels in the green channel and calculate the r_{mean} and b_{mean} for marked pixels from red and blue components

if $(P_i \rightarrow blue)$ mark the pixels in the blue channel and calculate the r_{mean} and g_{mean} for marked pixels from red and green components

- e. construct the color seed and eliminate P_i from the list
- f. Repeat the steps d and e until the desired number of color seeds has been reached

Figure 2 below shows a simple example of this algorithm. This example segments an image using a color space cluster with three components (red, green, blue). The input image contains a number of different region colors. When the adapted k-means





b) Food image segmented into 4 region.

Fig. 2. Examples of the image segmentation

algorithm is applied the objects can be separated. The choice of k is however important, if k is too small, some regions are grouped together.

3.2 Classification

The classification step allows grouping similar images into some class, for each class we compute a descriptor vector that depends much on descriptor vectors of the constituent images in the same classes. Image classification is also an active subdomain in the field of machine learning, in which it uses algorithms that map images of input, to set of labeled classes. These algorithms are called classifiers [5].

The classification of images involves two steps:

- (a) Segmentation step by k-means.
- (b) Classification step.

3.3 Similarity Measure

The distance between the query image and an M image from database is the distance between their descriptors. To compute this similarity distance, we use the following formula (8).

4 Experimental Results

In order to evaluate our proposed approach, the performance of our system is analyzed using the evaluation metrics including precision and recall. Furthermore, three datasets have been used to perform the proposed method.

4.1 Datasets Used

In this section, we examine the performance of the developed algorithm on a three scientific images database COIL-100, Wang, and Airplane.

Columbia Object Image Library (COIL-100) is a database of color images of 100 objects. The objects were placed on a motorized turntable against a black background. The turntable was rotated through 360° to vary object pose with respect to a fixed color camera. Images of the objects were taken at pose intervals of 5° . This corresponds to 72 poses per object (Fig. 3).



Fig. 3. Objects used in COIL-100.

We have also used the Wang database which is broadly used for CBIR field. This is why we implemented our test on it in this study. The database holds 1,000 images divided into 10 classes and in each class, there are 100 images. Images in the database are friendly implement for many assessing the CBIR systems (Fig. 4).



Fig. 4. Images example in wang classes database.

The airplane database contains 1074 images of an airplane in different positions with different sizes. These images have the JPEG format Collected by undergrads at California Institute of Technology from the web (Fig. 5).



Fig. 5. Images example in airplane database.

4.2 Performance Evaluation of CBIR

The most common measures to evaluate a system are response time and memory space used. A better system if the response time and the memory space are smaller this is the case of our system because we retrieve the query image in the appropriate class and not in the entire images database, in addition to these two measures, the user is interested in the relevant responses of the system. So information retrieval systems require the evaluation of the accuracy of the response. This type of evaluation concerns the system research performance. The indexation system and Image searching is an information retrieval system.

In this section, we will describe the two most important measures: recall and precision. These measures are interrelated. Therefore this relationship is often described by a recall and precision curve. Then we present other measures that are also used to evaluate information retrieval systems.

The recall is defined as the ratio of the number of relevant images retrieved to the total number of relevant images in the database.

$$recall = \frac{Ra}{R} \tag{9}$$

where Ra is the relevant retrieved images, R is the total number of relevant images in the database.

Precision is the ratio of the number of relevant images retrieved to the total number of images retrieved.

$$precision = \frac{Ra}{A} \tag{10}$$

where A is the total number of images retrieved.

In practice, we use several queries. In these cases, to evaluate a system, we calculate the average precision for all the queries corresponding to each callback level. This value is given by:

$$P(r) = \sum_{i}^{N_q} \frac{P_i(r)}{N_q} \tag{11}$$

Where Nq is the number of queries, and $P_i(r)$ is the precision for recall r with query i.

4.3 Image Retrieval

To check the retrieval performance and robustness of the proposed method, several experiments are conducted on COIL-100 and Wang database. We display the best 19 similar images retrieved to prove the performances of our proposed (Fig. 6).

Application to COIL-100 database.



Fig. 6. Retrieval results on the obj16_0 with the query image in the top-left corner

We have compared our proposed method with other methods of state of the art. The results obtained are given in the below table (Fig. 7).

Method	Average precision	Average recall
	(%)	(%)
GCBA (Kavitha et al. 2014) [11]	90.92	63.14
Transform based Haar (Seema et al. 2014) [12]	68.84	68.84
Cluster based using LBG Algorithm (Seema et al.	85.55	85.85
2014) [13]		
(El Asnaoui et al. 2015) [2]	92.87	18.20
The proposed method	100	19.00

 Table 1. Average precision and recall between the proposed system and some existing systems

• Application to Wang database.



Fig. 7. Retrieval results of a Bus class with the query image in the top-left corner



Fig. 8. Retrieval results of an out query image of database with the query image in the top-left corner



Fig. 9. Retrieval results of Food class with the query image in the top-left corner

We tested the proposed algorithm with several metric distances. From the below table, we found that the Euclidean distance is the suitable one for our system (Table 2).

Category	Distance (Average)				
	Euclidean	Manhattan	Swain	Chebyshev	
Africa	94	71.52	89.0	45.25	
Beaches	87	43.60	73.0	39.75	
Buildings	88,56	53.55	70.01	37.35	
Bus	98	85.30	83.8	74.10	
Dinosaur	100	99.55	99.8	91.45	
Elephant	98	59.10	79.5	30.40	
Flowers	99	90.95	89.75	85.15	
Horse	100	92.40	87.1	56.80	
Mountains	86,5	38.35	55.0	29.25	
Food	90	72.40	75.5	36.95	
Average	94,106	70.67	80,246	52.64	

Table 2. Precision comparison using different similarity distances



Fig. 10. Performance comparison of average Precision-Recall curve.

Based on the recall/precision curves, we find that accuracy decreases when irrelevant images are found. This is due to the fact that we have chosen good descriptors and the images grouping in classes (Fig. 10 and 11).

• Application to airplane database



Fig. 11. Retrieval results of a airplane in the sky with the query image in the top-left corner

192 M. Ouhda et al.



a) Personnale plane query in the airport



b) Airliners query in the airoprt

Fig. 12. Retrieval results of an airplane in the airport with the query image in the top-left corner

Based on the experimental result and on the recall and precision, we notice that the proposed approach can detect the position of the airplane, In addition it is able to differentiate between various types of airplanes: airliners, personal plane... (Fig. 12), due to the right choice of vector descriptor (Table 3).

Position	Precision %	Recall %	
Sky	100	19	
Airport	100	19	
Take off	96	18,24	
Landing	94,4	17,39	

Table 3. The recall and precision of the airplane in different positions

4.4 Response Time

In this section, we compare the response time of a query image. Therefore, the experiments are performed on a computer with processor: Intel (R) Core (TM) 2CPU T5200 @ 1.60 GHz, 1.60 GHz, 2 GB RAM running on the Microsoft Windows 7 Professional operating system (32-bit). For the simulation, MATLAB 2009a is used. The results obtained are given in the table below. The difference in response time compared to other systems will be remarkable when it comes to big database (Table 4).

Table 4. Response time in seconds.

Method	SIFTBoW (Romain	(El Asnaoui	(El Asnaoui	Proposed
	2013) [17]	2014) [18]	2016) [2]	method
Response	28	15	8,75	5,65
time				

5 Discussion

The image segmentation in NR regions depends on the content images in the database because the image can contain different objects. In our experiment, we have adapted NR = 2 for the COIL, airplane databases and NR = 4 for the Wang database.

In the COIL dataset (Table 1), a satisfactory retrieval of the expected images is provided faster. The proposed method could be considered as a solution for the development of visual information retrieval.

The result of this study is compared with the other methods like Jhanwar et al. [14], Rao et al. [15]; Vimina et al. [16], El Asnaoui et al. [2] and Zeyad et al. [10] (Fig. 9) show that the performance and the semantic level of our proposed system is better than other systems for all classes. Moreover, we have taken an image out of the databases to prove the efficiency of our method and finally, our system can successfully retrieve the similar images (Fig. 8).

6 Conclusion

In this paper, we have proposed an improved method of image indexing, by introducing a classification step in the CBIR system. We have tested our system in three image databases. The experimental result makes two important conclusions: low execution time and high accuracy relevance.

The main focus was to show the performance of image retrieval system with the classification. Database of images has been taken for experimentation. All the images are retrieved from the database on the basis of the RGB of the three different components of red, green, blue.

In terms of precision, we have shown that k-means clustering is quite useful for extracting a more relevant and efficient descriptor vector, moreover introducing steps of classification and verification give the efficient results.

References

- Shaila, S.G., Vadivel, A.: Indexing and encoding based image feature representation with bin overlapped similarity measure for CBIR applications. J. Vis. Commun. Image Representation 36, 40–55 (2016)
- El Asnaoui, K., Chawki, Y., Aksasse, B., Ouanan, M.: Efficient use of texture and color features in content based image retrieval (CBIR). Int. J. Appl. Math. Stat.[™] 54(2), 54–65 (2016)
- An, Y., Baek, J., Shin, S., Chang, M., Park, J.: Classification of feature set using k-means clustering from histogram refinement method. In: Fourth International Conference on Networked Computing and Advanced Information Management, NCM 2008, Vol. 2, pp. 320–324. IEEE, September 2008
- Rasli, R.M., Muda, T.Z.T., Yusof, Y., Bakar, J.A.: Comparative analysis of content based image retrieval techniques using color histogram: a case study of GLCM and k-means clustering. In: Third International Conference on Intelligent Systems Modelling and Simulation, pp. 283–286. IEEE, February 2012
- Subbiah, B., Christopher, S.C.: Image classification through integrated K- means algorithm. IJCSI Int. J. Comput. Sci. Issues 9(2), 1–7 (2012)
- 6. Dharani, T., Aroquiaraj, I.L.: Content Based Image Retrieval System Using Feature Classification with Modified KNN Algorithm (2013). arXiv preprint arXiv:1307.4717
- Sharma, S.: Fast retrieval of images using k-means & hierarchical clustering algorithm. VNN J. Comput. Technol. 2(1) (2016)
- Narasimhan, H., Ramraj, P.: Contribution-based clustering algorithm for content-based image retrieval. In: 2010 5th International Conference on Industrial and Information Systems, pp. 442–447. IEEE, July 2010
- Rejito, J., Wardoyo, R., Hartati, S., Harjoko, A.: Optimization CBIR using k-means clustering for image database. Int. J. Comput. Sci. Inf. Technol. 3(4), 4789–4793 (2012)
- Younus, Z.S., Mohamad, D., Saba, T., Alkawaz, M.H., Rehman, A., Al-Rodhaan, M., Al-Dhelaan, A.: Content-based image retrieval using PSO and k-means clustering algorithm. Arab. J. Geosci. 8(8), 6211–6224 (2015)
- 11. Kavitha, H., Sudhamani, M.V.: Content based Image retrieval based on global and region content of an image. Int. J. Adv. Comput. Commun. Syst. **78546**(1), 1 (2014)

- 12. Seema, A.C., Omprakash, Y., Vaishali, S.: Comparative evaluation of transform and cluster based CBIR. Int. J. Comput. Appl. **99**(6), 1–4 (2014)
- 13. Chaurasia, S.A., Suryawanshi, V.: Hybrid algorithm for Image retrieval using LBG and k-means. Database 7, 8 (2014)
- Jhanwar, N., Chaudhuri, S., Seetharaman, G., Zavidovique, B.: Content based image retrieval using motif co-occurrence matrix. Image Vis. Comput. 22(14), 1211–1220 (2004)
- Rao, M., Rao, B.P., Govardhan, A.: Content based image retrieval system based on dominant color and texture features. Int. J. Comput. Appl. 18(6), 40–46 (2011)
- 16. Vimina, E.R., Poulose Jacob, K.: A Sub-block based image retrieval using modified integrated region matching. Int. J. Comput. Sci. Issues 10(1), No. 2 (2013)
- 17. Romain, R., Burie, J.C., Ogier, J.M.: Structured representations in a content based image retrieval context. J. Vis. Commun. Image Representation **24**(8), 1252–1268 (2013)
- El Asnaoui, K., Aksasse, B., Ouanan, M.: Content-based color image retrieval based on the 2-D histogram and statistical moments. In: 2014 Second World Conference on Complex Systems (WCCS), pp. 653–656. IEEE, November 2014

Alignment of IT Frameworks for Corporate Governance

Hajar Ben Laadar^(III), Ilias Cherti, and Mohamed Bahaj

Laboratory of Innovation Technology in New Energy and Nano-materials (LITEN), Faculty of Sciences and Technologies, University of Hassan 1st, Settat, Morocco benlaadar.hajar@gmail.com

Abstract. Enterprises in the 21th Century, with their numerous different activities, have a vision for their brand image to be recognized, for their human resources to be innovative, for their investments to be fruitful, to their Data to be meaningful, to their predictions to be right and to their Decisions to be productive. All those expectations and more; made and set by the Management Board are what designs their purpose & effectiveness. However this can't be true in the Global Operational trending we are living in without having certain outlines to follow. Based on all the angles we pictured the workflow from, we organized and structured the Enterprise's services and we made a call for the IT as a tool to handle all the aspect and also a number of IT frameworks such as ITIL, COBIT, CMMI that will help do the follow up. This is going to be the subject of our article in which we try to project the essence of Governance in Enterprises context.

Keywords: IT governance · Frameworks · ITIL · COBIT · CMMI

1 Introduction

In a global trending world, investments today more welcomed by the decisive board if it will give them more control of the environment they are in.

We can't deny however the importance of every single service of the Enterprise to the development of the activity, but the investment still needs to go right, straight and in effective way.

Calling this as the most important aspect, companies need to give the lead to a consulting activity (internally speaking as from the outside) to happen in the backstage to diagnose the situation from now and then.

In a World where such terms as global economy, the world is a village, and globalization are very trending. In a world where investments done are likely re related to controlling their environments, In a context of which goods and services that companies are responsible of need to go true a number of processes and selection of services to develop and grow. Investments are a huge deal.

So how does a company structure its own services, & how does it get most advantage from it? How does companies handle the workflow, activities, assuring a good running investments? and how Management side of enterprises can get controlled?

2 Guide of Management for Enterprises Frame

2.1 What's an Enterprise

A purpose established out of a combination of several concepts and elements mentioning activities, processes, structure, information, human resources, objectives, environment

It is structure is graphically illustrated by most enterprises on a chart format dispersing all factors & efforts depending on facilities & the range of activities & their outcomes. We can't deny the multiple variations of context; still there is a common pattern from which a company can't move on from, calling accounting, legalities, human resources, Finance and Information Technology. From Operations side of perspective, the enterprise itself has to be defined and represented on a scale of what happens & how & when

Those details for an enterprise gives a state that speaks out for the decisions made, the Interior environment, It's relation with the world around it, and that can answer and solve all the problems, frame all investigation and add more sequences if needed. The way of gathering it is under operations or processes frame, in order to be analyzed, compared and measured with the world models of success, of quality, of security 1.2 Additional Information Required by the Volume Editor.

If you have more than one surname, please make sure that the Volume Editor knows how you are to be listed in the author index.

2.2 How the Workflow Is Handled

Indoors, data exchange happened on an intern level at the start of the Technology times, to the automated exchange of messages, to Information Systems, to Cloud computing, Big Data

All those were and still are tools for managers and leaders to receive, track and follow up the updates in their respective department and project running.

Information technology takes then a huge place in enterprises, as created in the first place to meet and answer a problem of connectivity and communication, now has gained a respectful grade managing the business as a main objective.

3 Literature Review

3.1 Management Side of View and Use

In the corporate world, all is run in a specific way following a specific strategy to achieve a certain goal, this process of leading, administrating and directing a company [1] the management is an art and actions undertaken help a business operate, with instruction and policies from the board director office. Management refers to the actions taken by a company to lead the business in a positive direction. [2] even responsibilities that follows it are set, decided by a higher positioned positions in every enterprise and the results achieved have to be reported to the respective positions for a later follow up on

performances. Still we can't underestimate the strong position the managers has inside every business, as they know and they reinforce the competencies, they make themselves schedules based on that, and know how to deal and increase performances based on that. So without this set of responsibilities there isn't a way big objectives SMEs (small & medium enterprises) and corporate are setting for themselves will be achieved.

3.2 A Call for Governance, a Specific One

What is governance; we can't go true this paper and the next ones without really having a set of definitions about Governance as a word, as a practice and behavior. For that matter we will start with a Basic definition that states it being an Establishment of policies, and continuous monitoring of their proper implementation [3] where its first known use was on the 15th century. Governance is exercised by a decisive board and it is about protecting a business [2] by setting a number of rules or what we call Road maps for managers to execute and act based on it and they maintain the right to oversea the performance of managers in following the strategy and attending objectives set. A corporation without governance is like a train without a track. No matter how much potential the business has, it will never undergo the business transformation needed to get to where it wants because it has nothing directing its progress. [4] The governance makes a way to get a vision for the organization, and translation of the vision into policy, management is all about making decisions for implementing the policies.

As to be specific, governance board are luckily these days opting for a specific well serving new kind of Governance is the Base of information, data, reports, databases..... all what happens in Enterprises in all levels, IT Governance.

3.3 IT Governance and the Choice of Frameworks

We can't dissociate both management & Governance, and thank to the tools proposed for IT governance, a discipline in itself that reunite both managers and leaders of the executive board. It's the interrelationship between the two previous concepts which the following "Fig. 1" simplifies.



Fig. 1. Interrelationship between governance and Management [5].

IT Governance gives control all the aspects in a business divide it to a number of processes, so can be calling it the small entity or instance that means a task or defines a task in IT, working in IT is like gathering small pieces of a puzzle, needs concentration and invoke a complementarily between each puzzle, always seeking for a complete picture and frame to have.

And from this vision what can make this IT Governance happen true is what exists under the label of IT Frameworks. Those to frameworks that are represented and predefined by international institutions in the Topic, as a conceptual support and guidelines that serves to realize something, depending on the subject, and IT frameworks objectives is definitely to be able to organize everything under process structure that can be evaluated, measured & trusted after everything. So following and implementing those IT framework assure at the exact moment and IT management internally by the managers, and give highly chance to governance positions to control and impact the external environment and that's what the "Fig. 2" is about.



Fig. 2. Explains how to achieve an IT Governance [5]

This introduction to the world of frameworks, is subjective and based on experts reviews and institution's publications, they are a floor to prepare in order to be able to set what's needed for an enterprise exactly to be on the right path for efficient and effective governance. From the big selection of IT frameworks we had to choose the most impactful in terms of Governance.

Following the Calder-Moir framework tool [6] "Fig. 3", we can see a collection of frameworks methodologies and standards in use in the market today, it's an approach like any other that gathers all the elements that impacts from near or far the IT governance art of work. In terms of our attempt here we have chosen to work with the three famous and well known IT frameworks to all community of IT; ITIL, COBIT and CMMI.



Fig. 3. Calder-Moir framework tool

Enterprises have been opting and investing in getting the right IT Framework and adapting it to the internal environment, but the fact is there isn't a right framework for business, and for your Business you need a Set of frameworks. How to gather them, and adopt them are issues we will share ongoing with the development of this paper.

3.3.1 ITIL V3

ITIL is a framework & a tool; that have been introduced by the UK Government to structure in the best way information systems within enterprises & administrations, by producing a series of books documenting specifically how IT services should be done and developed following a cycle of process; so it focuses basically on IT. Developed since 1980s and its acronym stands formally for Information Technology Infrastructure Library. ITIL has gone true updates thanks to the comments and the feedbacks of the community readers & experts and now settled up at the Third Version since 2011, To organize all services in the organization, the framework proposes this structure of processes and functions, as the "Fig. 4" showcases that ITILv3 is organized around a service lifecycle which includes service strategy, service design, service transition, service operation and continual service improvement [7].

We will focus on this map of processes & functions later after seeing the other frameworks.



Fig. 4. ITIL v3 lifecycle of process & functions

3.3.2 COBIT 5

COBIT stands for Control Objectives for Information and related Technology, and it is a framework for developing, implementing, monitoring and improving information technology. The COBIT framework is published by the IT Governance Institute and the Information Systems Audit and Control Association (ISACA). It's concerned about what IT should be ding for enterprises, so this framework is designed with an enterprise perspective.



Fig. 5. COBIT 5 process for Governance & Management [8]

In its 5th version today since 2012, a proof of progress and serious vision to undertake the Governance part as it should, It does split now to 5 domains and 37 processes based on the Business plan, like in the next figure "Fig. 5".

3.3.3 CMMI

The abbreviation stands for Capability Maturity Mode Integration, its mission is assuring the transition within a structure; in a growing sense, and describing it by 5 levels, growing up every structure hit from level one to the last one by validating a number of process in each level, it's a game to measure growth. And it is used to appraise the maturity of a system, of an IT strategy or model and it known with 26 process Areas that are a way to judge at what level of maturity the enterprise stands if we can find the elements that answers the process Areas, and the "Fig. 6" represents them with the direction of growth.

Level	Focus	Process Areas	Quality
5 Optimizing	Continuous Process Improvement	Organizational Performance Management (OPM) Causal Analysis and Resolution (CAR)	
4 Quantitatively Managed	Quantitative Management	Organizational Process Performance (OPP) Quantitative Work Management (QWM)	
3 Defined	Process Standardization	Capacity and Availability Management (CAM) (svc) Incident Resolution and Prevention (IRP) (svc) Service System Transition (SST) (svc) Service Continuity (SCON) (svc) Service System Development (SSD) (svc, optional) Strategic Service Management (STSM) (svc) Organizational Process Focus (OPF) Organizational Process Definition (OPD) Organizational Training (OT) Integrated Work Management (IPM) Risk Management (RSKM) Decision Analysis and Resolution (DAR)	
2 Managed	Basic Project Management	Service Delivery (SD) (svc) Requirements Management (REQM) Work Planning (WP) Work Monitoring and Control (WMC) Supplier Agreement Management (SAM) Measurement and Analysis (MA) Process and Product Quality Assurance (PPQA) Configuration Management (CM)	Risk
1 Initial			Rework

Fig. 6. 26 process areas of CMMI [9]

4 Compromising and Combining for a Better IT Governance

4.1 Alignment Method

Compromising, this is the idea of this paper is to prove that in diversity force, this force in our context is forgetting about being stubborn and adopting one framework for you business, the solution to the force we named is by being open minded and combining different frameworks.

Achieving alignment requires everyone & every structure to have the same objectives, goals and vision. in Every business there is a number of process, that accomplish one the other. The Alignment study goes on multi-dimensional space; since for every process we will have to make a how to do persecutes. We'll gather the practices that are similar or closer to each other in terms of functioning and objectives. This approach allows a specific analyzing to the information, and a specific complete treatment to the information and the services in hands on the enterprise.

4.2 Results of the Alignment Methods

True this section, we will try to demonstrate the complementarities that the alignment of the three previous frameworks ITIL, COBIT & CMMI allows to have, and how more beneficial to the enterprises the work becomes.

4.2.1 Manage the IT Strategy in the Business

The first thing to start with when it come to a business is the strategy setting, quite a mandatory step not to skip and to be taken seriously, and that's why this section is going to go around this process, and the next figure "Fig. 7" explains how to get to an efficient IT strategy process and results from the respective frameworks IITL, COBIT and CMMI in the versions we mentioned earlier true the Literature review.



Fig. 7. Strategy Management objective combined form the practices.

For COBIT the process Manage the Strategy, provides a view of the current business and IT environment, the future direction, and the initiative required to migrate to the desired future environment, and it's purpose is to align IT resources with business objectives by measuring a number of metrics to be sure that process Goal and IT goals are achieved in the Enterprise no matter it size it's an adaptable figure [10] that's what the next two figure defines as KPIs and metrics "Tables 1 and 2".

This process defines the main objectives behind a business, and also how to manage & direct all resources in line with business and do the follow up by assuring a list of KPIs to assure the right follow up.

Table 1. A set of Metrics that measures the achievement of IT the goals [10]

Related metrics

- Percent of enterprise strategic goals and requirement supported by IT strategic goals
- Level of stakeholder with satisfaction with scope of the planned portfolio of programs and services
- Percent of IT value drivers mapped to business value drivers
- Numbers of business disruptions due to IT service incidents
- Percent of business stakeholders satisfied that IT service delivery meets agreed-on service levels
- Percent of users satisfied with the quality of IT service delivery
- Level of business executive awareness and understanding of IT innovation possibilities
- Level of stakeholders satisfaction with level of IT innovation expertise and ideas
- Number of approved initiatives resulting from innovative IT ideas

Table 2. Related metrics to process goals in an enterprise [10]

Related metrics

- Percent of objectives in the IT strategy that support the enterprise strategy
- Percent of enterprise objectives addressed in the IT strategy
- Percent of initiatives in the IT strategy that are self-funding (financial benefits in excess of costs)
- Trends in ROI of initiatives included in the IT strategy
- Level of enterprise stakeholder satisfaction survey feedback on the IT strategy
- Percent of projects in the IT project portfolio that can be directly traced back to the IT strategy
- Percent of strategic enterprise objectives obtained as a result of strategic IT initiatives
- Numbers of new enterprise opportunities realized as a direct result of IT developments
- Percent of IT initiative/projects championed by business owners
- Achievement of measurable IT strategy outcomes part of staff performance goals
- Frequency of updates to the IT strategy communication plan
- Percent of strategic initiatives with accountability assigned

And from ITIL side, the process selected gives a guidance on how to design and implement IT service later, the Generation of strategy generates four main ac activities, and the ITIL process is about Gaining Clarity and studying all the parameters;

- Define the Market, it's more about external to define strategies to offer services to the market [11]
- Develop the Offering
- Develop Strategic assets, or otherwise said prepare and use of the tooling-up, the resources that will do the activity.
- Prepare for Execution, goes true self-analysis as a preparation for the implementation of the strategy.

Those steps of doing in this process, helps gaining of clarity and book of parameters that might and will affect the activity and the services.

And to the CMMI, WP stands for Work planning is about estimating costs, effort, and schedules, figuring out how you will attack the service engagement, and involving the right people [12], it's about establishing plan, getting commitment to the plan and finally maintaining the plan.

So to get a complete structure that allows us to set the IT strategy process only, we start with ITIL thanks to it we gain a clarity of the environment, and second we measure the goals by using COBIT's List of KPIs and we assure commitment and maintain of the plan thanks to CMMI framework. Thanks to the study done by the three we get a strong consisting plan, this is a showcase of real dependence and complementarities between ITIL, COBIL & CMMI.

4.2.2 Manage the IT Configuration in the Business

The urge to control the programs, and their interactions, their benefits and objectives and more of those details related to the IT programs, will be gathered under the configuration management section, and the "Fig. 8" is going to help us explicit and understand this process inside the enterprise.



Fig. 8. Configuration Management practice multi-frameworks

The configuration in COBIT5 seeks to manage The IT environment; it encompasses software, hardware, network appliances, data structures, process definitions, documentation, credentials, personnel and any other element that is part of the IT infrastructure [13]. True these steps:

- Managed business risk: by recording é updating all the information [13]
- Compliance with external laws & regulations: verification of the compliance of the enterprise landscapes with an established baseline [13]
- Business service continuity & availability: a system to resolve incidents [13]
- Information based strategic decision making: based on documented information [13]

- **Optimization of service delivery costs:** mapping the entire services and their independencies and redundant to be able to cut the extra costs [13]
- **Optimization of business process costs:** map supporting IT hardware and software component to the service delivery business processes [13].

And this how the configuration of the IT environment is taken care of true the use of COBIT.

True ITIL, we have in hands a set of tools and date used for collecting, managing, storing, updating, analyzing & representing data about all configuration items and their relationship [14]. This data is more centered on the process and so detailed covering the roles and the tasks of coordinators and managers that will be affecting its working well.

In the CCMI frame, the configuration management is located second of the Maturity levels; where the basics are just starting to be Established, noted and verified to give the integrity of work products using configuration identification (labeling), configuration control (known modifications and permission to modify), configuration status accounting (final status of work products), and configuration audits (checks to verify changes) [15].

CMMI's goal of establishing a baseline of identified work products for configuration management can be achieved through ITIL [16].

We can picture now the way of applying the three approaches, first by applying COBIT for the whole environment and when it comes to process configuration there isn't best than to adopt ITIL and it CMDB(Configuration management Databases stored) and to identify the products work for the CM giving a call to CMMI.

These two manipulation gives an insight about a new Multi-dimensional framework here made out of ITIL, COBIT and CMMI.

5 Conclusion

Proper governance requires time and though committed leaders who understand the benefits of aligning those three frameworks and how to benefit from them so that the organization produces the desired results. [4] Regardless of the type of venture, only good governance can deliver sustainable and solid business performance [4].

We tried truly to give an insight on how to take advantage of what's existing (the frameworks) and reduce the time of inspection about the best, because there isn't a best practice, there is a best combination of practices enterprises can take advantage of, and we proposed a beneficial way to have a fruitful implementation of IT frames and good practices by applying our suggestion on One process area.

Aligning IT with the services of the business will help to have control of the current situation and model the future & that's a fortune making for Enterprises.

This paper was an opportunity to open up our perspective on a more interesting topic where we will detail further how the Alignment can be done for all process areas.

References

- 1. BusinessDictionary, What is corporate management? definition and meaning Business Dictionary.com. http://www.businessdictionary.com/definition/corporate-management.html
- Askenazi, G.: The Difference Between Corporate Governance & Corporate Management | Chron.com. http://smallbusiness.chron.com/difference-between-corporate-governance-cor porate-management-61799.html
- 3. BusinessDictionary, What is governance? definition and meaning BusinessDictionary.com. http://www.businessdictionary.com/definition/governance.html
- Boutros, T.: 10 Principles that Promote Good Governance | BPM, Lean Six Sigma & Continuous Process Improvement | Process Excellence Network. http://www.processexcel lencenetwork.com/business-process-management-bpm/articles/10-principles-that-promotegood-governance
- IT Governance vs. Corporate Governance vs. IT Management | ServiceXen Thoughts about Software Strategy, Marketing and Management. https://servicexen.wordpress.com/2008/06/01/itgovernance-vs-corporate-governance-vs-it-management/
- 6. Calder Moir. https://www.itgovernance.co.uk/calder_moir
- 7. Service Operation | ITIL. https://itil2014.wordpress.com/tag/service-operation/
- 8. Understanding the Core Concepts in COBIT 5. http://www.isaca.org/Journal/archives/2013/ Volume-5/Pages/Understanding-the-Core-Concepts-in-COBIT-5.aspx
- 9. Cmmi Product Team and C. P. Team, CMMI® for Services, Version 1.3 CMMI-SVC, V1.3 Improving processes for providing better services (2010)
- 10. Pittsburgh ISACA, Practical approach to COBIT5 (2012)
- 11. Strategy Generation Process [Globaal-Een] (2009)
- 12. Forrester, E.C., Buteau, B.L., Shrum, S.: CMMI for Services: Guidelines for Superior Service. Addison-Wesley, Upper Saddle River (2011)
- 13. I. wwwisacaorg, Using COBIT ® 5 CONFIGURATION MANAGEMENT (2013)
- 14. bryantopham and hpecom, Organize your configuration data efficiently HPE Configuration Management System software Market drivers for CMS (2012)
- 15. MTarnowski, CMMI for Services CMMI-SVC Plays-In-Business (2014). http:// www.plays-in-business.com/cmmi-for-services-cmmi-svc/
- 16. Tegtmeier, K., Johnston, D.: First steps in implmenting the CMMI for Services Model and ITIL (2011)

A Design Requirements Framework for Mobile Learning Environment

Abdel Karim Aziz^(E) and Faddoul Khoukhi

Department of Computer Science, Faculty of Science and Technics Mohammadia, University Hassan II Casablanca, Casablanca, Morocco Karim.aziz@karimoosteam.com, khoukhif@gmail.com

Abstract. Whether online or in a classroom, learners have long suffered from a lack of interactivity, several mobile learning system provide recorded lectures, which only enhance passive learning. Thus and in order to improve interactivity and motivation into the learning process, we propose in this paper a conceptual framework to be efficiently integrated into the learning process. This framework describes the design requirements of a mobile learning environment that can deliver live broadcast of real-time classroom teaching to online student with mobile devices. It suggests five perspectives: generic mobile environment, mobile learning context, learning experience, social learning strategy and learning objectives. It also clarifies how to apply these last perspectives and build a mobile learning application efficiently.

Keywords: Mobile learning \cdot Design requirements \cdot Conceptual framework \cdot Theoretical framework \cdot Mobile learning theory

1 Introduction

Mobile devices are becoming increasingly popular and they are reaching all levels of society. Mobile learning has become a great tool for delivering educational resources anytime and anywhere. Therefore and under the umbrella of making every second of the human life and specially students count, the idea of Mobile learning approach, which refers to the use of mobile technology for educational purpose, these devices can offer learning opportunities that are: spontaneous, informal, contextual, portable, ubiquitous, pervasive, and personal [1]. Mobile learning is one of new learning mode that students can use mobile devices to assist them to learn. As usually students like to use their devices to communicate or to surf the net looking for breaking news or even to watch videos, while being out of the classrooms, so by providing them mobile learning system into their devices they can use it to enjoy the process of learning in order to be like a never ending lesson for students, however, mobile learning does not replace the traditional education, but enhance the effectiveness of traditional learning process, otherwise it's like another way to get more information, a space for students to learn more, to get so close to their teachers and to get a lifelong learning to every learner. Mobile learning could provide learners with the maximizing learning autonomy, and also provide the instructors with more flexible teaching and managing methods [2]. Furthermore mobile
learning could be used to support learners autonomy, likewise enhancing the traditional learning, as explained by pilling-Cormick and Garrison, learners take primary responsibility and control of their learning process, including setting goals, finding resources, determining strategies, and evaluating outcomes [3].

The objective of this paper is to present factors and requirements needed for mobile learning environment; it also includes how learning theory can be applied in mobile learning context for educational purpose.

This paper is organized as follows: Sect. 1 describe an introduction to the paper, Mobile Learning is proposed in Sect. 2, in Sect. 3, Mobile learning framework is discussed, finally we conclude in the last section.

2 Mobile Learning

Today, there is a global revolution in technology, where there is a swing from desktop to mobile technology. The mobile technology is changing the way people learn and work as well as the way they collaborate and interact with each other lives and of course the mode of accessing information. With mobile usage expected to double within five years and overtaking the laptop access to the web.

Mobile technology has been widely used in all sectors to provide goods and services to customers. For example, libraries are being digitized and information so it might be accessed by mobile technology [4], in the banking sector, customers can access to the bank services by using mobile technology. Furthermore, the healthcare system is also making significant progress by using mobile technology in order to provide care services to patients, and of course users use mobile technology in general for entertainment.

The use of wireless, mobile devices is gradually increasing and diversifying across all sector of education [5], in fact now there is a growing size of seminars and meeting all over the world dedicated to mobile learning. In this paper, we are creating a conceptual framework of designing a mobile learning environment which will be capable to answer the following questions:

- 1. How can any learner take full advantage of a mobile learning environment?
- 2. How can we design appropriate and attractive framework for learners?
- 3. How can be effectively implemented in both formal (inside classroom) and informal (outside classroom) learning?
- 4. How can be a social environment to learners?
- 5. How can we make learners enjoy the learning process?
- 6. What are the characteristics Hardware that we need to improve effective learning?

3 Designing a Mobile Learning Framework

Designing a learning framework for mobile usage might require the use of a number of learning theories; likewise the most of those learning theories frequently used is based on the supposition that teaching and learning take place in the classroom. However the

case in this study is to develop a theory of mobile learning that happen everywhere and this is a colossal part of the conceptual framework.

For the purpose to develop a theory of learning there is a number of crucial factors that need to be considered. To begin mobile learning is not limited to the classrooms, so learners can all the time bring up their devices anywhere and start the learning process. This theory of learning will promote successful learners. Successful learning is related to effective learning. According to the US national research council [6], effective learning is based on four elements:

- 1. Learner centered
- 2. Knowledge centered
- 3. Assessment centered
- 4. Community centered

Furthermore Learner centered focuses on positioning learners at the center of learning process, and consequently, learners will have the ability to construct new knowledge based on both their previous and current knowledge. Knowledge centered look at the curriculum which is built on validated knowledge, taught effectively and efficiently. Assessment centered, on the other hand focuses on evaluating learners competency, diagnosis difficulties and offering learning support. Finally, Community centered promotes sharing knowledge and collaborative learning. These four elements called upon to the constructivism approach, to the socio-cognitive approach, to the social learning space which learners are not just an empty box to be filled with information, in short, learners are responsible for acquiring, construct and discuss their knowledge. In order to conclude those arguments, Sharples et al. [7] propose five issues like a checklist for developing theory of learning:

- 1. Is it significantly different from the current theories of classroom, workplace or even social learning mobile environment?
- 2. Does it consider the mobility of the learners?
- 3. Does it include informal (outside classroom) and formal learning (inside classroom)?
- 4. Does it view learning as a constructive and a social process?
- 5. Is learning analyzed as personal and mediated by technology?

Parsons et al. [8] proposed three crucial requirements to design M-learning environment: generic mobile environment issues, M-learning contexts and learning experience and objectives.

To start, generic mobile environment enhances the study of mobility (the mobility can be conceptualized in different way, mobility of learner, mobility of the device, and the mobility of services), user interface, the use of rich media, and communication support. A study lead by Dewitt (Saedah, Dewitt and Norlidah Alias) [9] demonstrates that the use of short text messages among secondary school students has been identified as an effective tool to improve collaboration learning. On the one hand, Parsons et al. [8] classify this as learner roles and profile; on the other hand, Löwgren and Stolterman [10] categorize it as core, periphery and context. The second element proposed is M-learning contexts. Based on Wang's [11] dimensions, he suggested six dimensions of M-learning context: identity, learner, activity, collaboration, spatial-temporal, and

facility. Parsons et al. [8] place the first four, as situational context for M-learning and the last two as environmental context. Finally the last element proposed by Parsons et al. to design M-learning environment noted two useful metaphors, the cinematic metaphor and the game metaphor, the former deals with the story elements and narrative, likewise the latter deal with features of games, such as excitement, suspense and competition.

This paper explores a conceptual framework which is based on five elements:

- Learning theories
- Generic Mobile Environment
- M-learning Contexts
- Learning Objectives and Experiences
- Social Learning

3.1 Learning Theories

The key to success lies in finding the appropriate points for integrating technology into a new pedagogical practice. The task of translating learning theory into practical platform would be greatly simplified if the learning process itself was relatively simple and straightforward, unfortunately this is not the case; learning process is a complex strategy that has generated many interpretations and theories of how it's effectively accomplished. Therefore we find that is inevitable to include learning theory into the conceptual framework. Furthermore each theory has the specifics features; on the one hand, constructivism activities will promote learners to construct new knowledge based on both their previous and current knowledge, on the other hand, Socio-cognitive activities will promote to learners the acquisition of knowledge through observing others within the context of social interactions, experiences and emotions. Finally drill and exercise on the platform is a characteristic of behaviorism, especially when coupled with the fact that with this framework, feedback is immediate. The feedback can act as a motivator for learners.

3.2 Generic Mobile Environment

Under the matter of the general mobile environment, we might have to gather certain information to define the profile of people that are implicated in the process of learning; this element is classified as a user dimension. The understanding of this factor is necessary important because every learner has a different profile that is related to his background, for example a great highly self-esteem learner will improve his mechanism of learning by using the learning system, likewise a slow learner could enhance himself and be recover from his delay by benefiting from all what the system has to offer; furthermore the user dimension is categorized in three fractions; the first one is learners who are the core of the process of learning, then the periphery that include persons who are involved indirectly in the learning experience, however they have a part to play (e.g. instructor), finally there is the context which will be the scene where the leaning process would take a place. So to make a great design we must take under sight all parties that are involved and manage all the relationships. The awareness of the material design as a matter of fact is one of the environment mobile issues, knowing that mobile devices has small screens, a short life battery, a slow connection and restricted input methods, the design of this interface must take all of those obstacles under hand and responds to all the needs of the learners. As alternative this framework is the answer to resolve these issues. The design of the interface would be so flexible, not complicated, and will not consume excessively the life of the battery.

In another point if the learners faced any troubles of any kind in the process of learning, the learning system should have a support assistance that could guide them to solve it and make sure that they get all the help needed; the support also offer them a space to discuss it with other users and try to find the answers. This is all in term of collaboration learning.

3.3 M-learning Contexts

According to Parsons et al. [8] each user has different psychological properties that relate to their experiences. Koole [12] propose the - learner aspect - as learner profile. In this context, each learner has to be treated in different way, according to his learning style, knowledge level and concentration level.

M-learning context can collaboratively, perform activities of learners, which enhance learners' experiences. An example of such a project is the Ambient Wood Project conducted by Rogers and Price [13] confirms that placing students into the wood to perform scientific research has allowed an integrated approach in teaching of science. Student were doing extensive collaborative job with their peers using digital devices. The project also proves exchanging experiences and skills between students. That's way social learning is becoming a part of the learning process. Finally spatial-temporal issue is relevant to the sense of time and/or location of learning process.

3.4 Learning Objectives and Experiences

The Learning experience issue is related extremely to the learner background, according to Preece et al. [14], the framework should not concern just the usability goals, but also take into the consideration user experience. Also another statement by the same author, suggest that the framework should be enjoyable, satisfying and motivating, and also we propose that the content of the framework should be determined and organized. A determined content enhance understanding, likewise, organized content promote learning acquisition and learners feedback.

In considering framework objectives, the content should be organized in narrative mode (story mode), learners will promote new knowledge or improve skills, and also a big advantage of narrative mode is that it can be explorative, allows learners to reflect on what they have learnt, illuminating the process of learning, and construct an organized competence.

Another factor that could engage learners in dealing with the framework is challenge. So by providing competitive activities, learners will be more exciting and get more interest to the content.

3.5 Social Learning Strategy

Social learning is defined as working together with another person or group to accomplish shared objectives, also the term social deals with interaction between peers.

Social learning is a strategy in the framework to increase learner motivation, to enhance cooperation, to share common goals, and to recognize the reality of diversity between learners. According to Julita Vassileva [15] the new generation of learners has different patterns of work, attention, and learning preferences and that is due to the development of communication technologies. As a matter of fact, social learning promotes learners to synthesize, communicate, and discuss ideas in ways that advance conceptual understanding, also social interactions and uniquely mobile activities should be integrated into learning practices since they are part of learners life and because these activities can foster a better understanding of the environment around the learners.

Now, we are witnessing, how learning process is excessively in tension with more dynamic content, connected learners, and collaboration ways of finding and exchanging information. In this study we cannot ignore these new channels of communication and knowledge transfer. After all and designing uniquely mobile experiences is in order to help learners make meaning of the world around them and making them more motivate to learn better.

3.6 Discussion

At the end of our study, we have generated a figure (Fig. 1) to summarize all crucial factors and requirements that we found fundamentals to build a mobile learning system. By reflecting on synthesized information's as presented in the figure, we are able to present an understanding of each factor and the applicable reason to answer the research question.



Fig. 1. The requirements design of a mobile learning environment

All factors are labeled as a table; each factor is an intersection between column and line, for example recognize learner profile is for behaviorism theory and learning objectives of the mobile learning system.

The figure below summary all factors and requirement which seem to us as important in order to design a Mobile Learning Framework:

4 Conclusion

Mobile learning is an emergent phenomenon and it's still in search area, although in this paper we have proposed the factors and requirements which are seen as necessary in order to design a mobile learning environment. The integration of learning theory was for purpose to transform the basics of a classroom into a learning system.

As a perspective of this study, a mobile platform has been embarked upon using the proposed framework. It is hoped that when the platform is in production mode, learners will have increased their current skills, improved their knowledge with a new way of learning, acquired social skills and enhanced teamwork skills with others learners.

This platform will open the doors of new knowledge, collaboration work, social interactivities and ubiquitous access to information to everyone at anytime and anywhere. In addition to this, we wanted in this study to pursue the evolution of students believes and makes the process of learning more adequate to them.

References

- Pilar, R.-A., Jorge, A., Cristina, C.: The use of current mobile learning applications in EFL. Procedia Soc. Behav. Sci. 103, 1189–1196 (2013)
- Singh, A., Somani, R.K.: Cost effective model for e-learning. Int. J. Eng. Innovative Technol. (IJEIT) 2(6), 271–275 (2012)
- Pilling-Cormick, J., Randy Garrison, O.D.: Self-directed and self-regulated learning: conceptual links. Can. J. Univ. Contin. Educ. 33, 13–33 (2007)
- 4. Sanou, B.: ICT facts & figures. In: The World in 2015. Itu 150 Años (1865–2015), p. 6 (2015)
- 5. Traxler, J.: Defining, discussing, and evaluating mobile learning: the moving finger writes and having writ... Int. Rev. Res. Open Distrib. Learn. (2007)
- 6. National Academy Press, Washington, D.C. (2004)
- Sharples, M., Taylor, J., Vavoula, G.: Towards a theory of mobile learning. Proc. mLearn. 1, 1–9 (2005)
- Parsons, D., Ryu, H., Cranshaw, M.: A design requirements framework for mobile learning environments. J. Comput. 2, 1–8 (2007)
- DeWitt, D., Siraj, S., Alias, N.: Collaborative mlearning: a module for learning secondary school science. Educ. Technol. Soc. 17, 89–101 (2013)
- 10. Löwgren, J., Stolterman, E.: Thoughtful Interaction Design (2004)
- Wang, Y.: Context awareness and adaptation in mobile learning. In: International Workshop on Wireless and Mobile Technologies in Education, pp. 154–158 (2004)
- Koole, M.L.: A model for framing mobile learning. In: Mobile Learning: Transforming the Delivery of Education and Training, p. 39 (2009)

- Rogers, Y., Stanton, D., Thompson, M., Weal, M., Price, S., Fitzpatrick, G., Fleck, R., Harris, E., Smith, H., Randell, C., Muller, H., O'Malley, C.: Ambient wood: designing new forms of digital augmentation for learning out. In: Proceeding of the 2004 Conference on Interaction Design and Children Building a Community, IDC 2004, pp. 3–10 (2004)
- 14. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction. vol. 18, pp. 68–68 (2007)
- Vassileva, J.: Toward social learning environments. IEEE Trans. Learn. Technol. 1, 199–214 (2008)

Knowledge Management in Business, A Multi-desciplinar Science and A State of Mind

Ben Laadar Hajar^(☉) and Cherti Ilias

Department of Mathematics and Computer Science, Faculty of Sciences and Technologies, Hassan Ist University, Settat, Morocco benlaadar.hajar@gmail.com

Abstract. A complex concept that many structures tend to avoid and neglect, this is an opportunity to have an insight of the concept of Knowledge Management plus get advantage of learning about it from a perspective of different disciplines in Science, Their combination gives the recipe for a right and so called complete frame of Knowledge Management. In this paper we give an overview on the state of art in knowledge management that can handle governance of the business afterwards without having to worry about the wellness of the internal environment and how to capture it.

Keywords: Knowledge management · Disciplines · Future decision · Modeling

1 Introduction

Human been are in a continuous search for a better quality conditions at work, thus this is not just the wish of employees but also CEOs, managers and leaders in all organizations for a better now and a better future.

We all know that organizations can be visualized under different functionalities, conceptualized in diverse forms with their specific customers, and can exist under one land or split or spread worldwide. It's surely structured in a set of arrangements to be under a so called control and follow up, also operational & defined on a scale of representing what happens around, and when and how secured, managed and governed by IT services and tools to make life around easily.

With all those components, does Enterprises have a real hold of the state needed to make decisions, details about the internal environment, It's relation with the world outside the doors of the building, and that can answer and solve all the problems, frame all investigation and add more sequences if needed?

The urge need to define everything around us, took us to a level of hitting and searching for the smallest details to be able to see the big picture, and true this paper we will do the same.

2 Knowledge Management's Enigma

Scientists in different disciplines, seek to flash out the Concept of Knowledge in Enterprises, redefine it, structure it and organize it's components since the 1980's see even before. For a better understanding of the meaning of issues like data, information and knowledge need to be recognized. Generally, data are defined as raw facts, information is viewed as an organized set of data, and knowledge is conceived of as meaningful information [1]. We got to select those definitions out since they are close to our Local reality. Thus the same reality drove us to discuss the multiple faces of the knowledge itself.

An informal survey conducted by other authors identified over 100 published definitions of knowledge management, and of these, at least 72 could be considered very good! Clearly, Knowledge Management is a multidisciplinary field of study that covers a lot of ground [2]. We are in front of true puzzle we are going to be assembling, combined of efficient & effective sciences & practices.

Our community needs to explore more in the context of SME's as an implicit behavior just like operations for its the survival. And this something we showcase true our communication, and simplify thanks to the "Fig. 1" where we have gathered the most impactful fields under which we noted the impactful practices to complete.



Fig. 1. Contexts that impacts knowledge Management

This attempt was made to gives a hold of an almost complete Knowledge; every other attempt to represent a complete formula of Knowledge will stay doubtable. And in the next section of this paper, we will go true details of the concentrated idea of "Fig. 1".

3 Corporate Communication – A Right Setting of Values and Philosophy for a Strong Culture on Board

We can't go and dig deeper in the structure of a company without really feeling it's context from an internal and external perspective, and that's what most consultant miss in their consultancy services, defining the belonging environment is a mandatory step to go further with the analyses, being able to see the details in the big picture; the work-place as most of us call it, it showcased as organizational values, that mostly are communicated in a chart. It's the collective/personal values established, and acquired from around and used in daily exchange, true it they endure convictions that influences later their choices, decision, and behavior.

By defining Business Culture, you sit in a state of mind for the corporate, a collective spirit that all the employees lives permanently and impacts their daily interactions with colleagues, clients even far away to the community, Having a strong culture on board reflects and transmits credibility, ethic and coherence to the surrounding environment, and "Fig. 2" below describes the concept Circle of influence of **Stephen Covey** with the different layers of enterprise, where the CORE interest is the change in the individual, that will lead to the second sphere where two individuals starts sharing the values, and the third sphere where the Team has the culture needed, and later on with time where the organization is immune, strong & ready to impact the external environment.



Fig. 2. Circle of influence of the Culture in an Organization [3]

And from here, organization can be sure that sharing of competencies, transition process of tasks & projects, processes, ways of doings, methods, data, information. It is an assurance for companies for a smooth & successful running of the corporate true the path without any kinds of complications.

4 Talent Management for a Global Competencies Model

Competencies are what matters in a Company's context, they give an overall picture of the strengths & weakness of the Teams, updating them regularly by the Leader Role in every structure is essential to have hold always on row data & information, and how to

get to a complete picture of the state, is by taking care of those points more and more by Managers:

- Making a Structure of Jobs: Setting the Model of Competencies raw as imagined or decided by the CEO and his team, by Defining Responsibilities, Job descriptions, goals to attend for each and every position no matter it size, define also the interrelationship between the positions, how the serve or complete each other, and prepare manuals if need to make a schematization of this Grill.
- **Regular analysis:** Competencies Model, being developed at the beginning with the care of the Human Resources department or responsible, and aligned to it recruitment have been executed, this document needs to be updated regularly by putting in place Performance assessment to track individual's development of the Experience they are having, the work & innovation they are creating, to gain their trust also that it's more about creating the environment and the experience for them to grow more.
- **Defining roles of Leadership:** in different areas of interest inside the organization who'll help the HR developing their vision and reach their objectives in every department, to get notes of the practical know-how, processes.
- Education Plan: with Talent's education (talking theory & practice) organizations can improve the competencies of their Teams and give them an impulse of motivation to achieve their goals and fulfill their job and at the same time Structure have updates of the every stage.
- **Track Education:** it assures a continuous insurance that the aims after this investment is going right, also that teams are satisfied and performing at their best, that learning instaured during the sessions of education is still leaving and shared and used & making specially considerable results.

It's a non endless circle, that keeps on developing it's aspects, so it shouldn't be broken nor interrupted. This will be helping the advising board to measure later the effectiveness of their teams toward projects, and be in capacity to judge the efficiency of the company on what they do.

5 Information Technology's Share of the Knowledge of a Business

Information Technology (IT) is a small as an abbreviation that interacts with the needs of any structure in different shapes by purposing different solutions. It also acts like a support function to all departments of an Enterprise who seeks to organize their knowledge, their existence and in our context we are in front of a situation where we have to classify how IT can support individuals, teams, organization.

True the figure proposed by the SECI model matrix of knowledge creating process [4] is used in the Fig. 3.

	Tacit	Tacit	
Tacit	Socialization: Social networks Forums Communities of practice e-Mail Groupware Group decision support systems Conference systems Chat groups IPTV Etc.	Externalization: Expert Systems Blogs Wikis Good / bad Examples of god and bad practice Questions and answers Decision support systems Business modelling Knowledge warehouses Cognitive mapping tools IPTV Etc.	Explicit
Tacit	Internalization: • e-Learning • Using wikis • Using expert systems • Web browsers • Using decision support systems • Using decision support systems • Using analytical solutions • Using statistical analysis • Using data mining results • Using data mining results • Using neural networks • Using social networks • Using forums • Using communities of practice • Using case based reasoning systems • IPTV • Using other KMS	Combination: Wikis Content management systems Data bases Data warehouses OLAP analytical solutions Business intelligence Data mining Statistical analysis Machine learning Neural networks Intelligent agents Artificial intelligence systems Case based reasoning systems Document systems Workflow systems Yellow pages Knowledge maps Electronic bulletin boards Intranets Web portals Genetic algorithms solutions Etc.	Explicit
	Explicit	Explicit	

Fig. 3. IT and KMS support SECI model of knowledge creating process

All of those tools are used on a daily scale by employees to manage their contact with customers and suppliers and to prepare dashboards of information, data, processes to CEO's and Decision Makers.

6 Capture of the Future True Knowledge

After all these introductions in the multi world science windows true the previous part f this paper; the questions that goes to everyone's mind is how to get hold of that knowledge when everything is available under our hands, from tools to users to structure. The step is introduced simply under the concept of capture and treasure hunt for Data, but never explained as a complex concept that needs not only the presence and the expertise of an IT engineer or responsible to decrypt all what is collected, but also the mandatory expertise of behavior psychologist, coaches, Human Resources manager, and head chiefs of departments, it an combined interpretations based on the source of knowledge.

The Figure below was the closest to the idea, as we have in a first step identified the source, and we conceptualized the model we want to see, what goes after is the step where only IT Expert can interfere by codifying the model and organizing it use and implementation true different parts of the organization (Fig. 4).



Fig. 4. Key Knowledge acquisition phases [2]

Implicitly here, we described the resources we can invest in their presence to have hold of the Knowledge Management cycle and keep it alive running capturing more and more of updates the right way, and the "Fig. 5" displays the strength IT has as a tool to get to the objective.



Fig. 5. Cycle of Knowledge Management in companies [5]

As it shows IT interferes in all the steps of the Cycle of Knowledge Management as a Support & Major function, from its creation or caption from different sources, to its sharing and without forgetting the storage and later On the updating step that remains necessary.

7 KPIs of Maturity of the Ongoing Operations

Evaluating the performance of Knowledge management operations, it is a good practice and process to have evidence of what improvement happening indoors, and to be able to monitor the progress the structure will be making, and archives are always raw data that can be used at anytime later for research studies.

KPIs (key performance indicators) they are of huge importance to the evaluation as they allow to represent and end up with results, that can be stated, they are metrics that helps to quantify or qualify a strategy as a winner or success depending on the source or the formula.

In our context, what can be measured will be the performance in terms of the extraction of knowledge and how we use it to predict the future and Govern and manage changes that will happen, so when it will come to those elements that makes the activity of every enterprise how see things will be different and surely how we measure them also, so for:

- Handling projects: usually measured by reports with studies that showcases and determinate the time of treatment of a portfolio, number of resources used, budgeting for every project apart with the presence of the combined study we will be able to predict not only the deadline of achievements of results but also who would be best to be part of the team responsible of the project, and how tasks can be dispatched and how to make most benefit of it.
- The improvement of Learning & sharing program: the strategy set by the CEO & his leading team and executed by the managers to assure the development of the firm and the competencies in it, has to be measured too, usually by the number of profiles recruited, number of profiles going to trainings, thanks to the diverse expert team we will be on an other getting carriers designed inside a company from Day 1 of joining it and managing this programming true serious and numerous factors that can help see the future of an employee within the closed environment of an enterprise.
- Sharing of knowledge: it could be measured true the platforms of sharing by using always the IT skills to keep posted the managers about the efficiency of all the new implementations done and their efficacy in terms use, understanding, motivation, To check the involvement of the teams and the individuals and manage having it properly later.

8 Conclusion

Knowledge Management shaping is what this research is aiming to get as a result, by evolving a diverse team with different disciplines and backgrounds and can be explored from various perspectives [4]. It is before anything a state of mind, a number of practices and managed by a strong IT Science.

Everything in Life should be looked at from various and multiple perspectives and that a concept my paper is developing to have a big picture, matters on the state and concrete impactful results and predictions on the near future of enterprises.

It's more about proving a point that with all what technology can do, it stays incomplete if you don't combine it with Our sciences.

So at the end we should never forget to listen! Listen to your users, customers, and managers-whichever audience for which you are designing. They will tell you how you can meet their needs and have a successful KM initiative [6].

References

- Alipour, F., Idris, K., Karimi, R.: Knowledge creation and transfer: role of learning organization. Int. J. Bus. Adm. 2(3), 61–67 (2011)
- 2. Dalkir, K.: Knowledge Management in Theory and Practice, vol. 4. The MIT Press, Cambridge (2005)
- 3. Global AIESEC Experience leadership Competency Model Global Competency Model AIESEC Experience Leadership Competency Model Global AIESEC Experience leadership Competency Model

- 4. Natek, S.: Knowledge management systems support seci model of knowledge-creating process Abstract (2016)
- LogicalDOC: Knowledge Management System Software. http://www.logicaldoc.com/en/ solutions/knowledge-management. Accessed 06 Feb 2017
- 6. Hasanali, F.: Critical Success Factors of Knowledge Management (2002)

Advances Networking and Sensor Networks

Survey of Security in Software-Defined Network

Nadya El Moussaid^(III), Ahmed Toumanari, and Maryam El Azhari

LISTI, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco nadya.elmoussaid@edu.uiz.ac.ma, atoumanari@yahoo.fr, maryam.ensa@gmail.com

Abstract. The requirements of cloud computing are putting the traditional networks in tension which influence the quality of the services provided by cloud computing. Therefore, the application of software defined-network (SDN) within cloud computing reinforces the dynamicity and flexibility of cloud. Recently, SDN is the trend in networking and virtualized networks, where, SDN separate the network control plane from the data plane, which leads the management of the network routing from decentered architecture to centered architecture. Despite the advantages of merging the SDN paradigm within the cloud environment, the security issues still in the surface. This paper presents a survey on the security issues in software-defined networking and the challenges faced by admins and providers in order to guarantee a secure environment with a resume about the proposed solution.

Keywords: Software-defined network · SDN · SDN security · Privacy

1 Introduction

The cloud computing introduced the unlimited virtualized resources that changed the way of accessing and storing data. The cloud characterized with the five essential characteristics namely: (1) Resource pooling, (2) On-demand capabilities, (3) Broad network access, (4) Rapid elasticity and (5) Measured services. The providers offer these characteristics in the form of three major services such as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Because of this attractiveness of the cloud, several organizations migrate from adopting traditional warehouse infrastructure to utilizing services provided by cloud computing [1]. Also, attracts the attackers to seeking for any vulnerability that can help them getting access to sensitive data or to get benefit of the advantages of the cloud in order to exercise attacks from the cloud against other organization.

The virtualization and the shared resource between multiple tenants are the backbones of cloud computing, the virtualization can be as a virtual machine and a virtual network. Software-defined networking is part of the virtualization systems, the use of SDN technology may improve the performance of network routing within the cloud computing. However, it increases the sensitivity to security issues namely confidentiality, integrity and availability issues.

The SDN is an element of the software-defined system (SDS) package that contains:

- Software-defined Networking (SDN)
- Software-defined Cloud Networking (SDCN)
- Software-defined Storage (SDS)
- Software-defined Data Center (SDDC)
- Software-defined Radio (SDR)

SDN provides five major benefits [2], that we quote:

- Accuracy: The IT resources become automatic and programmable. Also, the requests of clients are independent of the hardware.
- Agility: The agility enables the components from migrating between environments in an easy and flexible way.
- Adaptability: this property provides no reliance on hardware resources of the vendors, which leads to the adaptability to new configurations and environments.
- Assurance: SDN provides an assurance that organizations are able to specify their own policy.

In addition to what is mentioned above, SDN characterizes with other advantages such as:

- Network's centralization: The SDN adopts the centralized monitoring and management of the network, as well as to the centralized security.
- Hardware optimization: The SDN reduce the use of physical hardware by the orientation toward virtualized network infrastructures. When we say optimization of the hardware, we say coast reduction as well.

2 SDN Architecture

The SDN architecture is characterized by the separation of control plane from the data plane. Control plane is the brain who takes the decision of traffic networking. Data plane or forwarding plane is responsible of forwarding traffic, according to the control plane to the next component. The architecture of SDN contains three layers namely application layer, controller layer and infrastructure layer as it's showed in Fig. 1.

2.1 Controller Layer

Controller layer contains a bunch of controllers that are responsible of controlling the network. In other words, the controller layer is the control plane which is the principal component that takes the decision about the optimal path that traffic will take and monitor the behavior of the forwarding network. The controller uses protocols in order to configure the network devices such as OpenFlow [4, 10, 15].



Fig. 1. SDN's architecture

Controllers communicate between them through east - west interfaces in order to maintain the synchronization and connectivity of the network [2, 3, 6, 8]. The controller layer communicates with other layers using north-bound API's and south-bound API's. Where, north-bound API's (e.g. REST, frantic, etc.) are used to communicate with application layer and south-bound API's (e.g. OpenFlow, NetConf, etc.) to communicate with infrastructure API's [8–10].

2.2 Application Layer

The application layer is built on the controller layer, which represents the first layer in SDN architecture. It contains a set of software related to business requirements such as intrusion detection systems (IDS) [11], network virtualization [12], load balancing [13], and so on. In the case of changes at the application layer, controller layers afford an abstraction of network's resources to be allocated to the software of the application layer, in order to avoid reconfiguration of the network's resources such as switches and routers.

2.3 Infrastructure Layer

The infrastructures layer is also known as the data forwarding layer, it contains virtual or physical network resources and devices. As its name mentions, it's responsible for the forwarding of packets of the network according to a set of rules within the flow table [10, 14, 15]. The flow table entries contain three section namely the pattern, action, and stats [2]. The pattern represents the header field of a packet; the action is executed according to the match of the rules, then stats, which are indications that indicates the network's status.

3 Security Issues

In this section, we present a set of security issues of different layers that may lead to a successful attack.

Open programmable APIs: As it's mentioned above, SDN communicates through programmable APIs, these APIs can be open which may cause security issues by making the layers open and the vulnerabilities of components of the SDN visible to attackers. This issue may lead to cross-site scripting attack (XSS) or injection of malicious code [16, 17].

The controller issues: Because of the central architecture of SDN, the configuration and the decision of the network is taken by the controllers. Therefore, an exploitation of vulnerability can gain the attacker to take control of the whole network which can cause huge damages [11].

The SDN switches issues: switches within SDN suffer from the limitation of entries of the flow table. This issue makes switches very sensitive to DDoS attacks.

4 SDN Attacks

SDN attracts attackers to look for vulnerabilities in order to use them to exercise attacks or a set of attacks. In this section, we classify the attacks according to target the layer.

Figure 2 shows the different attack point in the SDN architecture, which an attacker can exploit the existed vulnerability [11].

An attack can be exercised on the component of the application layer, against controllers of controller layer and at channels of communication between controllers. Switches are not excluded from these attacks. It also, can target the programmable API's that connects layers to each other.

4.1 Application Layer

The application layer may contain vulnerabilities related to software and the difficulties of modeling a global security policy that is able to manage the whole network without fails. Where most of the applications are developed by a third party, which doesn't take into consideration the mechanism of security standardization.

The rest of this section describes the major attacks faced by the application layer.

Unauthorized access: The large number of devices of a network may lead to the misuse of the application running on controllers by an intruder to gain unauthorized access to sensitive information such as network information. As most of the applications are made



Fig. 2. Attack point in SDN architecture

by third parties that have a limited knowledge about the security requirement such as the management of authentication systems, authorized accesses to an application, and the access of applications to the network information [11].

Malicious programs and application injection: Code, programs or applications injection is one of top ten attacks that target applications especially web applications according to OWASP project [16]. This attack may cause unauthorized access, data loss or information corruption. It's used by worms to propagate within the network. Also, it helps attackers to gain more privileges to accomplish their malicious tasks.

Insertion of rules: In order to get the benefit of the advantages of SDN, SDN paradigm is applied in various areas namely cloud computing, data centers, cellular networks, wireless networks, mobile networks, etc. where the number of devices is huge, with complex applications and services. Therefore, The insertion and the management of security rules is a big challenge for administrators and providers in order to prevent security rules conflicts between applications and services [5, 6, 11].

4.2 Controller Layer

The controllers are the brain of the SDN. Thus, because of its importance, attackers aim to get control of the whole network by exploiting the existing vulnerabilities. This section presents the well-known attacks faced by controllers.

Attacks from application layer: Applications are running on controllers, were any successful attack on application layer may lead to security issues in the controller layer. For example, application injection attack can gain access to network devices information and monitor the behavior of the network, or exercise other attacks for more serious effects.

DDoS/ DoS attack: Denial of service (DoS) and distributed DoS (DDoS) is the simplest attack exercised by attackers that target the availability of the network and services for the legitimate users [11]. This attack consumes the controller's resource such as CPU, memory, and bandwidth by rules installation and computation from the flooded flow requests [18]. Once the controller is saturated, the legitimate requests will be dropped and the switches connected to the affected controller will be affected as well [18].

Attacks against distributed multi-controllers: Because of the division of the main network into sub-networks, the need of using distributed multi-controller raises. This solution was proposed to overcome the DDoS/ DoS attacks and preventing the shutting down of the whole network. However, the SDN remains sensitive to DDoS/ DoS attacks, and to other issues related the management of the security policy and security conflict [11, 18].

4.3 Infrastructure Layer

Switches within infrastructure layer are divided into three part especially the OpenFlow switches: OpenFlow agent, packet buffer and table flow, which are a target for DoS attack.

DoS Attack: To perform the DoS attack, the attacker performs "the flow request flooding" by interrupting the performance of the three parts of OpenFlow switch. He/she sends a large number of malformed packets to saturate the OpenFlow agent since it generates a limited number of flow requests per second to be sent to the controller. Thus, the target switch is affected as well as to the hosts connected to the victim. In the case of a full packet buffer, the victim switch sends instead of packets headers, the entire packets to the controller that lead to the consumption of the bandwidth and channel congestion [11, 18]. Another drawback of the OpenFlow switch is the limited entries of a flow table, where the attacker aims to overflow it by installing new rules. This attack leads to dropping rules of legitimate flow [20].

Man-in-the-middle: The attacker of man-in-the-middle (MITM) monitors the traffic between controllers and switches, in order to intercept the information of communication without being detected. Controllers and switches are not directly connected to each other, which makes each entity doubtful to be a MITM node [19]. MITM attack

SDN layer	Security issue	Proposed solution	Description
Application layer	n Unauthorized access	OperationCheckpoint [23]	Presents a permission checkpoint to verify the authorization of applications
		SE-Floodlight [24]	Includes the digital authenticated northbound API for a minimum privilege
		AuthFlow [25]	Presents authentication and access control mechanism based on host credential
		NICE [22]	Verifies the correctness of OpenFlow application by automated the testing
		VeriCon, Verificare [26]	Verify the correctness of controller's applications and verifies the correctness of execution of any single network event
	Malicious programs and application injection	FortNox [27]	Provides role-based authorization and security constraint enforcement for the NOX OpenFlow controller
		LegoSDN [28]	Introduces fault-tolerant controller framework that allows SDN controllers to isolate and tolerate SDN application failures, in order to increase the availability of the network
		ROSEMARY [29]	Implements a network application containment and resilience strategy based on the notion of spawning applications independently
		NetPlumber [30]	Presents real-time policy checking and incrementally checks for compliance of state changes, using a dependency graph between rules
		Anteater [31]	Diagnosis problems through static analysis of the data plane in order to identify policy conflicts

 Table 1. The proposed solution to sdn security issues [11, 42]

(continued)

SDN layer	Security issue	Proposed solution	Description
		Flover [32]	Introduces a model of checking system which verifies that the aggregate of flow policies instantiated within an OpenFlow network does not violate the network's security policy
Controller layer	Attacks from application layer	SE-Floodlight [24]	Tracks the event flow of application to detect any attack that may come from applications
		FRESCO [33]	Implements different security function such as firew alls, scan detectors, attack de flectors, or IDS detection logic
	DDoS/DoS attack	FloodGuard [34]	Introduces a scalable, efficient and lightweight framework for SDN networks to prevent data-to-control plane saturation attack by using packet migration and data plane cache.
		CONA [35]	Analysis the content of requests made by the client to a server in order to reduce the harm of DDoS/DoS attacks
	Distributed multi-controllers	HyperFlow [36]	HyperFlow localizes decision making to individual controllers, thus minimizing the control plane response time to data plane requests
		McNettle [37]	Presents an extensible SDN control system based on multi-cores CPUs to control event processing. The processing of events related to the number of CPU cores
		DISCO [38]	Presents a distributed DISCO controller, where each one manages its own domain and communicates to each other to share and provide network services

 Table 1. (continued)

(continued)

SDN layer	Security issue	Proposed solution	Description
Infrastructure layer	DDoS/DoS attack	VAVE [39]	Provides a solution that verifies the validity of source address that causes DoS attack
		FlowVisor [40]	Presents a switch virtualization, where the same hardware forwarding plane can be shared between various logical networks, each with a distinct forwarding logic
	Man-in-the-middle	VeriFlow [41]	Presents a layer between a software-defined networking controller and network devices, and supports analysis over multiple header fields, and an API for checking custom invariants
		FortNox [27]	Verifies the legitimacy of the modifications through digital signatures or security constraints

 Table 1. (continued)

leads to the implementation of other attacks such us eavesdropping and black-hole attack [11, 20].

5 Countermeasures of SDN Attacks

This section deals with the solution that has been proposed to solve some of SDN security issues mentioned above. The following table (Table 1) summarizes the proposed solution with a description.

6 Conclusion

In this paper, we presented a review of the security in the software-defined system. The first part describes the different components of SDN architecture with their characteristics. The second part contains the security issues and a list of attacks faced by the elements of SDN. And in the third part, we gave a set of proposed solutions that aim to solve or mitigate the harm of attacks, these solutions are divided according to the three layer: Application layer, the controller layer, and infrastructure layer. DDoS/ DoS attack is one of the most common attacks that target the SDN at different levels (Application layer, the controller layer, and infrastructure layer).

In our future research, we intend to concentrate on the lack of visibility of the SDN state within cloud computing by proposing an approach that measures the security state of the virtual network and provides the appropriate countermeasure in case of an attack.

References

- Khalil, I.M., Khreishah, A., Azeem, M.: Cloud computing security: A survey. Computers 3(1), 1–35 (2014)
- Gong, Y., Huang, W., Wang, W., Lei, Y.: A survey on software defined networking and its applications. Front. Comput. Sci. 9(6), 827–845 (2015)
- 3. Cisco Inc.: Software-defined networking: why we like it and how we are building on it, White Paper (2013)
- McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Turner, J.: OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. 38(2), 69–74 (2008)
- 5. Ahmad, I., Nama, S., Ylianttila, M., Gurtov, A.: Security in software defined networks: A survey. IEEE Commun. Surv. Tutorials **17**(4), 2317–2346 (2015)
- Rawat, D.B., Reddy, S.R.: Software defined networking architecture, security and energy efficiency: A survey. IEEE Commun. Surv. Tutorials 19(1), 1–22 (2016)
- Kemmer, F., Reich, C., Knahl, M., Nathan, C.: Software defined privacy. In: IEEE International Conference on Cloud Engineering Workshop, pp. 25–29 (2016)
- Han, B., Gopalakrishnan, V., Ji, L.S., Lee, S.J.: Network function virtualization: challenges and opportunities for innovations. IEEE Commun. Mag. 53(2), 90–97 (2015)
- Yang, W., Fung, C.: A survey on security in network functions virtualization. In: IEEE NetSoft Conference and Workshops (NetSoft), pp. 15–19 (2016)
- 10. Hu, F., Hao, Q., Bao, K.: A survey on software-defined network and OpenFlow from concept to implementation. IEEE Commun. Surv. Tutorials **16**(4), 2181–2206 (2014)
- Shu, Z., Wan, J., Li, D., Lin, J., Vasilakos, A.V., Imran, M.: Security in software-defined networking: Threats and countermeasures. Mobile Netw. Appl. 21(5), 764–776 (2016)
- 12. Bernardo, D.V.: Software-defined networking and network function virtualization security architecture (2017). https://tools.ietf.org/html/draft-bernardo-sec-arch-sdnnvf-architecture-00
- Namal, S., Ahmad, I., Gurtov, A., Ylianttila, M.: SDN based intertechnology load balancing leveraged by flow admission control. In: IEEE SDN for Future Networks and Services, pp. 1–5 (2013)
- 14. Kreutz, D., Ramos, F.M., Verissimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S.: Software-defined networking: A Comprehensive survey. Proc. IEEE **103**(1), 14–76 (2015)
- 15. Stallings, W.: Software-defined networks and OpenFlow. Internet Protoc. J. 16 (2015)
- Top ten web application vulnerabilities (2017). https://www.owasp.org/index.php/Category: OWASP_Top_Ten_Project
- 17. Green, M., Smith, M.: Developers are not the enemy!: the need for usable security APIs. IEEE Secur. Priv. 14(5), 40–46 (2016)
- Zhang, P., Wang, H., Hu, C., Lin, C.: On denial of service attacks in software defined networks. IEEE Netw. 30(6), 28–33 (2016)
- Brezetz, S.B., Kamga, G.B., Balla, M.N., Criton, T., Jebalia, H.: SDN-based trusted path in a multi-domain network. In: IEEE International Conference on Cloud Engineering Workshop, pp. 19–24 (2016)

- Benton, K., Camp, L.J., Small, C.: OpenFlow vulnerability assessment. In: 2nd ACM SIGCOMM workshop on Hot Topics in Software Defined Networking, pp. 151– 152 (2013)
- Wen, X., Chen, Y., Hu, C., Shi, C., Wang, Y.: Towards a secure controller platform for openflow applications. In: The Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, pp. 171–172 (2013)
- 22. Canini, M., Venzano, D., Peresini, P., Kostic, D., Rexford, J.: A NICE way to test OpenFlow applications. In: The 9th USENIX Conference on Networked Systems Design and Implementation (2012)
- Yu, D., Moore, A.W., Hall, C., Anderson, R.: Authentication for resilience: The case of SDN. In: Security Protocols XXI. Springer, Berlin, pp. 39–44 (2013)
- 24. Security Enhanced (SE) Floodlight (2017). http://www.openflowsec.org/Technologies.html
- 25. Mattos, D.M.F., Ferraz, L.H.G., Duarte, O.C.M.B.: AuthFlow: Authentication and access control mechanism for software defined networking. Univ. Federal Rio Janeiro, Rio de Janeiro, Brazil (2014)
- Ball, T., Bjmer, N., Gember, A., Itzhaky, S., Karbyshev, A., Sagiv, M., Valadarsky, A.: Vericon: towards verifying controller programs in software-defined networks. ACM SIGPLAN Not. 49(6), 282–293 (2014)
- Porras, P., Shin, S., Yegneswaran, V., Fong, M., Tyson, M., Gu, G.: A security enforcement kernel for OpenFlow networks. In: 1st Workshop Hot Topics Software Defined Network, pp. 121–126 (2012)
- Chandrasekaran, B., Benson, T.: Tolerating SDN application failures with LegoSDN. In: Proceedings of the 13th ACM Workshop Hot Topics Network (2014)
- 29. Shin, S., et al.: Rosemary: A robust, secure, and high-performance network operating system. In: ACM Conference on Computer and Communications Security, pp. 78–89 (2014)
- Kazemian, P., Chan, M., Zeng H., Varghese, G., McKeown, N., Whyte, S.: Real time network policy checking using header space analysis. In: USENIX Symposium on Networked Systems Design and Implementation, pp. 99–111 (2013)
- 31. Mai, H., Khurshid, A., Agarwal, R., Caesar, M., Godfrey, P., King, S.: Debugging the data plane with anteater. ACM SIGCOMM Comput. Commun. Rev. **41**(4), 290–301 (2011)
- Son, S., Shin, S., Yegneswaran, V., Porras, P., Gu, G.: Model checking invariant security properties in OpenFlow. In: International Conference on Communications (ICC), pp. 1974– 1979 (2013)
- Shin, S., Porras, P., Yegneswaran, V., Fong, M., Gu, G., Tyson, M.: FRESCO: Modular composable security services for software-defined Networks. In: Network and Distributed Security Symposium, pp. 1–16 (2013)
- Wang, H., Xu, L., Gu, G.: FloodGuard: a dos attack prevention extension in software-defined networks. In: 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 239–250 (2015)
- Suh, J., Choi, H. G., Yoon, W., You, T., Kwon, T., Choi, Y.: Implementation of a content-oriented networking architecture (CONA): a focus on DDoS countermeasure. In: European NetFPGA Developers Workshop (2010)
- Tootoonchian, A., Ganjali, Y.: HyperFlow: a distributed control plane for OpenFlow. In: The 2010 Internet Network Management Conference on Research on Enterprise Networking. USENIX Association, p. 3 (2010)
- Voellmy, A., Wang, J.: Scalable software defined network controllers. In: The ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 289–290 (2012)

- Phemius, K., Bouet, M., Leguay, J.: DISCO: Distributed SDN controllers in a multi-domain environment. In: IEEE Network Operations and Management Symposium (NOMS), pp. 1–4 (2014)
- Yao, G., Bi, J., Xiao, P.: Source address validation solution with OpenFlow/NOX architecture. In: 19th IEEE International Conference on Network Protocols (ICNP), pp. 7–12 (2011)
- Sherwood, R., Gibb, G., Yap, K.K., Appenzeller, G., Casado, M., McKeown, N., Parulkar, G.: Flowvisor: a network virtualization layer. OpenFlow Switch Consortium, Technical Report (2009)
- Khurshid, A., Zhou, W., Caesar, M., Godfrey, P.: Veriflow: verifying network-wide invariants in real time. In: ACM SIGCOMM Computer Communication Review, pp. 467– 472 (2012)
- 42. Scott-Hayward, S., Natarajan, S., Sezer, S.: A survey of security in software defined networks. IEEE Commun. Surv. Tutorials 18(1), 623–654 (2015)

Weakness in Zhang et al.'s Authentication Protocol for Session Initiation Protocol

Mourade Azrour^(IM), Yousef Farhaoui, and Mohammed Ouanan

M2I Laboratory, ASIA Team, Department of Computer Science, Faculty of Sciences and Techniques, Moulay Ismail University, Errachidia, Morocco azrour.mourade@gmail.com, youseffarhaoui@gmail.com, ouanan_mohammed@yahoo.fr

Abstract. Authentication is the most security service required by Session Initiation Protocol (SIP). In recently years, Zhang et al. proposed for the first time an efficient and flexible authentication protocol for SIP using smart card and Elliptic Curve Cryptography. But, in 2014, Zhang et al. showed that their latest proposed protocol is vulnerable to impersonation attack. In order to improve their protocol, Zhang et al. proposed a second protocol. However, in this work we demonstrate that Zhang et al.'s protocol is vulnerable to server spoofing attack. Furthermore to overcome the weakness of Zhang et al.'s protocol we propose an improved and secured SIP authentication and key exchange protocol. The security analysis shows that our proposed protocol can resist to various attack including server spoofing attack.

Keywords: Session Initiation Protocol · Security · Authentication protocol · Elliptic Curve Cryptography · Smart card · Server spoofing attack

1 Introduction

The Telephony over IP (ToIP) is a service that allows to exchange multimedia flows (voice, text, video.) trough internet; ToIP is based on two types of protocols: signaling protocols and transport protocols. In recently decade, Session Initiation Protocol (SIP) [1] is the most signaling protocol used for establishing, altering and terminating session multimedia between different users. The architecture of SIP consists of a proxy server, redirect server, register server, location server, and User agents.

Authentication is the most security service required for SIP. Since, the original SIP authentication protocol (HTTP Digest Authentication [2]) was found vulnerable to deferent attacks; a large community has been participated by proposing the different protocols based on various mechanisms.

SIP authentication protocols proposed before 2013 [3–7] are based on the password verification using several mechanisms. Then, the password must be shared between the user and the server. The shared password is stored in the server database. Therefore, these protocols are vulnerable to stolen verifier attack. In addition to this attack these protocols suffer from the problem of managing of password's database. In 2013, Zhang et al. [8] proposed a first SIP authentication protocol using the Smart Card. Zhang et al.

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_22 have demonstrated that their protocol offers several advantages such as mutual authentication secrecy and password updating, and it is secure against replay attacks, server spoofing attack, stolen verifier attack, man in the middle attack, and offline password guessing attack. However, Wu et al. [9], Tu et al. [10], and Jiang et al. [11] showed that the protocol of Zhang et al. [8] is vulnerable to user impersonation attack. In order to solve this attack, Zhang et al. proposed a year after their second protocol [12]. Tu et al. [10] proposed a new secured SIP authentication protocol. Then, Tu et al. prove that authentication phase of their protocol reduce the computing time cost to 75%compared to the same Zhang et al.'s phase. Despite these advantages, Tu et al.'s protocol is demonstrated vulnerable to many attacks by Farash et al. [13], Mishra et al. [14] and Zhu et al. [15]. Then, Farah et al. [13] proposed their SIP authentication protocol. However, in 2015, Chaudhry et al. [16] demonstrated that the Farash et al.'s protocol is defenseless to replay attack and Denial of Service attack. As result, they proposed a new protocol and they have proven that is secure against known attacks. Also, Kumaris et al. [17] noticed that Farah et al.'s protocol is not secured against different attack. So, Kumari et al. proposed a new protocol that can resist user impersonation attack, Off-line password guessing attack, replay attack and man-in-the-middle.

In 2014, Arshad et al. [18] proposed a new SIP authentication protocol. However, in 2016, Lin et al. [19] have discovered that Arshad et al.'s protocol is vulnerable. To overcome this problem Lin et al. proposed a new protocol that is more secure and allows to users to update their password using a new method.

Recently, M. Azrour et al. [20] proved that Jiang et al.'s protocol suffer from server spoofing attack, in order to enhance the security of SIP, M. Azrour et al. proposed their protocol which is secured against various attack. For more information about related protocol please refer to [21–23].

In this paper, we will analysis the security performance of Zhang et al.'s [12] SIP authentication protocol. We will show that is vulnerable to server spoofing attack. Then, we propose our solution to overcome the weakness in Zhang et al.'s protocol. Performance analysis shows that our protocol is more secured if it is compared with Zhang et al.'s protocol.

The remainder of this paper is organized as follows. Section 2 delivers general information on the original SIP authentication protocol. In the Sect. 3, we briefly reviewed the Zhang et al.'s authentication protocol. Then, this protocol is analyzed in Sect. 4. The Sect. 5 presents our proposed protocol. The security and performance of our proposed protocol are analyzed respectively in Sects. 6 and 7. Finally the Sect. 8 concludes the paper.

2 Original SIP Authentication Protocol

HTTP Digest Authentication for SIP is based on the mechanism challenge/response. Before the protocol execution, the client and the server share the password, which is used to verify the client's identity. The messages exchanged between the server and the



Fig. 1. HTTP digest authentication

client during authentication procedure are illustrated in Fig. 1 and they are described as following.

- Step 1. Client → Server: REQUEST The client sends a REQUEST to the server.
- Step 2. Server → Client: CHALLENGE (nonce, realm) After receiving REQUEST; the server generates CHALLENGE that includes a nonce and the client's realm. Note that realm is used to verify username and password. Then, the server sends back CHALLENGE to the client.
- Step 3. Client → Server: RESPONSE (nonce, realm, username, response) After receiving CHALLENGE from the server, the client computes the response by using received nonce, username, secret password, and realm. response = F(nonce, username, password, realm). Note that F(.) is a one-way hash function. Next, the client sends back the original REQUEST with the computed response, username, nonce and realm.

• Step 4. According to username the server extracts the client's password. Then, the server verifies wither nonce is correct or not. If it is correct, the server computes F (nonce, username, password, realm) and uses it to compare it with the response. If they match, the server authenticates the identity of the client.

3 Review of Zhang et al.'s Scheme

In this section, we briefly review Zhang et al.'s protocol as follows. The notations used in this paper are shown in Table 1.

Notations	Explanations
U	The remote user
S	The remote server
$X \to Y{:}M$	X sends a message M to Y
username	The identity of user U
PW	The password of user U
$E_p(a,b)$	An elliptic curve equation with order n
S	The long-live secret key of server S
$P_{pub} = sP$	The long-live public key of server S
SK	A session key
$h(.), h_1(.), h_2(.)$	Three secure one-way hash functions
Z_q^*	Multiplication group of Zq
	The string concatenation operator
$E_s(\cdot)$	Symmetric key encryption under the key s

Table 1. Notations and their explanations

3.1 System Setup Phase

The server selects $E_p(a, b)$ with the order $n, P \in E_p(a, b)$, it chooses a random number $s \in_R Z_p^*$ as the secret key. Then, it selects two one-way hash functions, $h(.), h_1(.)$. Finally, the server publishes $\{E_p(a, b), P, h(.), h_1(.)\}$ and keeps *s* in secret.

3.2 Registration Phase

In this phase, the user registers on the SIP server through a secure channel. The details of this phase are as follows.

R1: The user U selects his/her *username*, password PW and a random number $a \in_R Z_p^*$. After that, U computes h(PW||a) and sends $\{h(PW||a), username\}$ to the server through a secure channel.

- R2: after receiving the registration information, the server computes $R = \frac{h(PW||a)}{h(username) + s}P$.
- R3: The server stores R into the smart card and issues it to U.
- R4: Upon receiving the card, U stores a in the card. Then, the card contains (R, a).

3.3 Authentication Phase

Whenever the user wants to login into the remote server, he/she performs the following steps.

A1: $U \rightarrow S$: REQUEST (username, V, W) U colorts a render number $h \in \mathbb{Z}^*$ and co

U selects a random number $b \in_R Z_p^*$, and computes V = bR and W = bh (PW||a)P. Next, the card sends a request message REQUEST(username, V, W) to the server.

- A2: $S \rightarrow U$: CHALLENGE(realm, Auth_s, S, r) After receiving the request message, the server S computes W' = h(username) + s)V = (h(username) + s) = bh((PW||a)P. Then, it checks $W \stackrel{?}{=} W'$. If true, the server chooses two random integers c, $r \in_R Z_p^*$, and computes S = cP, SK = ch(username)W' = cbh(PW||a)h(username)P, and $Auth_s = h_1(S||W'||SK||r)$. Next, it sends message CHALLENGE(realm, $Auth_s, S, r)$ to U over a public channel.
- A3: $U \rightarrow S$: *RESPONSE*(*realm*, *Auth_u*) Upon receiving message REQUEST, U computes SK' = bh(PW||a)h(username)S = cbh(PW||a)h(username)P. Then, it checks whether the equation $Auth_s \stackrel{?}{=} h_1(S||W||SK'||r)$ holds. If so, U computes $Auth_u = h_1(S||W||$ SK'||r+1) and sends RESPONSE(realm, Authu) back to the server. Otherwise, it deletes received information and the protocol stops.
- A4: After receiving the RESPONSE message, the server verifies $Auth_u \stackrel{?}{=} h_1(S||W'|| SK||r+1)$. If the message is authenticated, the server sets SK a shared session key with user U. Otherwise, it deletes received information and the protocol stops.

4 Cryptanalysis of Zhang et al.'s Scheme

In this section, we prove that the Server spoofing attack is still effective in Zhang et al.'s protocol. Suppose that A is an attacker. A can eavesdrops the message

REQUEST{username, V, W} transmitted between server S and user U. Then, he/she can execute server spoofing attack. The details of attack are presented as follows.

- Step1. U inputs his/her username and password PW, generates randomly a number $b \in_R Z_p^*$, and computes V = bR and W = bh(PW||a)P. Then it sends a request message REQUEST(username, V, W) to S.
- <u>Step2</u>. \mathcal{A} eavesdrops message REQUEST(username, V, W) and get username, V, W. He/she generates a random number $r \in_R Z_p^*$. Next, he/she get a value of base point P, and puts its value in $S'(S' \leftarrow P)$. Then, he/she computes SK = h(username)W and $Auth'_s = h_1(S'||W||SK||r)$. Next, \mathcal{A} sends message CHALLENGE(realm, $Auth'_s, S', r$) to U.
- Step3. Upon receiving message CHALLENGE (realm, $Auth'_s, S', r$), U computes SK' = bh(PW||a)h(username)S' and verifies if $Auth'_s = h_1(S'||W||SK'||r)$. The user will find true because: $W \leftarrow bh(PW||a)P$ and $S' \leftarrow P$

So

$$\begin{split} SK =& h(username)W \\ =& h(username)bh(PW||a)P \\ =& bh(PW||a)h(username)P \\ =& bh(PW||a)h(username)S' \\ =& SK' \end{split}$$

As result, user U authenticates attacker A and sends to him RESPONSE thinking that he/she communicate with a legal server S.

According to previous analysis, the adversary can easily impersonate identity of server at any time. The user U does not know whether the one he contacts is that the valid server or not. So the adversary can impersonate the server successfully. Therefore, Zhang et al.'s protocol is vulnerable to the server spoofing attack.

5 Our Proposed Protocol

In order to overcome weakness in Zhang et al.'s protocol, we propose an improved and secured authentication and key agreement protocol for SIP. Our protocol consists of four phases, which are system setup phase, registration phase, authentication and key agreement phase, and password changing phase.

5.1 System Setup Phase

In this phase, the server selects an elliptic curve equation $E_p(a, b)$, over a finite field F_q , an additive group G of order p and P a base point generator with order n over equation

 $E_p(a, b)$, *n* is a large prime of height entropy. Then, the server picks a random integer $s \in_R Z_p^*$ as its secrete key. Next, the server chooses three one-way hash functions $h(\cdot)$, $h_1(\cdot)$ and $h_2(\cdot)$. Finally, the server publishes all parameters except its private key *s*, which it is saved secretly.

5.2 Registration Phase

In this phase the user and server S perform the following steps over a secured channel.

- R1: The user U chooses freely his/her *username*, *password* W and a random number $a \in_R Z_p^*$. After that, U computes h(PW||a) and sends $\{h(PW||a), username\}$ to the S.
- R2: After receiving the registration information, the server computes $R = h(PW||a) \oplus h(username||s)P$ Then, the server stores R into the smart card and delivers it to U.
- R3: Upon receiving the card, U stores a in the card. Therefore, user card contains (R, a)

5.3 Authentication and Key Agreement Phase

As illustrated in Fig. 2, whenever U wishes to log into S, he/she have to inserts his/her smart card in card reader and inputs his/her *username* and password *PW*. Next, the following steps will be executed between server S and user U.

- Auth1: $U \to S$: REQUEST(username, V, W, T_1) After inserting the smart card in card reader and inputting the username and password. The user's smart card chooses a random $b \in_R Z_p^*$, and computes V = bR, = h(PW||a), and W = bXP. Then, he/she sends a message REQUEST(*username*, V, W, T_1) to the server over a public channel. T_1 denotes the current timestamp here.
- Auth2: $S \to U$: CHALLENGE(realm, $Auth_s$, S, r, T_3) Upon receiving the request message form U at time T_2 , S verifies validity of $T_2 - T_1 \leq \Delta T$. If it is OK, S computes Y = h(username||s) and $W' = V \oplus Y$. Then, it verifies $W \stackrel{?}{=} W'$. If it holds, the server S picks randomly two integers $c, r \in_R Z_p^*$. Then, it computes $S = cP, K = cW', SK = h_1(W'||r||YP)$ and $Auth_s = h_2(W'||SK|| r||K||S)$. Next, it sends message CHALLENGE(realm, $Auth_s, S, r, T_3$) to U over a public channel.
- Auth3: $U \rightarrow S$: RESPONSE(realm, Auth_u)

Once the user *U* receives the CHALLENGE message form *S* at time T_4 , *U* verifies validity of $T_4 - T_3 \leq \Delta T$. If is not fresh, *U* stops the process. Otherwise, *U* calculates K' = bXS and $SK' = h_1(W||r||(R \oplus X)P)$, and checks $Auth_s \stackrel{?}{=} h_2(W||SK'|| r||K'||S)$. If it is true, the server is authenticated. Then, user U computes $Auth_u = h_2(W||SK'||r+1||K'||S)$ and sends RESPONSE(realm, $Auth_u$) back to server S. Otherwise, it stops the protocol and deletes received and calculated parameters.
User U

Server S

 $b \in \overline{E_R Z_p}^*$ V = bRX = h(PW||a)W = bXPREQUEST(username, V, W, T,) Verify $T_2 - T_3 \leq \Delta T$ Y = h(username||s) $W' = V \oplus Y, W \stackrel{*}{=} W',$ If equation holds, $c \in_R Z_p^*$, $r \in_R Z_p^*$ S = cP, K = cW' $SK = h_1(W' || r || YP)$ $Auth_{s} = h_{2}(W'||SK||r||K||S)$ CHALLENGE (realm, Auth, S, r, T3) Verify $T_4 - T_3 \leq \Delta T$ K' = bXS $SK' = h_1(W || r|| (R \oplus X)P))$ $Auth_{s} \triangleq h_{2} (W || SK' || r || K' || S).$ If equation holds. $Auth_{u} = h_{2} (W || SK' || r + 1 || K' || S)$ RESPONSE (realm.Auth.,) Checks Auth_u \triangleq $h_2(W \parallel SK' \parallel r + 1 \parallel K' \parallel S)$

Fig. 2. Authentication phase of our proposed scheme

– Auth4:

After receiving the RESPONSE message, the server verifies $Auth_u \stackrel{?}{=} h_2(W || SK || r + 1 || K' || S)$. If it holds, the user U is authenticated and server S sets SK a shared session key with U. Otherwise, it stops the protocol and deletes received and calculated parameters.

5.4 Password Changing Phase

When the user U wants to update its password, it needs to agree on a session key with the server via the authentication phase in advance. The details of this phase are described as following.

- Pass1. U → S: (username, e, New_u) The user U chooses its new password PW* and two random integers $a^*, e \in_R Z_p^*$ and computes $h(PW^*||a^*)$ and $tag_u = h(username||e||h(PW^*||a^*))$, it then uses SK to encrypt the new parameters, New_u = E_{KS}(username||e|| $h(PW^*||a^*)||tag_u|$). Next, it sends message (username, e, New_u) to server. Pass2. S → U: (New_s) Upon receiving the information, the server decrypts the message and then checks the validity of the authentication $tag_u \stackrel{?}{=} h(username||e||h(PW^*||a^*))$. If it is valid, the server computes the new secret information $R^* = h(PW^*||a^*) \oplus h(username||s)P$ and $tag_s = h(username||e+1||R^*)$. Then, it sends encryption information New_s = E_{KS}(R^*||tag_s) back to user.
- *Pass3.* The user U decrypts received message and verifies the validity of $tag_s \stackrel{?}{=} h(username||e+1||R^*)$. If it is valid, U stores R^* and a^* in its smart card.

6 Security Analysis

6.1 Mutual Authentication

Mutual authentication means that both the user and server are authenticated to each other within the same protocol. In the proposed scheme the server can authenticate user after receiving REQUEST by checking W, and after receiving RESPONSE by checking $Auth_u$. Upon receiving message CHALLENGE user can authenticate the server by testing validity of $Auth_s$. As result, our protocol provides mutual authentication.

6.2 Session Key Secrecy

In our protocol the session key is computed in this way $SK = h_1(W'||r||YP) = h_1(W||r||(R \oplus X)P))$. Since, *PW*, *a*, and *s* are secret, the session key cannot be calculated by anyone except the server and the client. Therefore, our proposed protocol provides session key secrecy.

6.3 Server Spoofing Attack

Our scheme can resist against server spoofing attack. Suppose that an attacker A wants to impersonate the server and spoof user U, A has to computes $Auth_s$. However, A doesn't have any information about a server secret key s. Then, A can't compute K and SK. Therefore, he cannot forge a valid CHALLENGE message.

6.4 User Impersonation Attack

Assume that attacker \mathcal{A} wishes to connect to the server as legitimate user U. \mathcal{A} has to prove its validity by forging two messages REQUEST(username, V, W, T_1) and RESPONSE(realm, $Auth_u,T_3$). While \mathcal{A} need to know some secret information PW and a. Therefore, he/she is not capable to send the two validate messages. As result, our scheme can resist user impersonation attack.

6.5 Denning-Sacco Attack

In our scheme, the session key is calculated in this way $SK = h_1(W'||r||YP) = h_1(W||r||(R \oplus X)P))$. If an attacker has obtained it, he will have to break the one-way hash function to get *YP* or $(R \oplus X)P$. Then, he has to face Elliptic Curve Cryptography if he wants to guess the user password. So, the proposed scheme is secure against Denning Sacco attack.

6.6 Replay Attack

Suppose that the adversary Alice intercepts the messages REQUEST(username, V, W, T1) and RESPONSE(realm, $Auth_u$) and try to impersonate a legitimate user U. However, she cannot calculate V, W, and $Auth_u$. Since she don't know server secret key. Alice has to face the ECDLP, if she wants to get the correct one by guessing the secret key s from V or W. after replaying REQUEST or RESPONSE the server will detect the attack via comparing if $W \stackrel{?}{=} V \otimes Y$ or $Auth_u \stackrel{?}{=} (W||SK'||r+1||K'||S)$.

Now, Suppose that Alice intercepts the message CHALLENGE(realm, $Auth_s, S, r$) and try to replay it to impersonate the legal server. In order, to be authenticated by the user, Alice have to compute the value of $Auth_s = h_2(W'||SK||r||K||S)$ using secret PW, a, K, and SK. Since Alice don't have any information about secret parameters she cannot computes a valid $Auth_s$. As result, the proposed protocol withstands replay attack.

Attacks	Zhang et al. [8]	Zhang et al. [12]	Tu et al. [10]	Jiang et al. [11]	Ours
Stolen verifier	Yes	Yes	Yes	Yes	Yes
Denning-sacco	Yes	Yes	-	-	Yes
Password guessing	Yes	Yes	Yes	Yes	Yes
Replay	Yes	Yes	Yes	Yes	Yes
Man in the middle	Yes	Yes	No	-	Yes
Server spoofing	No	No	No	No	Yes
Impersonation	No	Yes	No	No	Yes
Mutual authentication	Yes	Yes	Yes	Yes	Yes
Session key secrecy	Yes	Yes	-	Yes	Yes

 Table 2.
 Security performances

Phase		Zhang et al. [12]	Ours
Registration	User	$1T_h$	$1T_h$
	Server	$1T_h + 1T_{pm} + 1T_{inv}$	$1T_h + 1T_{pm}$
Authentication	User	$4T_h + 3T_{pm}$	$4T_h + 3T_{pm}$
	Server	$3T_h + 3T_{pm}$	$4T_h + 3T_{pm}$
Total		$9T_h + 7T_{pm} + 1T_{inv}$	$10T_h + 7T_{pm}$

Table 3. Computational comparisons between our protocol and related protocols

6.7 Stolen Verifier Attack

In the proposed scheme, any user's secret is stored in server database. So, the attacker can't obtain the user's secret information from server. Therefore, our proposed protocol is secure against stolen verifier attack.

6.8 Offline Password Guessing Attack

Suppose that an attacker records all messages (REQUEST, CHALLENGE and RESPONSE) transmitted between user and server, then extract *username*, V, W, *realm*, $Auth_s$, S, r and $Auth_u$, and tries to guess the password PW^* and verifies its correctness. Since, the attacker does not know any information about values of s, a, b, and c. He/she can't compute K, SK. Then, he can't verify the calculated V, W, $Auth_s$ or $Auth_u$.

If attacker steals user card he can get *R* and *a*, However, he must to know *s* to check $h(PW||a) \oplus h(username||s)P$. Therefore, our proposed scheme is safe against password guessing attack.

6.9 Man-in-the-Middle Attack

In our protocol all messages are authenticated by server or user, to know their origin. In addition, at the end of authentication, the session key is shared between user and server, so the following messages will be encrypt using session key. To replay these messages, an attacker needs to know a session key. But, he cannot calculate it since he/she does not know s, a, X, PW and b. As result, our protocol is secure against Man-in-the-middle attack.

7 Performance Comparison

In this section, we will compare the performance and computation cost of our proposed protocol with Zhang et al.'s protocol. In this comparison a very lightweight operations like string concatenation operation, Exclusive-OR operation are not examined, because there computation cost is negligible. The notations used are illustrated as follows.

 T_h The computational cost of one-way hash operation

 T_{pm} The computational cost of elliptic curve point multiplication

 T_{inv} The computational cost of modular inversion

 T_{Eks}/T_{Dks} The computational cost of Encryption/Decryption algorithm

In the registration phase of our protocol the user uses one hash function and the server computes $1T_h + 1T_{pm}$. The computational costs of the user side and server side in our protocol's authentication phase are $4T_h + 3T_{pm}$ and $4T_h + 3T_{pm}$. In the password changing phase the user computes $3T_h + 1T_{EKs} + 1T_{DKs}$ and the server computes $3T_h + 1T_{Fm} + 1T_{EKs} + 1T_{DKs}$.

According to the Table 3 we can observe that modular inversion operation is not used in the registration phase of our protocol. So, this phase is faster than the same phase of Zhang et al.'s protocol. In the Table 2, we can see that our protocol is secured against different attacks especially server spoofing attack, which is effective in the protocol of Zhang et al. Moreover, Zhang et al.'s protocol consist on three phases: System setup phase, Registration phase, and Authentication phase; so it is impossible to change the password or it's not clear how it can be changed. Contrary, our protocol defined Password changing phase in addition to the last three cited phases. Therefore, we can say that our protocol is suitable for applications developed on the base of SIP.

8 Conclusion

In this paper, we have showed that Zhang et al.'s protocol is vulnerable to server spoofing attacks. In order to overcome this weakness we proposed an efficient and secure SIP authentication scheme. According to our analysis, our proposed protocol is secure against various attacks and can provide many security services.

References

- Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: Session Initiation Protocol. RFC 3261 (Proposed Standard). Updated by RFCs 3265, 3853, 4320, 4916, 5393, June 2002
- Franks, J., Hallam-Baker, P., Hostetler, J., Lawrence, S., Leach, P., Luotonen, A., Stewart, L.: HTTP Authentication: Basic and Digest Access Authentication, June 1999
- Durlanik, A., Sogukpinar, I.: SIP authentication scheme using ECDH. World Enformatika Soc. Trans. Eng. Comput. Technol. 8, 350–353 (2005)
- Huang, H., Wei, W., Brown, G.E.: A new efficient authentication scheme for session initiation protocol. In: Proceedings of the 9th Joint Conference on Information Sciences (2006)
- Yoon, E.J., Yoo, K.Y., Kim, C., Hong, Y.S., Jo, M., Chen, H.H.: A secure and efficient SIP authentication scheme for converged VoIP networks. Comput. Commun. 33(14), 1674–1681 (2010)
- Liu, W., Koenig, H.: Cryptanalysis of a SIP authentication scheme. In: 12th IFIP TC6/TC11 International Conference, CMS 2011. Lecture Notes in Computer Science, vol. 7025, pp. 134–143 (2011)
- Xie, Q.: A new authenticated key agreement for session initiation protocol. Int. J. Commun. Syst. 25(1), 47–54 (2012)

- Zhang, L., Tang, S., Cai, Z.: Efficient and flexible password authenticated key agreement for voice over internet protocol session initiation protocol using smart card. Int. J. Commun. Syst. 27(11), 2691–2702 (2013)
- Wu, K., Gong, P., Wang, J., Yan, X., Li, P.: An improved authentication protocol for session initiation protocol using smart card and elliptic curve cryptography. Rom. J. Inf. Sci. Technol. 16(4), 324–335 (2013)
- Tu, H., Kumar, N., Chilamkurti, N., Rho, S.: An improved authentication protocol for session initiation protocol using smart card. Peer-to-Peer Netw. Appl., 1936–6442 (2013). doi:10.1007/s12083-014-0248-4
- Jiang, Q., Ma, J., Tian, Y.: Cryptanalysis of smart-card-based password authenticated key agreement protocol for session initiation protocol of zhang et al. Int. J. Commun. Syst. 28(7), 1340–1351 (2014)
- Zhang, L., Tang, S., Cai, Z.: Cryptanalysis and improvement of password-authenticated key agreement for session initiation protocol using smart cards. Secur. Commun. Netw. 7, 2405– 2411 (2014)
- Farash, M.S.: Security analysis and enhancements of an improved authentication for session initiation protocol with provable security. Peer-to-Peer Netw. Appl. doi:10.1007/s12083-014-0315-x
- Mishra, D., Das, A.K., Mukhopadhyay, S.: A secure and efficient ECC-based user anonymity-preserving session initiation authentication protocol using smart card. Peer-to-Peer Netw. Appl. doi:10.1007/s12083-014-0321-z
- Zhu, W., Chen, J., He, D.: Enhanced authentication protocol for session initiation protocol using smart card. Int. J. Electr. Secur. Digital Forensics 7(40), 330–342 (2015)
- Chaudhry, S.A., Mahmood, K., Naqvi, H., Sher, M.: A secure authentication scheme for session initiation protocol based on elliptic curve cryptography. In: 2015 IEEE International Conference on Computer and Information Technology, Ubiquitous Computing and Communications. Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing (2015)
- Kumari, S., Chaudhry, S.A., Wu, F., Li, X., Farash, M.S., Khan, M.K.: An improved smart card based authentication scheme for session initiation protocol. Peer-to-Peer Netw. Appl. 10, 92–105 (2015)
- Arshad, H., Nikooghadam, M.: An efficient and secure authentication and key agreement scheme for session initiation protocol using ECC. Multimed Tools Appl. 75, 181–197 (2014)
- 19. Lin, H., Wen, F., Du, C.: An anonymous and secure authentication and key agreement scheme for session initiation protocol. Multimed Tools Appl. **76**, 2315–2329 (2016)
- Azrour, M., Farhaoui, Y., Ouanan, M.: A new secure authentication and key exchange protocol for session initiation protocol using smart card. Int. J. Netw. Secur. 19(6), 866–875 (2017). doi:10.6633/IJNS.201711.19(6).2)
- Azrour, M., Ouanan, M., Farhaoui, Y.: SIP authentication protocols based on elliptic curve cryptography: survey and comparison. Indones. J. Electr. Eng. Comput. Sci. 4(1), 231–239 (2016)
- 22. Azrour, M., Farhaoui, Y., Ouanan, M.: Cryptanalysis of Farash et al.'s SIP authentication protocol. Int. J. Dyn. Syst. Differ. Equ. (in press)
- Azrour, M., Farhaoui, Y., Ouanan, M., et al.: A server spoofing attack on Zhang et al. SIP authentication protocol. Int. J. Tomogr. Simul. 30(3), 47–58 (2017)

How Mobile Nodes Influence Wireless Sensor Networks Security and Lifetime

Mohammed Saïd Salah^(⊠), Abderrahim Maizate, Mohamed Ouzzif, and Mohamed Toumi

RITM-ESTC/CED, ENSEM, University Hassan II Casablanca, Casablanca, Morocco salahmedsaid@gmail.com

Abstract. To maintain the proper functioning of critical applications based on Wireless Sensor Networks, we must provide an acceptable level of security while taking into account limited capabilities of the sensors. In this paper we propose a mobile approach to secure data exchanged by structured nodes in cluster. The approach is based on mobile nodes with significant calculation and energy resources that allow cryptographic key management and periodic rekeying. However, mobility in wireless sensor networks aims to increase the security and lifetime of the entire network. The technical methods used in this paper are based on cryptography elliptic curves and key management through a balanced binary tree. To compare the performance of the proposed approach with other mobile algorithms, we focus on the following metrics: the energy consumed by normal sensors and cluster heads, the number of packets exchanged during key installation, time to generate and distribute cryptographic keys and the memory used by the different sensors to store keys.

Keywords: Wireless sensor networks · Mobiles nodes · Key management · Elliptic curve cryptography

Nomenclature

- AVL denotes a binary tree (authors names are: Adelson, Velskij and Landis).
- CH_i denotes a cluster head.
- d_{ii} denotes distance measure from node n_i to n_i .
- $d_{CHi,MN}$ denotes distance between the CH and MN.
- ECC denotes the cryptography based on Elliptic Curves.
- $e(n_i)$ denotes the readiness of node n_i to become a CH.
- f_s small fraction.
- f_r fraction of reduction.
- ID denotes a node identifier.
- k random number of the order of n (1 to n 1).
- KDC denotes a keys distribution center.
- MADR Mobile Algorithm for Key Distribution and Periodic Rekeying.

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_23

- MN denotes the mobile node.
- n_i or n_j denotes a node.
- $N(CH_i)$ denotes the CH list.
- $N(n_i)$ denotes the neighbors of node n_i .
- $Nw(n_i)$ is weight of neighbors.
- ORPHAN denotes that the node has no neighbor.
- P denotes a point of the elliptic curve E.
- S_i is cluster with I + 1 member nodes.
- $-t_w$ denotes waiting time.
- $w(n_i)$ denotes weight of nodes.
- WSN denotes a wireless sensor networks.
- \rightarrow send a message.
- − ← receive a message.

1 Introduction

A wireless sensor network (WSN) is composed of a set of nodes sensors. Each of these nodes has the capacity to collect and transfer data to a base station connected to the user.

Security plays a central role in critical applications of Wireless sensor network [1] such as smart cities applications, public lighting management, military applications, security solutions... during normal operation of the network, data can be threatened by external events something that should not happen. Security insurance [2, 3] is a serious challenge, especially when nodes are made up of electronic devices [4] with limited material capacities.

Two key management approaches are available: symmetric and asymmetric approach [5, 6]. The first focuses on the effectiveness of establishing cryptographic keys after network deployment. However, asymmetric cryptography offers better resistance against compromise nodes, but requires additional costs on the software and hardware nodes. Recently, experts in cryptography make use of techniques based on elliptic curves [7] that can replace integer calculations by calculations in groups associated with elliptic curve, which turns the cryptography based on ECC stronger.

Some approaches based on elliptic curve cryptography [8, 9] have been proposed by the scientific community, yet the limited capacity of captors considerably slow down the life span of the entire network and does not allow any complex calculations [10]. To maintain the proper functioning of critical applications, the key management and rekeying must be periodic after network deployment. Each node must have all necessary keys to communicate with neighbors or with nodes in the same cluster.

The remainder of this paper is organized as follows: in Sect. 2, we presented applications security in Casablanca Smart City. Then, we explained how mobility may

affect the safety and life of the WSN. In Sect. 3, we presented some algorithms used mobility in WSN. In Sect. 4, we discussed the technical methods used in this paper, especially the cryptography based on ECC and the AVL tree [11]. Thereafter we described the proposed approach and its added value by algorithms and explanations of each phase. In the last section, we analyzed the different algorithms and compared the performance of each approach [12, 13] by looking at: the average energy consumed per node, the memory used by node for storing ECC keys, the number of packets exchanged when installing keys and ECC keys computing time and rekeying.

2 Applications Security in Casablanca Smart City

The Casablanca has launched the study to develop a master plan for digital processing and information systems. This study has several phases, including analysis, development of digital development plan as well as the pane of intelligent applications. Then it launched concrete projects and starts their achievements.

Profits of applications based on wireless sensor networks [14] may enhance the management of road traffic; inter-modality between different means of transport, through the traceability of cleaning operations, energy efficiency... the objective is to put the applications for the benefit of citizens [15]. However, attacks can have negative impacts on all applications in city, thus minimizing security in those solutions. So it is important to secure these applications in order to maintain their effectiveness.

Applications will be used daily by the citizens of the city and will help to better manage the city and save energy resources [16]. To secure these applications we proposed nodes attached to the 6 existing tramway lines shown in Fig. 1, to ensure



Fig. 1. Tramay network in Casablanca.

coverage for the entire city (a diameter of 42 km). Each line contains 8 tramways with a 10 min passage frequency. Hence the rekeying frequency remains an organizational choice and depends on the risks to which the sensors are exposed.

Mobile nodes collaborate with each other to ensure full coverage of the city and the rekeying of all applications. Each mobile node by communicating with a cluster head, asks if the rekeying period has expired or not, if the period has expired the mobile node starts rekeying, if not the mobile node moves to the next cluster.

3 Related Works

3.1 Routing Driven Key Management (RECC)

The *RECC* method [12] is based on routing protocol for heterogeneous sensor networks. The main idea is that a node can communicate only with a few neighbors. Before installing keys, the protocol calculates the itinerary tree using a service to evaluate the location of each node, which requires several exchanges of messages between nodes. All these communications are not protected, allowing an attacker to capture and edit these messages. However, the routing protocol does not have the ability to add a new node or to update keys.

3.2 AVL-Headers and AVL-KDC

Network contains a large number of nodes which form clusters around the node elected Cluster-Head (CH). They receive and then transmit nodes messages to the base station. ECC keys generation and storage in the AVL tree are done by the base station while distribution is done either by a KDC server or by CH [13].

In AVL-KDC approach, when a node arrives or leaves the network, the KDC server starts updating public keys located throughout the node path to the base of the AVL tree and reconfigures the AVL tree. Then, it sends an update message to all Headers. Thereafter, each Header broadcasts the message to all nodes in the cluster.

In AVL-Headers, CH starts updating ECC cryptographic keys on the branch of the node that has just arrived or left the AVL tree, it reconfigures and rebalances the AVL tree. Finally, it broadcasts the updates to all nodes in the cluster which depletes its energy resources.

4 Technical Methods Used in the Proposed Approach

4.1 Key Management Based on the AVL Tree

After deployment, sensors need to establish cryptographic keys with their neighbors to communicate securely. However, static management is not appropriate after a long time.

In this paper, we proposed a management of cryptographic keys ECC based on the tree AVL [20] (the name comes from those who created this structure: Adelson, Velskij and Landis). In Fig. 2, the cryptographic keys are represented by circles and the sensor nodes by squares.



Fig. 2. Keys and sensors in the AVL tree

In the AVL tree, each sensor must have all the keys corresponding to the nodes of the AVL tree. In addition, the sensors do not know the keys they do not need. For the balancing factor, each node of the tree, the difference in height of its sub-tree is a maximum 1.

 $h(T_L)$: the height of the subtree in left

 $h(T_R)$: the height of the subtree in right

$$|\mathbf{h}(\mathbf{T}_{\mathrm{L}}) - \mathbf{h}(\mathbf{T}_{\mathrm{R}})| \le 1 \tag{1}$$

This obviously leads to the property that any sub-tree of a balanced binary tree under AVL tree is balanced [20] in the sense AVL.

The main question is to determinate the maximum height of a balanced binary tree AVL. To find out, it is necessary determines the minimum number (NB_h^{min}) of nodes a balanced binary tree AVL of a height (h) should be.

$$NB_0^{min} = 0, NB_1^{min} = 1, NB_2^{min} = 2$$
⁽²⁾

$$NB_{h}^{min} = 1 + NB_{h-1}^{min} + NB_{h-2}^{min} \quad for \ h \ge 2$$
(3)

This recurrence has as solution Fibonacci number (F) to order h + 2 minus 1. So:

$$NB_h^{min} = F_{h+2} - 1 \tag{4}$$

$$NB_{h}^{min} = 1/\sqrt{5} * \left(\left(\left(1 + \sqrt{5} \right)/2 \right)^{h+2} - \left(\left(1 + \sqrt{5} \right)/2 \right)^{h+2} \right) - 1$$
 (5)

$$NB_h^{\min} \approx 1/\sqrt{5} * \left(\left(\left(1 + \sqrt{5} \right)/2 \right)^{h+2} \right) - 1$$
(6)

So conversely the maximum height h of an AVL tree containing n elements is such that:

$$1/\sqrt{5} * \left(\left(\left(1 + \sqrt{5} \right)/2 \right)^{h+2} \right) - 1 \le n$$

$$\tag{7}$$

$$h \le \log_a \left((n+1) * \sqrt{5} \right) - 2 \quad \text{with } a = \left(1 + \sqrt{5} \right) / 2 \tag{8}$$

$$h \le log_a(2) * log_2(n+1) - log_a(\sqrt{5}) - 2$$
 (9)

$$h \le 1,44 * \log_2(n+1) - 0,328 \tag{10}$$

The worst case of complexity of balanced binary trees under AVL is log(n).

4.2 Cryptography Based on Elliptic Curves

Cryptographic systems based on elliptic curves allow to obtain a gain in efficiency in key management. In fact such cryptosystems require a small size [21] of keys (for example, a 160-bit key when RSA use a 1024 bit key, with an equivalent level of security) this represents a significant advantage for wireless sensor networks [22] whose memory space is very limited. In addition, the algorithms of calculations related to elliptic curves are faster, and therefore have an important debit for generate and exchange the keys [23].

The elliptic curve (E) is an algebraic curve [16] that can be represented by the Weierstrass equation:

$$E: y^2 + a_1 x y + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$
(11)

We suppose that the curve is defined in a field (K) and the parameters *a1*, *a2*, *a3*, *a4*, *a6* \in *K*. To obtain a smooth curve with no reversal point, it is necessary that the discriminant of the curve $\Delta \# 0$. The complete definition of Δ is represented in the formula (2).

$$\Delta = -d_2^2 d_8 - 8d_4^3 - 27d_6^2 - 9d_2d_4d_6 \tag{12}$$

Where:
$$d_2 = a_1^2 + 4a_2$$

 $d_4 = 2a_4 + a_1a_3$
 $d_6 = a_3^2 + 4a_6$
 $d_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$

The equation of an elliptic curve can be simplified if the curve is defined on a first field (Fp). Weierstrass equation can be transformed into a simple equation as the following:

$$E: y^2 = x^3 + ax + b$$
(13)

Where: $a, b \in Fp$. Actually, this is the curve's form used in this paper.

5 The Proposed Approach

In this paper we proposed architecture based on mobile nodes with significant capacities of calculations and no energetic constraints as they are fixed to movable empowered structures. Mobile nodes will be responsible for ensuring a good level of security by generating the ECC key [24], store them in the AVL tree and make periodic rekeying. The operations performed by the mobile nodes can minimize the cost of communication between the nodes and the base station which increases the lifetime of the entire network. Moreover, the results estimate how much messages we minimize during installing keys and how fast generating and storing ECC keys. In Sect. 6, we also compared the result with the following implementations specially RECC [12] and AVL-Headers/AVL-KDC [13]. The obtained results proved that the proposed algorithm can outperform other approaches based on ECC.

After deployment, the mobile nodes cooperate to secure all clusters that are on their ways. To maintain the security of the entire network [25], it is necessary to establish a strategy for coordination between mobile nodes and estimate the time intervals between rekeying passages. To evaluate the performance, the mobile approach is implemented as a prototype system on WSNs (Fig. 3).

The power of mobile nodes enables them to have a significant coverage radius and nodes located within that radius can exchange directly with the mobile node. However, nodes outside the cover (out of the radio coverage) may receive the messages from its CH, not directly from MN.

5.1 Assumptions and Requirements

This so-called approach is based on the following assumptions and requirements:

• The mobile nodes are powerful as a workstation capable of sending wide-range radio signals.



Fig. 3. Coverage by two mobile nodes

- Each MN has no constraints on energy, computing and storage capacity.
- Normal nodes have the same capacity in processing, energy and storage.
- The compromise of a node means that all information stored in its memory is known by the attacker.
- Communication channels are bidirectional, i.e., if a node 'x' can receive such a message from another node 'y' is possible likewise.
- An attacker can listen to all traffic; return of old messages, or injected their own messages.

5.2 The Main Phases of the Mobile Approach

Three phases those compose the proposed approach. During the deployment phase, all nodes are preloaded by the same network key (NK) which will later be the network broadcast key as well as it the root of the AVL tree. After the clusters formation and in the setting phase, the mobile nodes calculate their distances with cluster heads and establish the membership list. Thereafter, each cluster head form its members list and send it to the mobile node to which it is attached. Each mobile node generates and stores the ECC keys of its clusters in the AVL tree after; it sends the ECC key for each node of the selected cluster. During rekeying phase, each mobile node calculates its passage cycle based on the number of clusters and nodes per cluster. When an area is covered by several mobile nodes, the CH sends in the arrival of each MN the time elapsed since the last rekeying. If time is longer than the rekeying time, the MN starts the rekeying operations, if not the MN moves to the next cluster.

a - Deployment phase:

We propose that all nodes are randomly deployed and are preloaded with a network key NK to communicate with the CH and neighboring nodes. Then, each node broadcasts a "Hello" message with its ID to all neighbors within its listening range. The nodes in a node's neighborhood respond by sending their own ID. Each node prepares a list of its neighbors $N(n_i)$ from distances $d_{i,j}$ between them. The rest of this phase is described in the following algorithms:

```
Initialization:
  n_i \rightarrow broadcasts "Hello" message with its ID;
  IF no message received THEN n_i is ORPHAN;
             ELSE n; receives messages with ID and:
         collects IDs for received messages;
  ni
      :
             the neighbor set N(n_i);
  n_i computes d_{ij}, for all neighbors n_i \in N(n_i);
      : computes e(n<sub>i</sub>) for node to become CH;
  ni
  n_i \rightarrow (d_{ij}, N(n_i), e(n_i)) to all n_j \in N(n_j);
             computes weight w(n<sub>i</sub>);
  n_i \rightarrow w(n_i) to all n_i \in N(n_i);
             collects Nw(n_i) weights of distance 1 nodes;
Cluster Formulation:
       IF (e(n_i) = 1) and (Nw(n_i) \neq 0) THEN
              IF (w(n_i) \ge w(n_i)) for all n_i \in N(n_i) THEN
  n<sub>i</sub> : initiates a cluster; ELSE
  n_{i} : joins a cluster formed by n_{i} \in N\left(n_{i}\right) ;
         ELSE each node computes |N(n_i)| for all n_i \in
  N(n_i);
         IF |N(n_i)| \ge |N(n_i)| for all n_i \in N(n_i) THEN
  n<sub>i</sub> : initiates a cluster; ELSE
  n; : joins cluster created by n;
  n_i \rightarrow broadcasts invitation ;
  n_i \leftarrow \text{the responses to form } S_i = \{n_i, n_{i1}, n_{i2}, \dots, n_{iT}\};
  n_i \rightarrow Si to all neighbors;
Cluster Reduction:
  IF (w(n<sub>i</sub>) = 0) and (f_s \star |S_i| \leq |N(n_i)|) THEN
```

IF $(w(n_i) = 0)$ and $(I_s^*|S_i| \ge |N(n_i)|)$ THEN Enquire: n_i asks members of S_i of their alternate cluster choices and receives responses;

```
Reduce: If > (f_r * |S_i|) members have choice to join
other cluster then send message to accept invitation;

<u>Cluster Head Selection:</u>

IF |S_i| | N(n_j) | THEN
n_i : is confirmed as a CH; ELSE IF
n_{ij} : can be reduced THEN
create connectivity by reducing n_{ij};
ELSE IF e(n_{ij}) = 1 and (S_i - n_{ij}) \in N(n_j) THEN
n_{ij} : is confirmed as CH;
```

b - Configuration phase:

During the first pass, the mobile nodes establish the membership list from the calculated distances between them and the CH. After each MN_i selects the nearest cluster head and asked it to send its list of members. Once received by the MN, then it begins to generate the cryptographic keys depending on the number of node by cluster. When key generation ends, the MN_i stores the keys in the AVL sub-tree of the cluster. At this moment the MN_i sends the tree branch that contains all ECC keys allowing to each node to communicate with its neighbors and the CH. When sending the key ends, the CH is responsible for sending the ECC keys required for nodes belonging to the cluster and which are outside the coverage of all MN. Hereinafter, the algorithms of this phase.

```
Collecting Information:
  MN \rightarrow message "Hello" to CH<sub>i</sub> with its ID;
  MN \leftarrow messages from CH<sub>i</sub>;
      IF no message received THEN MN falls asleep for tw
      and sends the message "Hello";
      ELSE MN receives messages with ID and an estimate of
      the distance based on the received signal power.
  MN : collects (ID, d<sub>CHi,MN</sub>) for received messages;
        the CH_i set N(CH_i);
Generating ECC Keys:
 MN : computes for each cluster all public and private
     keys required for all nodes, following this
     algorithm:
     k_0=1 and k \leftarrow (k_{n-1}, ..., k_1, k_0) 2
     Set P_1 \leftarrow P, P_2=2P
     For i from i-2 down to 0 do
              If k_i=1 then Set P_1 \leftarrow P_1 + P_2, P_2 \leftarrow 2P_2.
              Else
                              Set P_2 \leftarrow P_2 + P_1, P_1 \leftarrow 2P_1.
     end for
            RETURN ( Q = P_1)
```

```
Insert the keys in AVL tree::
 MN : inserts the keys generated in the AVL tree, T
 denotes the tree pointer and x denotes ECC keys:
      If T=null THEN
                   T:=new tree; T.data:= x; height:=0;
          case
                 T.data = x : return ;
                 T.data > x : return Insert(T.left, x)
       If ((height (T.left) - height (T.right)) = 2) {
             If (T.left.data > x ) THEN
                 T = RotateFromLeft (T);
                     T = DoubleRotateFromLeft (T); }
             Else
                     T.data < x : return Insert (T.right,
                     X)
             Similar to the left case
             Endcase
Send the necessary keys:
 MN: computes the neighbors of each node based on its
 position in the AVL tree.
 MN \rightarrow all keys to communicate with them in the selected
 cluster.
```

c - Rekeying phase:

The need for security is different from one application to another and the risks to which the nodes are exposed should be studied in advance. When an application requests a high level of security, it is recommended to enforce the periodic rekeying T (for example every hour), in order to limit the impact of key compromise on data flows.

Among the advantages of the proposed approach is that a mobile node to renew the cryptographic keys of all clusters during each pass T_i , greatly increasing the security of the entire network. During the ΔT_i interval, the mobile node compares the nodes list during the last pass and the list sent by the CH to check if a node has left or has just arrived in the cluster. Hereinafter, a chronology of rekeying cycles ΔT_i for each cluster and periods of passing T_i for three mobile nodes (Fig. 4).



Fig. 4. Example of rekeying by three mobile nodes

6 Simulation

6.1 Simulation Parameters and Metrics

The different algorithms are implemented with software TOSSIM (TinyOS Simulator) and we used IEEE 802.15.4 as a communication model for the network.

In this part of the paper, we evaluated the performance of the proposed mobile algorithm for distribution and periodic rekeying MADR by comparing simulation with both methods RECC and AVL-KDC/AVL-Headers. Hereinafter, the simulation parameters for a normal node (Table 1).

Parameters	Values
Deployment surface	$(0,0) \times (10000 \text{ m}, 10000 \text{ m})$
Eelec	50 nJ/bit
Eamp	10 pJ/bit/m ²
Initial energy of a normal sensor	10 J
Active mode	0,073 J/s
Sleep mode	0,005 J/s
Data traffic	CBR

Table 1. Simulation parameters of a normal sensor

In this section we present results that correspond exactly to the values obtained by simulations of the studied methods. In order to evaluate each method performance we will focus on the following four metrics:

- Energy consumption per node/cluster head
- ECC Keys computing time and rekeying
- Number of stored keys
- Number of exchanged messages.

6.2 Comparison of Metrics

- Comparison of Energy consumption by sensor

The lifetime of a network depends heavily on the energy consumed by the sensors that compose it. Several studies showed that a large part of energy consumed in a WSN is due to wireless communications and the calculations performed by the processing units.

In the following simulation we focus on the operations performed by the mobile nodes that have no energetic constraint to ensure proper security and increase the lifetime of the entire network. The following graph shows the energy consumed by each method during the distribution of ECC keys (Fig. 5).



Fig. 5. The energy consumption per node

The results show that the RECC method consumes more energy and this because of the large number of packets exchanged during distribution and ECC key installation. By against the power consumed by normal nodes in both approaches AVL-KDC and AVL-Headers does not vary with the size of the network.

We can see after this comparison as mobile nodes and their permanent energy can drastically save energy consumed and that is the strong point of the proposed approach.

- ECC Keys computing time and rekeying

The more the regeneration and distribution of cryptographic keys take time, the more the network becomes vulnerable. When the number of cluster per node is large, the generation will take more time because of complex calculations to achieve to generate the ECC keys. After deployment of the RECC approach, the nodes evaluate their positions in the network and calculate the routing table. All these operations require several message exchanges and more time before installing keys.

A header of AVL-Headers takes at least 15 min to calculate and manage ECC keys and that because of its limited computing capabilities, which puts the entire network at risk. In the proposed approach, the diffusion and rekeying of cryptographic keys are ensured by the mobiles nodes. The significant power calculation allows the mobile node to manage keys in a cluster of 50 nodes in 122 s and rekey the same cluster every 5400 s.

- Number of Stored Keys

The memory used by the nodes and cluster heads to store the cryptographic keys has an important role in the proper functioning of the network. When the used size is large it requires more treatment and more energy is consumed. According to the size of the network, Fig. 6 shows memory used by the normal node to save ECC keys.



Fig. 6. The size of memory used to store the keys by normal node

The cluster head is preloaded with all of its cluster keys (use almost 24000 bytes); which makes the RECC approach the most gourmand in terms of memory. AVL-Headers use memory space that varies between 1200 and 2500 bytes. AVL-KDC and MN use a very limited memory space, because the CH is not loaded with the keys of the nodes in the cluster. The server KDC and the MN are responsible for the AVL tree, which offers to the Cluster Head an economy of storage memory.

The Fig. 6 shows a management methods based on AVL tree, the number of keys stored in the normal node depends only on the height of the AVL tree. Each node has

the keys how he needs to communicate with its neighbors and its Cluster Head. However, the number of keys stored in a normal node in the RECC approach depends on the size of the network, because each node has all the network keys.

Number of exchanged packets

The last metric to discuss in this paper is the number of packets exchanged during distribution and Installation of ECC keys. A large number of packets exchanged implies more energy consumption and more risk of capture the exchanged packets.

In Fig. 7 we compare the number of exchanged packets by the nodes during distribution and rekeying. RECC approach has no renewal mechanism of cryptographic keys that explain its small number comparing to other methods.



Fig. 7. The energy consumption per cluster head

However, the rekeying in the AVL-KDC and AVL-Headers methods are performed after the outbreak of an external event. The MN approach offers a periodic rekeying for all clusters. For a cycle of 5400 s the number of exchanged messages during distribution and rekeying is smaller than the remaining methods.

6.3 Security Analysis

The time to Generate and send cryptographic keys for both methods AVL-Headers and AVL-KDC is very long. Thus, the energy cost of exchanges and calculation has a great impact on the lifetime of the entire network.

A normal node in the RECC approach, communicates with all the nodes to be located in the network, after it calculated its routing table. All these operations take a long time before the generation and distribution of cryptographic keys, which makes applications based on WSN vulnerable to attacks.

We find that the MADR approach has three strong points. The first deals with the lifetime of the network, mobile nodes that have no energy constraints ensure all complex calculation which can deplete the energy of a normal sensor. The second is

security [26]; the rekeying is done periodically which significantly reduces the risk of attacks. The last point is on the network coverage, the significant power of mobile nodes allows a large radius of coverage and even the nodes that are outside coverage of mobile nodes are managed by their cluster heads.

7 Conclusion and Perspectives

To maintain the effectiveness of critical applications based on wireless sensor networks, we must ensure a good level of nodes security taking into account their limited energy and computing.

In this paper, we proposed an approach based on mobile nodes for ECC keys generation, distribution and periodic rekeying of all clusters. The generation is done by the mobile node which stores subsequently the generated ECC keys in an AVL tree before sending keys to each node it needs to communicate with its neighbors and its cluster head.

Collaboration between powerful mobile nodes provides better coverage and a good key management. Due to the significant capabilities of the mobile nodes we can use them to secure critical applications at the same time if needed in applications which require difficult operations.

References

- Othman, M.F., Shazali, K.: Wireless sensor network applications: a study in environment monitoring system. Procedia Eng. 41, 1204–1210 (2012)
- Jain, A., Kant, K., Tripathy, M.R.: Security solutions for wireless sensor networks. In: 2012 Second International Conference on Advanced Computing & Communication Technologies, pp. 430–433. IEEE (2012)
- Pandey, A., Tripathi, R.C.: A survey on wireless sensor networks security. Int. J. Comput. Appl. 3(2), 43–49 (2010)
- Adnan, A., Hanapi, Z.: Geographic routing protocols for wireless sensor networks: design and security perspectives. Int. J. Commun. Antenna Propag. (IRECAP) 5(4), 197–211 (2015)
- Chandra, S., Paira, S., Alam, S.: A comparative survey of symmetric and asymmetric key cryptography. In: International Conference on Electronics, Communication and Computational Engineering (ICECCE), pp. 83–93. IEEE (2014)
- Sasi, S., Dixon, D., Wilson, J.: A general comparison of symmetric and asymmetric cryptosystems for WSNs and an overview of location based encryption technique for improving security. IOSR J. Eng. 4(3), 1 (2014)
- Kandasamy, R., Krishnan, S.: Enhanced energy efficient method for WSN to prevent far-zone. Int. J. Commun. Antenna Propag. (IRECAP) 4(4), 137–142 (2014)
- Munivel, E., Ajit, M.: Efficient public key infrastructure implementation in wireless sensor networks. In: International Conference on Wireless Communication and Sensor Computing, February 2010
- Zhang, Y., Yang, W., Kim, K., Park, M.: An AVL tree-based dynamic key management in hierarchical wireless sensor network. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing, August 2008

- Setiadi, I., Kistijantoro, A., Miyaji, A.: Elliptic curve cryptography: algorithms and implementation analysis over coordinate systems. In: Proceedings of the 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA) 2015, pp. 1–6. IEEE (2015)
- 11. Qin, Z., Zhang, X., Feng, K.: An efficient key management scheme based on ECC and AVL tree for large scale wireless sensor networks. Int. J. Distrib. Sens. Netw. **2015**, 198 (2015)
- Du, X., Guizani, M., Xiao, Y., Chen, H.: A routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks. In: IEEE International Conference on Communications, ICC 2007, Glasgow, August 2007
- Boumerzoug, H., Boucif, A.: A lightweight key management scheme based on an AVL tree and ECC cryptography for wireless sensor networks. Concurr. Comput. Pract. Exp. 28(6), 1831–1847 (2016)
- Schleicher, J.M., Vögler, M., Dustdar, S., Inzinger, C.: Enabling a smart city application ecosystem: requirements and architectural aspects. IEEE Internet Comput. 20(2), 58–65 (2016)
- 15. Walravens, N.: Mobile city applications for Brussels citizens: smart city trends, challenges and a reality check. Telematics Inform. **32**(2), 282–299 (2015)
- Jalali, R., El-Khatib, K., McGregor, C.: Smart city architecture for community level services through the internet of things. In: 2015 Proceedings of the 18th International Conference on Intelligence in Next Generation Networks (ICIN), pp. 108–113. IEEE (2015)
- 17. Cai, W., Chen, M., Hara, T., Shu, L., Kwon, T.: A genetic algorithm approach to multi-agent itinerary planning in wireless sensor networks. Mob. Netw. Appl. **16**(6), 782–793 (2011)
- Hao, C.H.U., Cheng-dong, W.U.: A Kalman framework based mobile node localization in rough environment using wireless sensor network. Int. J. Distrib. Sens. Netw. 2015, 73 (2015)
- 19. Anderson, R., Seitz, S.: CSE 326: Data Structures AVL Trees (2014)
- Salah, M.S., Maizate, A., Ouzzif, M.: Security approaches based on elliptic curve cryptography in wireless sensor networks. In: Proceedings of the 27th International Conference on Microelectronics (ICM), pp. 35–38. IEEE (2015)
- Gura, N., Patel, A., Wander, A.: Comparing elliptic curve cryptography and RSA on 8-bit CPUs. In: International Workshop on Cryptographic Hardware and Embedded Systems, pp. 119–132. Springer, Heidelberg (2004)
- 22. Enge, A.: Elliptic Curves and Their Applications to Cryptography: An Introduction. Springer Science & Business Media, Boston (2012)
- Anitha, R., Nawaz, G.: Development of a secure, energy efficient and reliable routing protocol for mobile wireless sensor networks. Int. Rev. Comput. Softw. (IRECOS) 9(3), 487–494 (2014)
- Kazienko, J.F., Moraes, I.M., Albuquerque, C.V.: On the performance of a secure storage mechanism for key distribution architectures in wireless sensor networks. Int. J. Distrib. Sens. Netw. 2015, 44 (2015)
- Tsou, Y.T., Lu, C.S., Kuo, S.Y.: MoteSec-aware: a practical secure mechanism for wireless sensor networks. IEEE Trans. Wirel. Commun. 12(6), 2817–2829 (2013)
- Geetha, R., Kannan, E.: Secure communication against framing attack in wireless sensor network. Int. Rev. Comput. Softw. (IRECOS) 10(4), 393–398 (2015)

A Novel Smart Distribution System for an Islanded Region

Youssef Hamdaoui^(运) and Abdelilah Maach

Department of Computer Science, Mohammedia Engineering School (EMI), Mohammed V University, Rabat, Morocco youssefhamdaoui@research.emi.ac.ma, maach@emi.ac.ma

Abstract. Region outage is one of distribution system and most of time caused by weather or a fault in transmission lines. Smart grid is a new concept that can be a solution for few people existing in a far region to avoid islanded problem. Thanks to IT evolution and the integration of the renewable resources in the existing electrical grid and the bi-directional communication ensured in distribution network, we can distribute power smartly to response all demands or response the emergencies demands. In this paper we propose a potential outage management and a new approach of distribution based on some collected region parameters and based on the potential of distributed generation like solar panel, wind turbine and storage batteries. Islanding region can be detected with some existing islanding detection methods. Smart loads selection is done through an intelligent controller who collects information from different controllers and sensors. The regional data center analyzes and decides a layered tree that contains consumers who will be covered by the outage operation. An algorithm is developed to implement our approach with some distribution simulation to discuss and analyze results.

Keywords: Distribution network · Outage · Loads · Efficiency · Demands · Smart region · Distribution planning · Storage batteries

1 Introduction

The concept of Smart Grid (SG) [1–3] has comprehensively overhauled the scene of existing electric grid and the integration of prosumers, where an entity can consume and produce simultaneously extemporized in a complete paradigm shift [4]. Electricity is generated by power plants and moves through a complex system of electricity substations, transformers, and power lines that connect electricity producers and end users. Transmission and distribution systems require regular investment in transmission to maintain high degree of uninterrupted power system service quality, reliability for users and the safety [5]. The Distribution system provides major opportunities for smart grid because of its important role of distributing electrical power from the main grid to users, it has the potential to supply electricity based on renewable resources during grid outages. The evolution of IT technologies and through a fast wired/wireless communication, the smart grid can detect, collect real time information, analyze and take a smart decision to manage the network and control consumption and production.

SG integrates advanced technologies and supports two way information flows, it is possible for a smart grid to enable demand-side management [6, 7] in a residential area, also flexibility of tailoring costumer energy demand.

A Smart grid presents the new concept of the existing electrical grid

- Intelligent- capable of sensing system overloads, bi-directional communication and rerouting power
- Efficient- capable of meeting increased consumer demand without adding infrastructure
- Accommodating
 – accepting energy from any resources including solar and wind as
 easily and transparently and capable of integrating any ideas and technologies like
 energy storage technologies
- Motivating- enabling real-time communication between the consumer and utility so consumers can tailor their energy consumption based on individual preferences, like price and/or environmental concerns
- Resilient increasingly resistant to attack and natural disasters as it becomes more decentralized. Resiliency is very important because is required to improve reliability and reliability refers to the ability of a power system to provide an adequate power to the load at any time and any context
- "Green" slowing the advance of global climate change and offering a genuine path toward significant environmental improvement

Since the society becomes increasingly concerned the efficiency, saving energy and preserve the environment, the interest toward the distributed generation systems, such as solar panel, wind turbines and ocean resource, increases year after year. But generally Distributed generations (DG) will have affects in the network that one of these influences is an islanding phenomenon. Islanding or an outage is the situation in which part of the power grid, consisting of load and DG is isolated from the main utility grid. The duration of interruptions might be from minutes to hours depending on the severity of the fault that occurred and the costs of an outage are considerable and not recoverable (outage Costs). The major power outage data shows major systematic causes are mostly related to weather and to infrastructure [8] or when the demand in the peak hours exceeds the supply [9, 10], but weather caused 80% of all outages (Kenward and Raja 2014). The integration of distributed energy resources (DER) can be a solution for outage problem through the distribution network and also are green resources and allow more monitoring and control and flexibility management.

In Outage case, the best way to manage the electrical distribution grid is to break the distribution system down into small clusters or micro-grids with coordinated control of multi micro grids [11]. We suppose the islanded region like a micro-grid power by his local resources. Microgrid has to detect the outage and automatically continue to provide a constant power to the islanded region [12].

Prosumers or the end user who can produce power is an important actors in the new concept of SG [4], because the prosumers have the ability to inject his extra power to the grid or to charge electrical vehicles (EV) when his storage batteries are full charged. In our approach we integrate the prosumers [13] as resources and we compare his consumption with the existing energy to make decision and to add it to the selection tree. In This paper, we propose a solution to islanding problem through a smart

selection algorithm that focus on distribution network based on local DER and the prosumers existing in the islanded area. In the first and second section we present a review of outage causes, islanding detection methods and distribution system based on DER. In the third section we propose our distribution network approach in an islanded smart region with a layered tree selection model; the planning should cover the majority of local costumers and response first to emergency loads. In Sect. 4, we present the technical part with simulation and discussion of simulation results.

2 Islanding and Distributed Generations

2.1 Power Outage

Islanding occurs when a section of the distribution system containing DGs is disconnected from the main utility grid, while the DGs continue to supply fully or partially the load in the isolated section, termed as island. Most of outage are caused by weather as shown in Fig. 1.



Fig. 1. Major weather-related power outages—those affecting \geq 50,000 customers—increased dramatically in the 2000 s (Kenward and Raja 2014)

The major power outage are Snow storms, summer and winter storms, high winds with lightning, heavy rains, hurricane, heat wave, maintenance problems, deteriorating infrastructure, substation failures and transmission lines failures. We should rethink our concept to include autonomous micro-grid and new distribution concepts with DGs to provide a constant power to the islanded zone load. An energy outage costs for both the customers and the utilities [14]. Economic losses are called 'outage costs' [15, 16].

2.2 Islanding Detection Methods

Outage detection is achieved by using islanding detection methods (IDM) [17–23]. Generally, IDMs are divided into local and remote methods communication based and is more expensive. Methods are based on measurement of some parameters or variables

on the MG side (frequency and voltage).Local methods include passive methods, active methods and hybrid method.

- Passive IDMs are based on measuring a local parameter-index and comparing it with a preset value. So, they detect possible sags of frequency and voltage at the point of common coupling (PCC), voltage phase jump or total harmonic distortion (THD) indices, that serve as indicators of the grid operational mode. In general, the main drawback of passive IDMs is the large NDZ especially in cases where the Microgrid production is matched with the load [17–23].
- Active IDMs impose additional perturbation signals to cause power mismatches, so that a certain system parameter drifts, once islanding occurs. This may include phase shifting, active power variation, reactive power variation or harmonic power variation [17–23].
- A hybrid method is an active and passive combination, used for complex systems and can improve multiple performance indices but Islanding detection time is prolonged [17].

2.3 Islanding Distributed Energy Resources

The major roadblock of a DER-based distribution is system complexities and technical difficulties of managing and the main benefits are

- Improving reliability
- Allow integration of renewable resources
- Dynamic islanding
- Distributed control
- Improving efficiency
- Environmental protection
- Reduce distance between generation and loads
- Minimize distribution and transmission problem
- Less loss and costs
- Response to emergency

DERs are very important for environment first and all consumers that have an emergencies. The Utilization of renewable energy is different for each outage case, it depends on location, time and the resources exist in this area, so a combination between DERs become the best way to power islanded regions. And to maintain and improve the power quality, additionally, renewable energy sources can maximize their production efficiency when paired with an energy storage system. Table 1 present some Characteristics of renewable energy technologies.

The majority of studies confirm that using a single renewable energy generation, such as solar or wind is difficult to provide a continuous power all the time [25] and the advantages in our approach is the integration of prosumers, for example PV and wind turbine depends on weather and can have sometimes a low generation. A hybrid solar-battery system can provide 100% of power supply for end users, thus greatly decreasing the energy costs and increasing the reliability of power supply [26].

		Table 1.	Characteristics of ren	lewable energy t	echnologies [2	4]		
Type of	Hydropower:	Hydropower:	Solar PV:	Solar PV:	Solar	Geothermal	Wind:	Wind:
electricity	grid based	off grid/rural	ground-mounted	rooftop	thermal:	power	onshore	Small
generation			utility-scale		domestic			scale
technology					hot water			turbine
					systems			
Plant size	1-18,000 MW	0.1–1000 kW	2.5-250 MW	3-5 kW	7-10	1-100 MW	1.5-3.5 MW	<100 kW
				(residential)	kWth			
					(single			
					family)			
Capital costs	750-4000	1175-6000	1200-1950	2150-7000	147–2200	1900-5500	925-1950	6040
(\$/kW)								(SU)
Typical	2–23	5-40	9-40 (non	28-55	1.5–28	4–19	4-16	15-20
energy costs			OECD)	(non	(China)			
(cent/kWh)				OECD)				

techn
energy
renewable
of
Characteristics
Η.
Table

3 Proposed Approach

3.1 Smart Region Model

When an outage is detected, the islanded region becomes a Microgrid with a regional data center (RDC) who control and monitor the state of power. A smart region is supposed contain some local resources like a PV/WIND farm with storage batteries. The main controller is the regional data center and its connected to all entities in Micro-grid as shown in Fig. 2, and his role is to maintain a high degree of power quality and a continuous power to all demands or if we don't have enough energy, we have to response to the emergencies loads and our approach is for this case of response when the loads exceeds the power existing. Each costumer in the region has a controller to control smart meters sensors and actuators inside house and periodically information are transmitted to the main controller to update costumer's information. RDC become like cyber physical systems that use physical information, analyze, plan and make actions for execution.



Fig. 2. Micro-grid model

Our approach is to define the critical costumers that need a uninterrupted energy and define it as spanning tree, like hospitals or house contain people with a health problem, the idea is when the main grid is down, we lunch the emergency process who collect data from a deferent controllers to collect information and start our selection method based on this real information to define the selected emergencies. The advantage in our method is that it is dynamic and not static and can change each outage case because the context and state of the energy in the region in not static. The smart selection method and smart distribution is based on the deferent DERs installed in the micro grid as principal sources with a householder's energy as second resources based on his decision to inject or not into the electrical grid.

3.2 Costumer Parameters

We define some decisive parameters that our selection will use for islanded operation like geographic position, uninterrupted load, emergency classification, the costumer production and his decision to inject into the grid or be in off main grid use. Table 2 presents the parameters details.

Parameters	Description
(X,Y)	Costumer Geographical coordinates.
С	Emergency classification.
Р	Uninterrupted load
UseCpFlag	To know if the costumer want inject his extra power or not (on/off electrical
	grid use)
Ср	Energy produced by costumer renewable resource.

 Table 2.
 Parameters description

3.3 Emergency Classification

It's very important parameters because we classified all entities connected to the micro grid into three categories

- C1: Very high emergency and especially for health machine load.
- C2: Medium emergency like small industries to not have an important production loses.
- C3: low emergency level like a house loads with a normal appliances.

3.4 Layered Tree Selection Algorithm

When a outage is detected and all information for our islanded operation are collected and we collect also storage batteries level to know the energy available for distribution. The main controller start analyze for making decision, The main controller have the possibility to switch connected controllers to 'on' or 'off use' of the electrical grid. A layered tree selection is an independent vectors that contain at first the entity with high level of emergency and second the entities with medium level of emergencies and the end will be the low emergency entities.

The first step in our approach is to select the C1 entities and for each one we start build a new tree based on this entity, so we will have number of trees or vectors equal to number of entities with C1 Classification.

If we take for example the first tree T1 that contain the entity E1 classified as C1, we look for neighbors but in square mode, if we find an entity neighbors with C2 classification we control his P if it can be covered by the existing energy, if the condition is verified we layered the entity to E1. We update the existing energy with P = initial(P) - P(E2) and we search for next neighbor with C2 classification and we do the same things until finish with C2 classification. We return to the E1 entity and we start scanning but this time we search for C3 classification.

Figure 3 is just an example of results and present a example of 4 Sub Microgrid (Layered Trees) selected by the islanded operation based on costumers parameters and local DER. Figure 3(a) present the electrical load model with different loads and DGs installed and after lunching the islanded operation we have as result 4 independents trees that will be covered during outage. Each Microgrid (Mi in the Fig. 3(b)) have one entity with high level of emergency and others are a combination between C2 and C3 classification. And the selection is based on algorithm steps.



Fig. 3. Islanded operation progress

The classification can change because we collect data every day from deferent entities through sensors and controllers connected to RDC to have dynamic classification. It's an advantage to have an adaptive algorithm to the context and the off grid area. When an islanding case happens, intentional islanding have to make the supply of emergency load to reduce outage costs and restore the general loads power as much as possible, the condition is the total loads not exceed the generation capacity of DGs plus the costumer extra energy injected. The algorithm is described below and by steps:

- 1. Compute the sum of P existing P(total) = P(DERs).
- 2. Identify the special entity with high level of emergency E (C1).
 - Verify if the P(total) > P(E(C1)), if ok add entity to the selection list and update flag to '1' to avoid double selection.
 - Update P(total) = P(total)-P(E(C1)).
 - Repeat action 2 until defines all C1 entities and defines all vectors (trees).
- 3. Loop for each tree and do
- 4. Grow scanned area on square mode
- 5. Search for neighbors entities with C2 classification in the next scanning area then if exist we control if the P(total) > P(E(C2)) or P(total) > P(E(C2)-Cp if the entity allow injection of his self-production 'useCp = 1'), if ok then we add entity to the tree, for example if our loop is in Tree1 we should add this entity to this tree to have a independent trees at the end of outage operation, and we update P after covering this entity.
- 6. We repeat 4 and 5 actions until finish all C2 demands.
- 7. Return to the first C1 entity for each tree and start searching with the same concept for the entities with C3 classification with growing area each time found and add entities.
- 8. The program exit when covering all demand or depletion of existing energy.

As result of this dynamic algorithm, we will have a vectors (one or more) with different entities that will be covered in our outage operation.

4 Simulation and Results

The implementation is done in ECLIPSE environment with JAVA language and a MYSQL.

4.1 Case of Study and Discussion Simulation Results

As shown in Table 3, we limited simulation to 10 entities with a dynamic classification generation with a random function. We suppose that we have as P(DER) = 200 Kwh. The test system adopts reactive local compensation, so all power loss is ignored. To be in a dynamic context, we use the random function to generate a different classification.

Entity	Coordinate	P (Kwh)	CP (kwh)	Classification	useCpFlag
E(1)	(10,1)	10	15	C3	1
E(2)	(5,1)	10	9	C3	1
E(3)	(6,6)	50	10	C2	0
E(4)	(3,5)	10	6	C3	0
E(5)	(6,5)	50	25	C2	1
E(6)	(4,2)	10	15	C3	0
E(7)	(2,9)	10	3	C3	0
E(8)	(8,9)	10	2	C3	0
E(9)	(8,4)	20	8	C3	0
E(10)	(7,3)	100	0	C1	0

Table 3. Load characteristics

Following the steps cited in the layered tress algorithm, the first step is to search for all entities with C1 classification, in or simulation we have just one, so we will have one vector and one tree with a E10 as spanning tree and based on localization of E10 we start search C2 neighbors with growth of scanning area on square mode, in the next growth we found E9 but E9 is C3 classification so we ignore for the moments and we continue looking for C2, we find E5 and we verify his P, CP and his useCpFlag, we have P = 50, Cp = 25 and UseCpFlag is 1 so we have to allow the difference P-Cp = 25Kwh and we add it to the selected tree layered to E10, we update the P and search for next we find E3 and we verify his parameters, E3 don't inject then we have to cover his full P, we don't have others C2 then we return to E10 and start looking for C3 entities with the same concept, we find E9 and the P existing is after covering E10 and E5 and E3 is P = 200-100-25-50 = 25, the P(E9) = 20 with a fag null then we have still energy then we cover also E9 and we stop here because we don't have enough energy to cover more. The vectors results is V = [E10,E5,E3,E9].the simulation is done for a simultaneous tree and for different context.

The simulation is available to not only one entity with C1 classification but to manage a lot of separated selection with the same algorithm and a result is dynamic and we control the selected entity to not be selected twice by the flag in database.

5 Conclusion

The idea of the distribution network in outage condition is to increase consumer service quality, propose a solution to resolve shortage outage problem in a regular intervals, to response the maximum of demand and have a continuous power supply during outage duration by taking decisions that depends on the context parameters. The objective is to make order in the distribution system to response emergencies loads first and creates some maturity through integration of prosumers power using and smartness in the region controlled by a regional data center. Smart grid interacts with smart devices, sensors and smart end user who need be able to control his consumption and identify exactly his daily energy needing. The prosumers is a very important actor on the smart grid and his integration is a key for the success of the new smart grid concept. Our approach is not enough optimized and need be compared to others approach to validate the effectiveness and his adaptation for different real context.

References

- Misra, S., Krishna, P.V., Saritha, V., Obaidat, M.S.: Learning automata as a utility for power management in smart grids. IEEE Commun. Mag. 51, 98–104 (2013)
- 2. Fang, X., Misra, S., Xue, G., Yang, D.: The new and improved power grid: A survey. IEEE Commun. Surv. Tutorials (2012)
- 3. Gellings, C.W.: The Smart Grid: Enabling Energy Efficiency and Demand Response. The Fairmont Press, Inc., Lilburn (2009)
- Gillani, S., Laforest, F., Picard, G.: A generic ontology for prosumer-oriented smart grid. In: EDBT/ICDT Workshops, pp. 134–139 (2014)
- Alepuz, S., Busquets-Monge, S., Bordonau, J., Martinez-Velasco, J.A., Silva, C.A., Pontt, J., Rodriguez, J.: Control strategies based on symmetrical components for grid-connected converters under voltage dips. IEEE Trans. Ind. Electron. 56(6), 2162–2173 (2009)
- Alizadeh, M., Wang, A., Scaglione, A.: Demand side management trends in the power grid. In: International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 141–144 (2011)
- Conejo, A.J., Morales, J.M., Baringo, L.: Real-time demand response model. IEEE Trans. Smart Grid 1(3), 236–242 (2010)
- Sundar, D.J., Kumaran, M.S.: A comparative review of islanding detection schemes in distributed generation systems. Int. J. Renew. Energy Res. 5(4), 1016–1023 (2015)
- Jayaweera, D., Galloway, S., Burt, G., McDonald, J.R.: A sampling approach for intentional islanding of distributed generation. IEEE Trans. Power Syst. 22(2), 514–521 (2007)
- Fuangfoo, P., Meenual, T., Lee, W.J., Chompoo-inwai, C.: PEA guidelines for impact study and operation of DG for islanding operation. IEEE Trans. Ind. Appl. 44(5), 1348–1353 (2008)
- Hamdaoui, Y., Maach, A.: A smart approach for intentional islanding based on dynamic selection algorithm in microgrid with distributed generation. In: 2017 International Conference on Big Data, Cloud and Application (BDCA), pp. 1–7. ACM (2017)
- Haque, M.E., Negnevitsky, M., Muttaqi, K.M.: A novel control strategy for a variable-speed wind turbine with a permanent-magnet synchronous generator. IEEE Trans. Ind. Appl. 46(1), 331–339 (2010)

- 13. Hamdaoui, Y., Maach, A.: Prosumers Integration and the Hybrid Communication in Smart Grid Context. Networked Systems. LNCS, vol. 9466. Springer, Cham (2015)
- Küfeoğlu, S., Lehtonen, M.: Interruption costs of service sector electricity customers, a hybrid approach. Int. J. Electr. Power Energy Syst. 64, 588–595 (2015)
- Baarsma, B.E., Hop, J.P.: Pricing power outages in the Netherlands. Energy 34(9), 1378– 1386 (2009)
- 16. Modarres, M.: Power outage planning. Eur. J. Oper. Res. 49(2), 254-265 (1990)
- Hamdaoui, Y., Maach, A.: Smart islanding in smart grids. In: 2016 IEEE Smart Energy Grid Engineering (SEGE), pp. 175–180. IEEE (2016)
- Li, C., Cao, C., Cao, Y., Kuang, Y., Zeng, L., Fang, C.: A review of islanding detection methods for microgrid. Renew. Sustain. Energy Rev. 35, 211–220 (2014)
- Papadimitriou, C.N., Kleftakis, V.A., Hatziargyriou, N.D.: A novel islanding detection method for microgrids based on variable impedance insertion. Electr. Power Syst. Res. 121, 58–66 (2015)
- Teoh, W.Y., Tan, C.W.: An overview of islanding detection methods in photovoltaic systems In: International Conference of World Academy of Science, Engineering and Technology, WASET 2011, Bali (INDONESIA), pp. 674–682 (2011)
- Hamdaoui, Y., Maach, A.: Dynamic balancing of powers in islanded microgrid using distributed energy resources and prosumers for efficient energy management. In: 2017 IEEE Smart Energy Grid Engineering (SEGE). IEEE (2017)
- Kyritsis, A., Papanikolaou, N., Tselepis, S., Christodoulou, C.: Islanding detection methods for distributed PV systems overview and experimental study. In: Karampelas, P., Ekonomou, L., (eds.) Electricity Distribution, pp. 63–79. Springer, Heidelberg (2016)
- Renewable Energy Policy Network for the 21st century (REN21). Renewables 2014 global status report (2014). http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/ GSR2014_full%20report_low%20res.pdf
- Hamdaoui, Y., Maach, A.: An intelligent islanding selection algorithm for optimizing the distribution network based on emergency classification. In: Wireless Technologies, Embedded and Intelligent Systems (WITS), pp. 1–7. IEEE (2017)
- Li, C., Cao, C., Cao, Y., Kuang, Y., Zeng, L., Fang, B.: A review of islanding detection methods for microgrid. Renew. Sustain. Energy Rev. 35, 211–220 (2014)
- Ma, T., Yang, H.X., Lin, L.: A feasibility study of a stand-alone hybrid solar-wind battery system for a remote island. Appl. Energy 121, 149–158 (2014)

Taxonomy of Routing Protocols in MANETs

Younes Ben Chigra^(K), Abderrahim Ghadi, and Mohamed Bouhorma

Computing, Systems and Telecommunications Laboratory, UAE-FSTT, Tangier, Morocco younesbenchigra@gmail.com, Ghadi05@gmail.com, bouhorma@gmail.com

Abstract. Mobile Ad Hoc network (MANET) is a collection of smart mobile nodes, which form a dynamic and autonomous system. These nodes communicate wirelessly in a self-organized, self-configured and self-administered manner. Routing protocol is the main building block in route establishment and traffic delivery, which must be accomplished anywhere and anytime, between a pair of source and destination. Therefore, research interest in MANETs has been growing, and particularly the design of MANET routing protocols has gained a lot of interest. Furthermore, constantly changing network topology, limited bandwidth and energy issues make the task of routing in MANETs a challenging one. In this paper we provide the taxonomy of routing protocols for MANETs, which constitute the main key behind the design of routing protocols process.

Keywords: Mobile Ad Hoc Networks (MANET) \cdot Wired routing protocols \cdot Wireless routing protocols \cdot Taxonomy

1 Introduction

Nowadays, the development of mobile technology applications such as web browsing, online banking, online gaming and social media, has stimulated the wide spread usage of wireless network. Therefore, wireless networks have become almost a necessity and a vital component of contemporary daily life.

In telecommunication field, networks can be classified in two categories. The first category is wired networks, which rely on physical links such as wires and optical fibers. The second category is wireless networks, which use radio transmission techniques to establish links between nodes. In addition, wireless networks can be split into two classes, as presented in Fig. 1, Infrastructure based wireless networks, which use fixed access points as gateways between wired and wireless area. For example, cellular networks (2G, 3G, and LTE), WiFi (IEEE 802.11), WiMax (IEEE 802.16) and infrastructure less networks broadly known as Ad Hoc networks do not rely on any pre-established infrastructure consequently Ad Hoc networks are self-organized, self-configured and self-administered. Furthermore, Ad Hoc Networks are single-hop like Bluetooth or multi-hop like Wireless Sensor Networks (WSN), Wireless Mesh Network (WMN) and Mobile Ad Hoc Network (MANET). In the following paragraphs will focus on routing in Mobile Ad Hoc Network [1].

Ad Hoc networking is a multi-level issue because of its autonomous operations; hence network layer should adapt its routing operations to several network constraints,



Fig. 1. Network categories.

such as nodes mobility, nodes Energy, scarce bandwidth and network size to establish efficient paths for data communication. In this context, many routing protocols has been designed in order to deal with different constraints and guarantee the quality of service required by mobile Ad Hoc network applications [2].

In the flowing, we will present a brief history and different issues in the design of mobile Ad Hoc routing protocols. In Sect. 3, we are going to present taxonomy of routing protocol in both wired network and Mobile Ad Hoc network in order to give a deep understanding of different strategies used in designing routing protocols and gain the proper knowledge about MANET routing protocols.

2 History and Challenges

2.1 Brief History

Back to 1972, the first Ad Hoc network generation called: Packet Radio Network (PRNET) was used by, The Defense Advanced Research Project Agency (DARPA), as trial to provide networking facilities in combat environment. Mainly, PRNET uses the combination of Areal Location of Hazardous Atmospheres (ALOHA) and Carrier Sense Multiple Access (CSMA) for multiple access and distance vector routing.

The second generation emerged in 1980, when PRNET was integrated into the Survival Adaptive Radio Networks (SURAN) project. This provided an infrastructure less packet switched network to the mobile battlefield. SURAN significantly improved upon the radios by making them smaller, cheaper, with scalability of algorithms, and more resilience to electronic attacks.

In the 1990s, Mobile devices like laptops, notebooks, PDAs and Software development became widely available for an easy interconnection of computers. At that time, commercial need for mobile interconnection has triggered the interest of several companies; therefore, the Institute of Electrical and Electronics Engineers (IEEE) established a subcommittee, IEEE802.11, to standardize technologies to be used for wireless Local Area Networks WLANs. Since the subcommittee was established involving experts to address the need of both infrastructure based and infrastructure less based communications [1–3].
2.2 Networking Challenges

Mobile Ad Hoc network is a set of smart mobile nodes, which form a dynamic and autonomous system. In mobile Ad Hoc networks, each node within the network has the ability to change its location and configure itself on the fly. These nodes are able to establish routes, anywhere and anytime, between a pair of source and destination using routing protocols. In addition, routes are generally multi-hop due to limited transmission range of mobile nodes, so routing protocols must be able to route data packets through intermediate nodes until reaching targeted destination [4].

In addition, nodes mobility, bandwidth-constrained, energy-constrained, and limited security of shared medium of Mobile Ad Hoc network make designing process of routing protocols most important and difficult instead of design process in fixed network. Due to the dynamic nature of mobile nodes, MANET experiences frequent link failures, which cause frequent network topology change, this has led to the design of various routing protocols. The purpose of each protocol is to solve problems for a specific MANET topology condition; therefore, the designer should have a prior knowledge about the condition or the context of the network targeted with routing protocol design. Therefore, different routing protocols perform differently with different networks' conditions such as level of mobility, size of the network in terms of connected nodes number or type of packets being routed through the network.

3 Taxonomy of Routing Protocols

The main function of routing in networking world is to make possible the transfer of information and communication between two parties whether in wired or wireless networks.

3.1 Routing in Wired Network

There are two main and commonly used routing protocols for wired networks namely: distance vector and link state algorithms as shown in Fig. 2.



Fig. 2. Wired Routing

3.1.1 Distance Victor Routing

In distance victor routing (DVR), which is commonly known as Bellman Ford Algorithm, route computing between the source and destination is accomplished using different metrics such as number of hop, queue length and delay. Each router maintains a routing table or vector indicating the best known distance to each destination and which route to use to get there. Every neighboring router exchanges the necessary information with each other to keep these tables updated. The main disadvantages of Distance Vector routing protocol are loops and count-to-infinity problems. The best known DVR protocol is Routing Information Protocol (RIP) which solves the issue of count to infinity by introducing the maximum hop count of 16 and looping issue by Time to Live (TTL) [5].

3.1.2 Link State Routing

Link state routing (LSR) has been developed to overcome DVR drawback. It use Dijkstra's algorithm to calculate the shortest path. Routing operation consist of a periodic flooding of link state information through the network to update the current status of links. In case of any topology change a notification will be flooded throughout the entire network to re-compute new routes and update topology information. The main routing protocols in this category are Open Short path First (OSPF) and Intermediate System to Intermediate System (ISIS) [5].

3.2 Routing in Wireless Ad Hoc Network

The main building bloc in Ad Hoc networking is routing. Therefore, designing routing protocols have attracted the interest of researchers. Since several routing protocols have been proposed in order to meet required functionalities related to a specific application field. As a result, there is no routing protocol that could fit to all Ad Hoc networking contexts [6, 7]. Routing protocols can be classified using several approaches, depending of the purpose or the goal for which the protocol is designed. See illustration in Fig. 3. There are different criteria for classifying routing protocols in ad hoc networks:

- Communication Model
- Network structure
- Scheduling model
- State Information
- Route establishment
- Type of Cast
- Type of path



Fig. 3. Routing classification

3.3 Communication Model

Routing protocols can be designed to work on different wireless communication schemes. Wireless communication models can be single channel or multi-channel [8]. Firstly, single channel schemes were designed for Medium Access Control (MAC) to address physical layer deficiencies and provide reliable information to upper layers, these schemes suffer from hidden and exposed terminal problems, as illustrated in Fig. 4, fairness and power consumption issues related to radio communication in mobile ad hoc networks. Most of designed protocols have partially solved the intrinsic problems of wireless communication. On the other hand, multichannel schemes have shown better capabilities to handle the hidden and exposed terminal problems due to the usage of



Fig. 4. Hidden and exposed terminal problems

more than one channel in their network. Multichannel protocols are generally used in Time Division Multiple Access (TDMA) or Carrier Sense Multiple Access (CSMA) based networks. In contrast, single channel are Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA) scheme based.

3.4 Network Structure

Depending on how nodes participate in routing task, routing protocols can be flat or hierarchical [6, 7, 9]. In flat protocols, all nodes have the same responsibilities in the entire routing process. Hence, routing control messages are globally managed in a uniform manner, which cause scalability problem in large networks. For example, Destination sequenced Distance Victor (DSDV) and Ad Hoc On demand Distance Victor (AODV). In hierarchical protocols, the main concern is reducing routing control messages in order to scale to large networks. For example, Zone-based Hierarchical Link State (ZHLS). Thus, nodes are dynamically organized into clusters. In hierarchical scheme, only cluster head nodes know the topology information. Other nodes just send data to these cluster head nodes, which in turn will execute the entire routing process such as finding optimal path to the destination.

3.5 Scheduling Model

In the literature, most routing protocol classification in Ad Hoc network are based on their route establishment and maintenance strategies [2, 10–12]. Routing protocol is considered proactive or table-driven if nodes maintain route information all the time to all destinations. Thus, when a source node has data to send it looks up its routing information data base and start immediately sending data to the destination, which guaranty lower transmission delay. The drawback of proactive routing is periodic routing table updates, which cause routing overhead problem. In reactive or on demand routing protocols, like AODV, route information is acquired and maintained only when a source has data traffic to send. Generally, reactive protocols have two main operations: the first operation is route discovery, which is initiated when a source needs a route to a destination. The second operation is route maintenance, which consists of repairing failed links through active route due to topology changes.

3.6 State Information

Routing protocols may be described in terms of the state information acquired at each node. Two main categories are distinguished: topology based protocols and destination based protocols that are broadly used in traditional wired routing protocols. In topology-based protocols, nodes maintain global view of network topology [13]. This approach is known as link State, where link sate packets (LSP) are exchanged between the whole network's nodes. Every node constructs and maintains a global network topology from the LSPs it receives, and computes the best routes to all other nodes using Dijkstra's algorithm. In destination based protocols nodes do not maintain large scale topology information. The main destination based protocols are Distance victor where every node

periodically exchanges distance vector with its neighbors. When a node receives distance vector information from its neighbors, it computes new routes and updates its distance vector database. The complete path then established, in a distributed scheme, by combining the next hop of nodes on the path from the source to the destination. Distance vector routing protocols have less computational complexity and message overhead.

3.7 Route Establishment

Routing protocols can be distinguished according the way a data packet is forwarded from the source to the destination. There are two approaches [1, 2]: First, source routing protocols, such as Dynamic Source Routing (DSR), which place the entire route information in the packet header then intermediate nodes only forward the packet according to route information stored in the header. In this approach intermediate nodes do not need to compute and maintain updated routing information, as a result much less time is needed for traffic delivery and much less control traffic is generated. However, Source routing do not scale very well in large network and dynamic topology, especially when the route is long, data packet header become large and consume too much of scarce bandwidth. Second approach is hop by hop, which use next hop information stored at each node involved into an active path, like OLSR. Thus, when a node receive a packet, it lookup the routing table and forward the packet to the next hop. The advantage of this strategy is that routes are adaptable to the dynamically changing environment. The drawback of hop-by-hop routing is that each intermediate node has to maintain routing information for each active route and each node may require being aware of their surrounding neighbors through the use of beaconing messages.

3.8 Type of Cast

Routing protocols can be classified depending on their type of cast. For example: OLSR, AODV and DSDV. Unicast Routing Protocol is the most developed for MANET applications. In the unicast routing one separate copy is sent to each receiver from the source node. Thus, data packet is replicated at the source node and then delivered to each destination node; see Fig. 5a. Unicast process consumes more much bandwidth due to redundant data packets. Multicast routing protocol, like on demand Multicast Routing Protocol (OMRP), has become very important in multimedia communications. To send simultaneously the same data packet to multiple receivers, the simplest way is broadcasting. However, broadcast consumes considerable bandwidth and power, which should be avoided as much as possible in mobile Ad Hoc networks due to scarce bandwidth and limited nodes' energy. In Multicast process, the network replicates data packet instead of replicating by source node in Unicast process, which lead to optimal use of scarce bandwidth. See Fig. 5b.



Fig. 5. Types of cast

Another type of routing protocols is geo-cast routing protocols illustrated in Fig. 5c. This routing scheme has been adopted in VANET routing; it consists of sending data packet to a set of nodes inside a specific geographical area [13, 14].

3.9 Type of Path

Some routing protocols are able to find multiple paths to a destination, like Multipath Ad hoc on demand Distance Victor protocol (AODVM), which make routing efficient in case of frequent links break due to nodes mobility. In contrast, others routing protocols are simple and find only one path to a destination. Single path routing protocols should re-compute new route each time a link failure is detected which become more complicated in highly dynamic environment [15].

4 Conclusion

In this paper, we highlight different challenges that researchers are facing in designing Mobile Ad Hoc routing protocols, such as high node mobility and hence dynamic topology, restricted bandwidth and limited energy. Furthermore, we present taxonomy of routing protocols to understand different strategies that could be flowed to develop a routing protocol according to specific network context. In next phase, we are going to review common routing protocols and discuss their advantages and disadvantages in a highly dynamic network.

References

- Loo, J., Mauri Lloret, J., Hamilton Ortiz, J.: Mobile ad hoc networks: current status and future trends, pp. 19–33. CRC Press (2012)
- 2. Sarkar, S.K., Basavaraju, T.G., Puttamadappa, C.: Ad Hoc Mobile Wireless Networks Principles, Protocols, and Applications, 2nd edn., pp. 1–158. CRC Press (2013)
- Reddy, G.R.M., Kiran, M.: Mobile Ad Hoc Networks Bio-Inspired Quality of Service Aware Routing Protocols, pp. 15–36. CRC Press (2016)
- Siva Ram Murthy, C., Manoj, B.S.: Ad Hoc Wireless Networks Architectures and Protocols, pp. 206–227. Prentice Hall (2004)
- 5. Majdi, M.S.: Intelligent Mobile Ad Hoc network management system. Thesis, Brunel University, London, October 2016
- 6. Mohandas, G., Silas, S., Sam, S.: Survey on routing protocols on mobile adhoc networks. IEEE (2013). 978-1-4673-5090-7/13
- Chandni, A.B., Sharma, K.: Qualitative evaluation of routing protocols of MANET in wireless sensor network. Int. J. Comput. Commun. Technol. 38–42 (2013)
- Nagaraja, A., Mangathayaru, N., Rajashekar, N., Kumar, T.S.: A survey on routing techniques for transmission of packets in networks. IEEE (2016). 978-1-5090-5579-1/16
- 9. Pang, K.L., Qin, Y.: The comparison study if flat and hierarchical routing in AdHocwireless networks. 0-7803-97460/06/2006 IEEE
- Dorronsoro, B., Ruiz, P., Danoy, G., Pigné, Y., Bouvry, P.: Evolutionary Algorithms for Mobile Ad Hoc Networks, pp. 3–98. Wiley (2014)
- Chen, J.T., Boreli, R., Sivaraman, V.: Improving the efficiency of anonymous routing for MANETs. Comput. Commun. 35, 619–627 (2012)
- Patel, P.V., Kadhiwala, B.: Broadcasting techniques for route discovery in Mobile Adhoc Network - a survey. IEEE (2016). 978-9-3805-4421-2/16
- 13. Ashoor, A.: Performance analysis between distance vector algorithm (DVA) & link state algorithm (LSA) for routing network. IJSTER 4(02), 101–105 (2015)
- Singh, P.: Comparative study between unicast and Multicast Routing Protocols in different data rates using vanet. In: International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) (2014)
- Manohari, P.K., Ray, N.: Multipath routing protocols in MANETs: a study. In: 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH) (2016)

Cloud, Parallel, Distributed and High Performance Computing

Allocation Strategy for Cloud Datacenter Based on Multi Agent and CP Approach

Merzoug Soltane^{1(SC)}, Kazar Okba², Ezziyyani Mostafa³, and Dardour Makhlouf⁴

 ¹ University of El_Oued LINFI, El-Oued, Algeria merzoug-soltane@univ-eloued.dz
 ² University of Biskra, Biskra, Algeria kazarokba@yahoo.fr
 ³ University of Abdelmalek Essaâdi IBDD-TC, Tangier, Morocco ezziyyani@gmail.com
 ⁴ University of Tebessa LRS, Tebessa, Algeria m.derdour@yahoo.fr

Abstract. The massive diffusion of Cloud services in the internet led to increasing demands of services and cloud infrastructure are conduct to an increasing in energy consumption in data centers, which has presents interesting great challenge. In this paper, we propose new strategy of allocation resources, with intelligent management of resources our approach based on constraint programming (CP), according to the autonomous resources allocation using multi-agent systems (SMA). At first, we make a review of various solutions that have been proposed in order to optimize Cloud energy consumption; second, we offer a logical solution to manage physical and virtual resources in smarter data center. At the end, we conclude this paper by future work.

Keywords: Cloud computing \cdot SMA \cdot Multi agent system \cdot Allocation resources \cdot Resources management \cdot Energy consumption \cdot Smart data center

1 Introduction

A cloud computing it is one of the latest technology used via the Internet, the major advantage of cloud service is the elasticity that allows a dynamic change in the number of resources based on the varying demand from a customer as well as a pay-as-you-go opportunity, both of which can lead to substantial savings for the customers. Appropriate management of resources in clouds is essential for effectively harnessing the power of the underlying distributed resources and infrastructure. The problems range from handling resource heterogeneity, allocating resources to user requests efficiently as well as effectively scheduling the requests that are mapped to given resource, as well as handling uncertainties associated with the workload and the system. As a researcher, one can understand the opportunities to dig further to carry-on with more innovations to contribute better solutions to the existing problems. On other hand, the remarkable increasing demands on cloud services increasing expansion of datacenter to huge datacenter that led to a rise in energy consumption by datacenter the create negative factors affect the environment. According to (Barroso 2013) [1], a typical data center needs 10

megawatts of power to operate. In 2011, the energy consumption of Datacenters is expected to exceed 100000000.00 kilowatt per hour (kWh). By 2013, the Datacenter in the US consumed 91 billion of electricity (kilowatt per hour); equivalent to the annual output of 500 MW, costing US businesses 13\$ billion per year for electricity bills and generating nearly 100 million tons of carbon pollution per year.

The purpose of our research is to serve a smart configuration of the cloud infrastructure based on energy consumption; it mainly has a scientific interesting in the field of cloud to show the concept of smart management of resources in cloud datacenter and its importance on minimizing energy consumption, and maximizing the benefits of cloud service provider. In this paper, we propose new allocation strategy for cloud data center based on multi agent system and constraint programming. At first, a management system design of cloud infrastructure is put forward based on allocation of resource and energy consumption constraint. After that, we involve a multi-agent system to shift traditional resources management systems to autonomous resources management systems.

The rest parts of this paper are organized as follows. In Sect. 2, we present the resource management in Cloud Computing and related works. In Sect. 3, describes our cloud architecture with integration multi agent system. Section 4 describes our Allocation strategy based constraint programming. While the Sect. 5, last section concludes our paper and presents some perspectives.

2 Resource Management in Cloud Computing

When discussing resources, a difference should be made between data-center resources such as available servers, storage space and network bandwidth and computing resources directly available to mobile devices. At data-center level there are multiple tiers for resource allocation and optimization: at cluster or supercomputer, virtual machine and operation system disk image levels. Moreover, for any of those levels different objectives can be pursued: increasing power usage efficiency, increasing or insuring a predefined level of availability for provided services, increasing performance, lowering the data-center air conditioning costs or a combination of them [2] (Fig. 1).

Resource allocation is the process of distributing available resources between the various applications running in a cloud environment. There are several problems addressed by an optimal resource allocation [3]:

- Resource contention: multiple consumers are trying to gain access to the same resources at the same time.
- Scarcity of resources: there are limited resources. Different cloud implementations
 must cope with different levels of resource scarcity. The problem is of bigger importance for private clouds.
- Resource fragmentation: the resources are isolated. Even when there are enough resources, their potential consumers cannot gain access to them.
- Over-provisioning: resources are reserved for a client's exclusive use in quantities exceeding its needs.
- Under-provisioning: opposed to over provisioning, not enough resources are reserved for exclusive use of a client.



Fig. 1. Intelligent architecture for cloud system

3 Literature Review

This section presents our interested research about cloud data center architecture and allocation algorithms, according to the literature, the power consumption has attracts a lot of research s in the past few of years. Whereas, Liu et al. [4], present Green Cloud architecture, which aims to reduce data center power consumption while guaranteeing performance from a user's perspective. Using the recommendations developed in its open-source Cloud standards' incubator. Tian et al. [5], proposed a dynamic and integrated load balancing algorithm for resource scheduling in Cloud data centers. Fumiko Satoh et al. [6], also focus on reducing the usage of energy in data centers. But for the future energy management they develop an energy management System for cloud by the use of sensor management function with an optimized VM allocation tool. This system will help to reduce the energy consumption in multiple data centers and results shows that it will save 30% of energy. This system also used to reduce the energy in carbon emissions. Rasoul et al. [7], proposed software architecture that calculates the energy consumption in datacenter and provide services to the users which uses energy efficiently. Beloglazov et al. [8], focuses on virtual machine for the reduction of the energy consumption. An author proposed the dynamic reallocation technique for VMs and toggles off the unused servers which results, considerable energy saving in the real Cloud Computing data centers. Hulkury et al. [9], used The Green Broker uses these directories and chooses the green offer and energy efficiency information and allocates the services to the private cloud. And finally give the result to the users. Garg et al. [10],

proposed a new architecture called as integrated green Cloud architecture (IGCA). It smartly includes client oriented in the Cloud Middleware that verifies the cloud computing is better than the local computing with QoS and budget (Fig. 2).



Fig. 2. Communication between agents

4 Cloud Datacenter Architecture

This section, we will explain the combination of our multi-agent system under the Cloud environment. Our SMA is dedicated, where 3 agents: Analyzer agent, agent scheduling, Controller agent. The figure above, showed the role of each agent in different layers of Cloud Datacenter. Indeed, the proposed architecture consists of three layers of the Cloud: the application layer, the network layer and the Datacenter layer. In which the figure expresses the relationship between the Cloud layers and the function of each agent in each layer.

If the user of service cloud select service and send request to our system, our system in time ti launches analyzer agent (AA) that last analyze user request and identify all resource needed like (CPU speed, RAM capacity, disk storage ...), when this agent finished her job send anther request to scheduling agent (SA), this agent depends on knowledge base to select and allocate resource, so at first, to select server for final allocation, this agent needed to create or update their knowledge base to select the appropriate servers for allocation, so to create knowledge base SA agent launched research process for available servers that satisfies the service request, the SA agent sends msg to each AC agent hosts in server Si, Each AC agent hosts in a server must be following all operation on their server and send msg contain all the information about resources status in this server to the agent SA, The agent SA receive all msg of each AC agent, from this msg the SA agent build their knowledge base and applies our allocation strategist to determine the final server from the list of appropriate servers.

4.1 Agents Description

This subsection includes a brief description of agent's components that are used for an efficient cloud management and the resources allocation. Some properties must be considered to adequately perform the cooperative activity Autonomy: An agent is said to be autonomous if it is able to make decisions for performing additional actions, or for changing its current task. Perception: the capacity of an agent to perceive its environment and consequently to update its mental state. Auto-Organization: an agent is able to auto-organize himself, when it has the capability to evaluate his interactions with others and add organizational links or remove some of them.

Analyzer Agent (AA)

An analyzer agent examines and extracts specifications from client requests to identify the major parameters: CPU speed, RAM, storage capacity, etc. Subsequently, these specifications will be communicated to the network agent. Figure 3 shows the architecture of the analyzer agent, on which the input is for client requests an acquiring, while output interface is designed to release the aforementioned specifications to the SA agent.



Fig. 3. Analyzer agent configurations

Scheduling Agent SA

To achieve an efficient allocation, the scheduling agent figures prominently position when synchronization processing and task performance presents its main purpose. Along with planning, these specifications of the allocation must comply with constraints defined by the allowance algorithm in order to maintain both energy consumption and the execution cost. Likewise, the SA agent collaborates with other agents in which planning decisions are endowed. The Fig. 4 shows a schematic representation of the SA agent. Where it has two communication interfaces, including external interface to communicate with the AC agent, while the external interface to communicate with the AA agent.



Fig. 4. Scheduling agent configurations.

Controller Agent CA

This agent hosts in all work server in datacenter, his is intended to trigger monitoring method when resource availability statements are requested. For providing monitoring data based on the arrival a request, the AC agent provides a communication interface that carries monitoring method outcomes to the SA agent. Figures 5 and 6 shows a schematic representation of the CA agent. Whose internal external enables the communication with the SA agent, while the internal interface ensures communication and controlling and monitoring all resources in server Si.



Fig. 5. Controller agent configurations.



Fig. 6. Show interaction between agents

5 Strategy of Allocation

1. Problem description

The hardware architecture consists of a set $P = \{p1,..., pk,..., pm\}$ of m identical processors with fixed memory capacity mk and identical processing speed. They are all connected to a network with bandwidth δ .

A task T_i is defined through its temporal characteristics and resource needs: its period Ti (as a task is periodically activated), Tasks are periodically activated in an independent way, and they read and write data at the beginning and the end of their execution.

Finally, each processor is scheduled with a fixed priority strategy. A priority P_i is given to each task.

An allocation is a mapping A: $T \leftarrow \leftarrow \leftarrow \leftarrow \leftarrow \leftarrow P \leftarrow$ such that the image of a task Ti_w is a processor p_k :

$$Ti \rightarrow A(Ti) = p_k$$

The allocation problem consists in finding the mapping A which respects the whole set of constraints described in the immediate below. There are three classes of constraints the allocation problem must respect: timing, resource, and allocation constraints.

• Allocation constraints:

This set of constraints deal with the position (or relative position) of the tasks on the processors. Some tasks require specific processor characteristics to be executed (signal processor, compression processors, databases, etc.) and can only reside on a subset of the available processors. Others must not be put together on the same processor. Two sets of tasks may also have to be disjoint in any assignment [11].

Resource constraints:

The memory usage of a processor cannot exceed a fixed capacity [10].

• Timing constraints:

A hard real-time system must respect all timing constraints to assure the security of the process. Temporal constraints define a schedulable allocation according to deadlines or due dates requirements [11].

An allocation is said to be valid if it satisfies allocation and resource constraints. It is schedulable if it satisfies timing constraints. Finally, a solution to our problem is a valid and schedulable allocation of the tasks.

2. Our solution for Problem of allocation

Constraint programming (CP) techniques have been widely used to solve a large range of combinatorial problems. A *constraint satisfaction problem* (CSP) consists of a set $V\psi$ of variables defined by a corresponding set $D\psi$ of possible values (the so-called *domain*) and a set $C\psi$ of constraints. A solution to the problem is an assignment of a value in $D\psi$ to each variable in $V\psi$ such that all constraints are satisfied. This mechanism coupled with a backtracking scheme allows the search space to be explored in a *complete way*. For a deeper introduction on CP, we refer to [12].

We propose an approach based on multi agent system and integrate constraint programming, for this, we suppose a data center D consists of a group of physical servers S; wherein each server $Si \in S$ characterized by a number of processors (nbCpu); RAM capacity (RamC); storage unit (Disk); limited bandwidth(Net BW); a set of virtual machines (Vij). At virtual level, for each virtual machine $vj \in Vij$ includes virtual resources such as nb VCpu; RAMs and Disks.

Nevertheless, to achieve to an effective management, we associate these parameters by precondition equations as the following:

The Cpu allocation for the VMs must not be exceeded the entire number of physical Cpu on the server, as shown on the Eq. 1.

$$\sum_{j=1}^{n} (vj.nbVCpu) \le (S_i.nbCpu)$$
(1)

On only one server, memory utilization rate by the VMs must not exceed the total rate of available RAM capacity on this server

$$\sum_{j=1}^{n} (vj.RamC) \le (Si.RamC)$$
(2)

The utilization rate of storage space by VMs on the server must not exceed the available memory space on the hard disk.

$$\sum_{j=1}^{V_j} (vj.VDisk) \le (Si.Disk)$$
(3)

The number of Vji allocated by the server Si must not be exceeded approved VMs allowed by this server.

$$V_{ij} \le \left(S_i.\mathrm{max}\mathrm{V}_j\right) \tag{4}$$

```
Step 1: inputs initialization
  Nbr_ PM: number of physique server in the data center;
  Nbr_ VM: number of virtual server in the data center;
  LV : list of running VMs in each server on data center;
  LV.V : list of available VMs in each server on data center;
  LS.V : list of available servers in data centers;
L: list of available server Ready for VM (Ri) allocation and Table of
all tuple M;
  M: The constraint to selecting server for allocation VM ;
  Free (): searching free resource in Server Si
  AC_Request: denotes a function that extracts the allocation request
                     specifications such as the identifying needed
                     resources;
Step 2: Analyzed Request
  If user send their request (user.request)
  AA agent (AC_Request)
  AC_Request= Ri;
  AC_Request(Ri);
  Send(Ri)to SA agent ;
  }
Step 3: VM Allocation
  If (SA Receive Request (Ri)) SA launched process to create their
knowledge base
       {
  Send msg (Ri) (Discovery available resources in datacenter for Ri
request)
       }
  If (CA Receive msg (R;)) launched search process
  Nbr_ PM();
Nbr_ VM ();
  LV();
  LV.V ();
  LS.V ();
  Send L(Ri)to SA agent
  }
  If (AC.free ()not null)
  {
  Sand msg (AC.Free) to SA agent SA agent update knowledge base
  3
  If (SA Receive L(R_i)) receive list of available S_i ready to allocate
R<sub>i</sub> request
  If there is a list of servers suitable for the allocation then
  For each S<sub>i</sub> in Selected SerList
   {
  PiSi = Calculate power consumption (R;);
  Min Energy Consumption(P_iS_j)
  Allocate a VM with lowest energy consumption.
  Else
  There is no ready server for allocation;
```

6 Conclusions

In this paper, we introduce to the smarter cloud resource management, where it mainly based on multi-agent system and constraint programming (CP). Before allocation, we were proposing a precondition constraint aimed select a ready server for the request allocation. Whereas, energy consumption criteria has been selected to identify a preferment server in order to satisfy the user request. As key feature, we involve 3 agents for autonomous cloud resources configuration and improve the allocation policies. In the future work, we attempt to simulate a real scenario by utilizing OMNET++ simulator and iCanCloud framework.

References

- Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The stanford digital library metadata architecture. Int. J. Digit. Libr. 1, 108–121 (1997)
- Vinothina, V., Sridaran, R., Ganapathi, P.: A survey on resource allocation strategies in cloud computing. Int. J. Adv. Comput. Sci. Appl. 3(6), 97–104 (2012)
- Ionescu, A.: Resource management in mobile cloud computing. Informatica Economică 19(1), 55 (2015)
- Liu, L., Wang, H., Liu, X., Jin, X., He, W.B., Wang, Q.B., et al.: GreenCloud: a new architecture for green data center. In: Proceedings of the Sixth International Conference Industry Session on Autonomic Computing and Communications Industry Session, ICAC-INDST 2009, New York, pp. 29–38. ACM (2009)
- Tian, W.H., Zhao, Y., Zhong, Y.L., Xu, M.X., Jing, C.: A dynamic and integrated load balancing scheduling algorithm for cloud data centers. In: The Proceedings of CCIS 2011, Beijing (2011)
- Satoh, F., Yanagisawa, H., Takahashi, H., Kushida, T.: Total energy management system for cloud computing. In: Satoh, F., Yanagisawa, H., Takahashi, H., Kushida, T., (eds.) Proceedings of the IEEE International Conference of the Cloud Engineering (IC2E), March 25–27, Redwood City, CA (2013)
- Beik, R.: Green cloud computing: an energy-aware layer in software architecture. In: Beik, R., (ed.) Proceedings of the Spring Congress of the Engineering and Technology (S-CET), May 27–30, Xian (2012)
- Beloglazov, A., Buyya, R.: Energy efficient allocation of virtual machines in cloud data centres. In: Beloglazov, A., Buyya, R., (eds.) Proceedings of the 10th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid), May 17–20, Melbourne, Australia (2010)
- Hulkury, M.N., Doomun, M.R.: Integrated green cloud computing architecture. In: Hulkury, M.N., Doomun, M.R., (eds.) Proceedings of the International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Washington DC, USA (2012)
- Garg, S.K., Yeo, C.S., Buyya, R.: Green cloud framework for improving carbon efficiency of clouds. In: Garg, S.K., Yeo, C.S., Buyya, R., (eds.) Proceedings of the 17th International European Conference on Parallel and Distributed Computing (EuroPar), Bordeaux, France, August–September 2011
- 11. Barták, R.: Constraint programming: in pursuit of the holy grail. In: Proceedings of WDS 1999 (1999)
- 12. Hladik, P.-E.: How to solve allocation problems with constraint programming (2005)

A Trusted Way for Encryption Key Management in Cloud Computing

Saad Fehis^{1(x)}, Omar Nouali², and Mohand-Tahar Kechadi³

¹ Higher National School of Computer, Algiers, Algeria s_fehis@esi.dz

² Research Center on Scientific and Technical Information CERIST, Algiers, Algeria onouali@cerist.dz

³ School of Computer Science and Informatics, University College Dublin, Dublin, Ireland tahar.kechadi@ucd.ie

Abstract. We propose an approach to provide the cryptography key management system (*CKMS*) as a trusted security service in Cloud Computing, based on the trusted platform module (*TPM/vTPM*). In this approach we have used the TPM's capabilities/functions as a secure way and a root of trust for this kind of services. Therefore, and as an application case, we have used TPM's key generation component as a trusted way to generate and to sign any encryption/signing keys by the CKMS for their customers.

Keywords: Cloud computing \cdot Security as a service \cdot Cryptographic key management system \cdot Trusted platform

1 Introduction

Providing an IT services today, favored by the evolution in computing architectures, known by a Cloud Computing (*IaaS*, *PaaS*, *SaaS*) [1]. The Cloud computing has a many of advantages, as the flexibility of resources using (*CPU*, *storage*, *network*). Consequently, the numbers of cloud-based services' users have been increasing, (*such as* salesforce.com or Google Apps,) means that many mobile IT users will be accessing business data and services without traversing the corporate network. This will increase the need for enterprises to place security controls between mobile users and cloud based services. Therefore, using of the Security as a service (*SecaaS*). The SecaaS refers to the provision of security applications and services via the cloud either to cloud-based infrastructure and software or from the cloud to the customers' on premise systems. This will enable enterprises to make use of security services in new ways, or in ways that would not be cost effective if provisioned locally [2].

Gartner is predicting the cloud-based security services market, which includes secure email or web gateways, identity and access management (*IAM*), remote vulnerability assessment, security information and event management to hit 4.13 billion by 2017 [3]. According to its "Market Trends: Cloud-based Security Services Market, Worldwide, 2014," Gartner is predicting growth is likely to come because of the adoption of these cloud-based security services by small-to-mid-sized business (SMB) in particular. Certain market segments mentioned in the report will see higher overall sales and yearover-year growth.

However, the shared resources (*multi-tenant environment*) create many security issues, so, a real challenge in adoption of Cloud-Computing [4]. Therefore, there are new problems add to the traditional threats, will be addressed such as: vulnerability due to virtualization of the physical infrastructures [5], data isolation, management of privacy and confidentiality of data, consequently; the creation of the trust between the customers and services provider.

In this work, we have focused on the providing to the clients a trusted software components (*as a virtual Platforms*), using they for providing a security as a service (*SecaaS*) [2]. These components provide to customers a several of a security services features as: the e-mail/web content filtering, identity and access management, cryptography or cryptographic key management system (*CKMS*). The latest service is the main subject matter in this work.

The CKMS as a service, can be provided it as a dedicated or shared platforms (*to companies or users group*), and this depending on the software architecture adopted. Their security in a virtualized environment requires the assurance of its integrity at the disk level (*Source, data, executable*) and at memory level (*running*) firstly, and secondly, providing to the owner of CKMS service or a trusted third party the control, the check or the audit, by verification mechanisms of the trusted chain.

However, the safety of this type of solution, does not only depend on the integrity of software, but also of its operation and its interaction with a multi-tenant environment as; the access control (*CKMS's Metadata*) and the control of information flow between instances of the same CKMS provided by the same service provider for different customers (*companies or group of users*) or with other kind of software hosted in the same cloud (*the same shared physical infrastructure using the virtualization*). In this context and to deal with those challenges, the CKMS needs to manage the encryption keys in secure way, and in all of its life cycle, from the key generation with its related data (*Metadata*), to delivering they for their customers (*Owners of the key*). Therefore, we need to a trusted component to answer this capability as the Trust platform Module (*TPM*) [6], (vTPM) [7].

In this work, we propose an approach to provide the CKMS as a trusted security service (*CKMS_SecaaS*) in the cloud computing based on the virtual trusted platform module (*vTPM*). The vTPM's capabilities/functions used as a secure way for the protection of the CKMS_SecaaS service; therefore, the vTPM used as a root of trust, and as an application case, we have used the vTPM's key generation component as a secure and a trusted way to generate and to certificate any encryption/signing keys for the client for the CKMS_SecaaS.

In order to treat these challenges, we have structured this paper as follows: In Sect. 2 we will present the CKMS, its architecture, their security challenges, and its deployment in Cloud Computing. In Sect. 3, we will present a background on the TPM as a root of trust and related works. In Sect. 4, we will present our approach TPM based CKMS_SecaaS, with returned results. In the Sect. 5 we will present a conclusion with the future direction of our work. Then the list of references used in this work.

2 Cryptographic Key Management System

In this section we will present firstly a set of definitions related to the CKMS. Secondly, the CKMS as service and its related to Software as service model; and finally, the security challenges related to a CKMS as a security service (*CKMS_SecaaS*).

2.1 CKMS Presentation

A CKMS consists of policies, procedures, components and devices that are used to protect manage and distribute cryptographic keys and certain specific information, called (*associated*) metadata herein. A CKMS includes any device or sub-system that can access an un-encrypted key or its metadata. A CKMS will be a part of a larger information system that executes information processing applications. While the CKMS supports these applications by providing cryptographic key management services [8].

A CKMS can be as simple as a software program running on a single-user computer and supporting user applications. It can also be as complex as a variety of sub-systems, each containing many devices that provide key management services to numerous networked users and applications. Therefore, it may be widely distributed geographically and connected with a myriad of communications networks [8].

A key is associated with metadata that specifies characteristics, constraints, acceptable uses, and parameters applicable to the key. For example, the key type, how it was generated, when it was generated, its owner's identifier, the algorithm for which it is intended, and its crypto period. Therefore, the metadata elements may be generated by the same entity that generates the key or they may be received from another trusted entity [8].

2.2 CKMS in the Cloud Computing

In Cloud computing, the CKMS_SecaaS can be viewed as a SaaS [1]. The SaaS is "Software deployed as a hosted service and accessed over the Internet". From an application architect's point of view, a designed SaaS application is: Scalable, Multi-tenantefficient, and Configurable [9]. An application SaaS maturity can be expressed using a model with four distinct levels (Fig. 1). Each level is distinguished from the previous one by the addition of one of the three attributes listed above.

Architecturally, SaaS applications are largely similar to other applications built using service-oriented design principles [9]. However, the real challenge is: How to keep a separate customer's data for each level, and at each data state: as in storage (*memory/disk*); running (*processor*) and in transit network? Therefore, the development and the use of a configurable metadata is not an easy task!

In our case, the CKMS as security service architecture can be developed, deployed and provided as any SaaS applications (Fig. 1). However the nature of this service as security service needs a real isolation of each CKMS's instance from any other software hosted in the same shared platform physique. Therefore, the using of the second level (*configurable*) is an interesting for the isolation but not for a good performance. Consequently, using of the 4th level load balancing between the same customer's instances is



Fig. 1. Four-level SaaS maturity model [9]

the best choice (*see* Fig. 2). And we can apply the isolation between the set of the same tenant's instances and the other tenants. This isolation can be viewed as the built of a wall around each tenant's instances.



Fig. 2. Multi-tenant Isolation

2.3 CKMS_SecaaS's Security Challenges

A CKMS must be designed in a manner that supports the goals of each organization using the CKMS. Several types of policies and their relationships will influence the design of a CKMS, as information management policy (*IMP*), information security policy (*ISP*) and CKMS security policy. A CKMS security policy is created to establish and specify requirements for protecting the confidentiality, integrity, availability, and source authentication of all cryptographic keys and metadata used by the organization. These protection requirements cover the entire key life cycle, including when they are operational, stored, and transported. A CKMS Security Policy includes the selection of all cryptographic mechanisms and protocols to be used throughout the organization's automated information systems [8].

Like keys, metadata needs to be protected from unauthorized modification and disclosure and have its source adequately authenticated. Therefore, a key and its metadata need a trusted association and supporting processes between them [8]. Therefore, a CKMS requires different types of security controls to protect its components, devices, and the data that they contain. To do this, the CKMS will likely require computer systems to perform functions in a secure way, such as key generation, key storage, key recovery, key distribution, cryptographic module control, and metadata management.

However, how protect and provide the CKMS hosted in the Cloud Computing? How perform these functions in a secure way? In this context, our work has focused on the security approaches used to provide a CKMS as a trusted component.

To protect the CKMS's metadata (*Its dictionary D-CKMS*), we proposed an encrypted data's model for the DCKMS [10]. The model is an encryption of the D-CKMS, which provides the D-CKMS management without decryption. However, and to introducing the trust concept in this kind of service we need to answer that the CKMS_SecaaS is trusted software and not a malware, and the keys are generated by a trusted component. The trusted platform module (*TPM*), can act as a root of trust for those challenge, therefore CKMS_SecaaS based on TPM, where we will be presented in the next section a background on TPM.

3 TPM Background

The Trusted Platform Module (*TPM*) is a hardware chip designed to enable commodity computers to achieve greater levels of security than was previously possible. The TPM stores cryptographic keys and other sensitive data in its shielded memory, and provides ways for platform software to use those keys to achieve security goals. Application software such as Microsoft's BitLocker and HP's Protect Tools use the TPM in order to guarantee security properties.

TPMs are manufactured by chip producers, including Atmel, Broadcom, Infineon, Sinosun, STMicroelectronics, and Winbond. It is specified by the Trusted Computing Group (*TCG*) and other industry consortium in three documents [6] totaling about 800 pages. In this section we will give an overview on the TPM and the trust, its architecture and its operation.

3.1 The TPM and the Trust

The TPM has been designed specifically to support trusted computing platforms. Therefore, in order to understand the TPM design requirements, it is first necessary to understand what the desirable features of a trusted platform are. To do this, a definition is required as to exactly what is meant by the term "trusted platform":

- The TCG defines trust to be: The expectation that a device will behave in a particular manner for a specific purpose [11],
- Another definition of a trusted platform is provided by Pearson [12] who states that: A Trusted Platform is a computing platform that has a trusted component, probably

in the form of built-in hardware, which it uses to create a foundation of trust for software processes,

• Or by Balacheff et al. [13] who say: A trusted platform is defined as a computing platform that has a trusted component, which is used to create a foundation of trust for software processes.

It is perhaps appropriate at this point to make a subtle distinction between what is meant by a trusted component, such as the TPM, and a trustworthy component. Anderson [14, 15] given definition of these terms who states that: "The proper definition is that a trusted system or component is one whose failure can break the security policy, while a trustworthy system or component is one that won't fail". By implementing this trusted component, the TPM, as a tamper proof integrated circuit; and binding it to the platform, usually on a printed circuit board containing a more powerful processor capable of running software applications; the TPM can be used as the foundation of trust for higher level processes that run on the main processor.

In order to establish this foundation of trust, the TPM is expected to provide a fundamental set of security features which have been defined by the TCG. The features that a trusted platform should have are: Protected Capabilities, Integrity Measurement and Reporting, Confidentiality and Integrity Protection, Secure Storage and Process Isolation.

3.2 TPM's Components

The TPM has a set of components (Fig. 3); which we classified them as flowing:

- The first class related to a set of tools, for managing the encryption/signing keys and related algorithms. In this set we have the Cryptographic Co-Processor implements cryptographic operations within the TPM. Those operations include the Asymmetric key generation (*RSA*), Asymmetric encryption/decryption (*RSA*), Hashing (*SHA-1*) and the Random number generation (*RNG*). The TPM uses these capabilities to perform generation of random data, generation of asymmetric keys, signing and confidentiality of stored data. The TPM may symmetric encryption for internal TPM use but does not expose any symmetric algorithm functions to general users of the TPM.
- The second class related to the storage, where we have two types of memory. The first is a Non-Volatile Memory that used to store persistent identity and state associated with the TPM as the endorsement key (*EK*) and storage root key (*SRK*); the second is a volatile memory contains the platform configuration register (*PCR*), is a 160-bit storage location for discrete integrity measurements and the loaded keys created by the TPM.
- The third class related to the Execution of the code, the powering of the TPM and the In/Out bus; where the Execution Engine runs program code to execute the TPM commands received from the I/O port. The execution engine is a vital component in ensuring that operations are properly segregated and shield locations are protected. The Opt-In component provides mechanisms and protections to allow the TPM to be turned on/off, enabled/disabled, activated/deactivated. The Opt-In component

maintains the state of persistent and volatile flags and enforces the semantics associated with these flags. The last one is the I/O component it manages information flow over the communications bus.



Fig. 3. TPM Architecture

3.3 TPM's Keys Management Concept

Each TPM has a unique public/private key pair called the endorsement key (EK), set at manufacture time and usually certified by the manufacturer. The EK can be taken to be the identity of the TPM. In addition to EK, when ownership of the TPM is taken, the TPM generates a public/private key pair called the storage root key (SRK) which is the root of the tree of storage keys; and it also generates a secret random value called tpmProof which is used by the TPM to identify blobs that it creates.

For platform authentication, one may create signing keys known as application identity keys (*AIKs*). These may be used to sign (*appropriately tagged*) application-specific data, and to sign PCR values. For such signatures to be useful, AIKs need to be certified as belonging to a TPM. For reasons of user privacy, the certificate will not specify which TPM they belong to; it will just specify that they belong to a TPM. There are two ways to obtain a certificate on an AIK: using privacy CAs, or using Direct Anonymous Attestation (*DAA*) [6].

Now how about the measurement and reporting operations? The TPM contains a number of 160-bit registers called platform configuration registers (*PCRs*) intended to enable a relying party to obtain unforgeable information about the platform state. We think of the platform as consisting of several "components", which may receive control and pass on control to another component. Typical components are the BIOS, the master boot record, boot sectors, the boot loader, and ultimately the operating system and application software. A component can "measure" another component (*compute its*)

hash) and insert that measurement into a PCR (*for example, before passing control to it*). This insertion is an irreversible process, known as "extending" the PCR.

A given PCR can be extended with any number of measurements. The current value of the PCR represents the accumulation of them all. A secure chain of trust can be established by ensuring that the very first code segment executed on power-up is measured and that measurement is extended into a PCR. Then, every component "A" that loads another component "B" and passes control to it ensures that "B" is first measured and the measurement is extended into a PCR. The PCRs then represent an accumulated measurement of the history of the executed code from power-up to the present.

A TPM signing key (AIK) can be used to sign the values of the PCRs. In this way, application software can send assurance about the state of the platform to a third party. Additionally, PCR values can be used to ensure that certain data is accessible only to authorized software.

3.4 TPM's Key Related Tools

The TPM has a set of tools and components (Fig. 3) used for the keys generation [6]. The TPM key types are defined at key creation time by the User, the different key types are:

- Non-Migratable Key (*NMK*): A key which is bound to a single TPM. This is a key that is (*statistically*) unique to a single TPM and cannot be migrated or exported from the TPM.
- Migratable Key (*MK*): A key which is not bound to a specific TPM, and with suitable authorization, can be used outside a TPM or moved to another TPM.
- Certifiable Migratable Key (*CMK*): A key whose migration from a TPM is highly controlled and the TPM can attest/certify it properties.

The migration destinations are defined and authorized by the TPM Owner. To use the key generation capabilities, the TPM provides a set of functions, partitioned into two classes:

- Migration functions: In this class there are many functions (*commands*) used to create and transfer migratable objects from one TPM to another for backup, upgrade or to clone a key on another platform. To do this, the TPM needs to create a data blob which holds the encrypted key exported from a TPM [21]. As an example, the TPM_CMK_CreateKey command both generates and creates a secure storage bundle for asymmetric keys whose migration is controlled by a migration authority. The resultant key must be a migratable key and can be migrated only by TPM_CMK_CreateBlob command; the command is Owner authorized via a ticket. The migrationAuth is an HMAC of the migration authority and the new key's public key, signed by tpmProof (*instead of being tpmProof*).
- Cryptographic Functions: In this class also there are many functions: as the TPM_Sign command, can be used to sign data (*as a blob*) and returns the resulting digital signature. The TPM_GetRandom command returns the next bytesRequested bytes from the

random number generator (*RNG*) for any caller. The TPM_CertifyKey operation allows one key to certify the public portion of another key.

3.5 TPM in Cloud Computing

In Cloud Computing, all hardware components are shared resources by a technical virtualization. However, the TPM is designed for a single physical machine with a single operating system. Therefore, several TPM virtualization approaches and use to multiple instances of virtual TPM (vTPM) have been proposed, with the aim that each vTPM is designed to secure a single virtual machine [7, 16–20].

However the establishment of a vTPM in the virtual machine manager layer (*VMM*) may be venerable to attack different types of software, including the attacks of VMM itself (Fig. 4). Therefore, securing a vTPM is directly depend on securing VMM and a virtualization techniques, which involve many challenges remain to be addressed: as the protection of vTPM at Storage, their secrets across Reboots, attestation (*Credentials, Deep Attestation Layer Bindings*), vTPM backup, restore and migration.



Fig. 4. Three layer architecture with privileged [7]

4 TPM Based CKMS_SecaaS Approach

The CKMS used to protect manage and distribute cryptographic keys and certain specific information, called (*associated*) metadata. Therefore, the CKMS needs a key generated and signed by a trusted component, and delivered it in a secure way to this CKMS, where then, the key will be delivered to he/she (*any user, thing or application*) requested the key. The key can be symmetric or asymmetric, and can be used for any purpose encryption/signing in any way as in the cloud computing, internet of things, user mobile outside of his company's network or inside of any company's network.

For this purpose, the trusted platform module can be used by the CKMS as a trusted component (*see* Fig. 5). However, how can use it to provide a CKMS as a trusted service in the cloud computing.



Fig. 5. CKMS as a security services

TPM has a set of capabilities, where we can use them, to create, certify and sign any keys inside itself, therefore, the TPM as a trusted and physical platform. However, in the cloud computing all physical components are shared between tenants, using the virtualization technical. Therefore, the using of the vTPM by the CKMS, consequently the inheriting of its related challenges! In this section we will give the returned results, and the deployment approach of the CKMS based on TPM in cloud computing.

4.1 Deployment Approach

From the previous section, we have captured the following main points (Fig. 5):

- Using of such components (*TPM*, *vTPM*) as a root of trust for the CKMS_SecaaS.
- Inspiring from the architecture components and theirs deployment [6, 7, 18–20] for the CKMS solution, where these components contain the most elements of the features platform of a CKMS. This implies their interactions between them or their extension for communication with the external environment.
- Exploitation of the security solutions already proposed or improvement they for the CKMS, because we have the most same challenges with these components and our module (*migration and protection*, ..).
- Creation of the trust between the customers and providers, by implementing the check of trusted chain, based on integrity measures of the CKMS instances service.

However, these solutions have many challenges:

• All these solutions based on the physical hardware TPM; therefore, the inheriting of the visualization problem related to migration and upgrading.

- Protection of the vTPM instances from the attacks at the VMM level; therefore the information flow control problem between the instances and from the external attacks!
- The protection of the CKMS's data dictionary (*Metadata*) needs others solutions as encryption to protect it.

4.2 TPM vs CKMS_SecaaS

Now, what about the benefit of TPM from the CKMS: The TPM has a limited storage capability. However, in Cloud computing there are important number of software hosted and running in the same (*shared*) physical platform, and they need an important number of keys from TPM, for encryption, signing, authentication. Therefore; any enterprise key management systems/solutions (*as CKMS_SecaaS*) will need to:

- Support key management life cycle for TPMs
- Support the TPM as a platform identity token
- Support TPM-enabled (TPM-aware) applications
- Manage the TPM as a generic secure key store and trust anchor store

Consequently; we can resume the benefit relation between the TPM and the CKMS in the following (Fig. 6):



Fig. 6. TPM vs CKMS

5 Conclusion

Outsourcing the CKMS to a trusted platform in Cloud Computing remains a real challenge. Because, the security of all applications (*solutions*) based on the security of the encryption/signing keys, which are themselves dependent on the security of CKMS's operations itself. It is within this context that reside the challenge of our research work and our contributions.

For treating the challenges related to the trust in a multi-tenant environment (*Cloud-Computing*), we investigated in the solutions that can serve as a root of trust [6, 7, 18-20]; Those solutions can ensure the integrity software, the trust chain, or using their architecture as a basis of inspiration for the development of CKMS as a security service.

In this paper, we have proposed an approach to provide the CKMS as a trusted security service in the cloud computing based on the TPM/vTPM. The TPM's capabilities/functions can be used as a secure way for the protection of this type of service. Therefore, the TPM used as a root of trust, and the TPM key generation component used as a secure and a trusted way to generate and to sign any CKMS's encryption/signing keys ordered by customers.

However, these solutions are based on a hardware module (*TPM*), therefore the inheritance of the challenges related to the TPM virtualization! In the future work we believe to treat the following challenge:

- Implement of the deep attestation from the CKMS's instance to the TPM (*as a physical component*). This to answer the authentication problem.
- Using the vTPM instance as a CKMS instance, because the vTPM instance is a software component running on the VMM platform level; and this to protect the CKMS_SecaaS at this level.
- Protection of the CKMS instance by the creation of the walls around the CKMS instance as a Chinese wall [22], and this to control the information flow between the CKMS's instance themselves and with any other software hosted inside the same shared physical platform, or from the Cloud Computing outsider.

References

- 1. Mell, P., Grance, T.: The nist definition of cloud computing (2011). http://csrc.nist.gov/ publications/nistpubs/800-145/SP800-145.pdf
- Jerry, A., Alan, B., Dave, C., Nils, P., Paul, K., Jim, R.: Defined categories of service 2011. In: Cloud Security Alliance, Security as a Service Working Group (2011). http:// www.cloudsecurityalliance.org/guidance
- Janessa, R.: Gartner says cloud based security services market to reach 2.1 billion in 2013. Gartner, Technical report, October 2013. http://www.gartner.com/newsroom/id/2616115
- Rafal, L., Dave, S., Bryan, S., Luciano, J.S.: The notorious nine: cloud computing top threats in 2013. In: Cloud Security Alliance, Top Threats Working Group and Others (2013). http:// www.cloudsecurityalliance.org/topthreats
- Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., Warfield, A.: Xen and the art of virtualization. ACM SIGOPS Oper. Syst. Rev. 37(5), 164– 177 (2003)
- TCG: Tpm main part 1 design principles, specification version 1.2 revision 116. Trusted Computing Group, Copyright (c) 2003–2011 Trusted Computing Group, Incorporated, Technical report, March 2011
- TCG: Virtualized trusted platform architecture specification version 1.0.26. Trusted Computing Group, Copyright (c) 2003–2011 Trusted Computing Group, Incorporated, Technical report, 27 September 2011
- Barker, E., Smid, M., Branstad, D., Chokhani, S.: A framework for designing cryptographic key management systems, special publication 800-130. U.S. Department of Commerce, National Institute of Standards and Technology (NIST), Technical report, April 2012
- Frederick, C., Gianpaolo, C.: Architecture strategies for catching the long tail, application architecture software-as-a-service (saas). Microsoft Corporation, Technical report, April 2006. https://msdn.microsoft.com/en-us/library/aa479069.aspx
- Fehis, S., Nouali, O., Bentayeb, S.: Meta-data's protection in ckmsas-a-security services. In: Proceedings 4th International Conference on Information Systems and Technologies Conference ICIST 2014, 22–24 March 2014, Valencia, Spain, pp. 195–206 (2014)

- TCG: Tcg specification architecture overview, tcg specification revision 1.4. Trusted Computing Group, Copyright (c) 2003 Trusted Computing Group, Incorporated, Technical report, August 2007
- 12. Pearson, S.: Trusted computing platforms, the next security solution. HP Laboratories Bristol, Technical report HPL-2002-22, November 2002
- 13. Balacheff, B., Pearson, S., Chen, L., Plaquin, D., Proudler, G.: Trusted Computing Platforms: TCPA Technology in Context. Prentice Hall Professional, Upper Saddle River (2003)
- Anderson, R.: Cryptography and competition policy: issues with 'trusted computing'. In: Proceedings of the Twenty-Second Annual Symposium on Principles of Distributed Computing, pp. 3–10. ACM (2003)
- 15. Anderson, R.: Security Engineering a Guide to Building Dependable Distributed Systems. Wiley, New York (2001)
- Sadeghi, A.-R., Stüble, C., Winandy, M.: Property-based TPM virtualization. In: Wu, T.-C., Lei, C.-L., Rijmen, V., Lee, D.-T. (eds.) ISC 2008. LNCS, vol. 5222, pp. 1–16. Springer, Heidelberg (2008). doi:10.1007/978-3-540-85886-7_1
- Danev, B., Masti, R.J., Karame, G.O., Capkun, S.: Enabling secure vm-vtpm migration in private clouds. In: Proceedings of the 27th Annual Computer Security Applications Conference, pp. 187–196. ACM (2011)
- Krautheim, F.J., Phatak, D.S., Sherman, A.T.: Introducing the trusted virtual environment module: a new mechanism for rooting trust in cloud computing. In: Acquisti, A., Smith, S.W., Sadeghi, A.-R. (eds.) Trust 2010. LNCS, vol. 6101, pp. 211–227. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13869-0_14
- Chang, D., Chu, X., Qin, Y., Feng, D.: TSD: a flexible root of trust for the cloud. In: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 119–126. IEEE (2012)
- Krautheim, F.J., Phatak, D.S., Sherman, A.T.: Private virtual infrastructure: a model for trustworthy utility cloud computing. University of Maryland Baltimore County, Baltimore, MD, Technical report (2010)
- TCG: Tpm main part 3 commands, specification version 1.2 level 2 revision 116. Trusted Computing Group, Copyright (c) 2003–2011 Trusted Computing Group, Incorporated, Technical report, March 2011
- Fehis, S., Nouali, O., Kechadi, T.: A new chinese wall security policy model based on the subject's wall and object's wall. In: 2015 First International Conference on Anti-Cybercrime (ICACC), pp. 1–6, November 2015

Use of Cloud Computing Technologies for Geographic Information Systems

Ahmed Ziani^{1(\mathbb{I})} and Abdellatif Medouri²

 ¹ Information Systems and Telecommunications Laboratory, Faculty of Science, Abdelmalek Essaadi University, Tetuan, Morocco zianiahmed@gmail.com
 ² Modeling and Information Theory Group Polydisciplinary, Faculty of Tetuan, Abdelmalek Essaadi University, Tetuan, Morocco

amedouri@uae.ma

Abstract. Geographic Information Systems (GIS) plays an essential role in wide range of areas and is extensively adopted nowadays. They combine social, economic and topographical data that is used for a variety of purposes including flood defense planning, healthcare and road traffic management. GIS is a collection of tools that captures, stores, analyzes, manages, and presents data that are linked to geographical locations. This technology continues to change the way to manage infrastructure, emergency response, and planning [1, 2]. Several efforts are being made to upgrade the conventional GIS applications in order to improve decision making, deliver better service, and reduce operating costs. "Cloud computing" a term which has become popular in recent years, has appeared as technologies by its focus on large-scale asset sharing and reduced cost for largescale data storage technology to be applied to solve and overcome the challenges in GIS applications [3]. Many efforts are being made to upgrade the GIS applications in order to improve decision making, deliver better service, and reduce operating costs. "Cloud computing" a term which has become popular in recent years, has appeared as technology able to solve and overcome the challenges in GIS applications, by its focus on large-scale asset sharing and reduced cost for large-scale data storage. In this paper, a brief evaluation of Cloud Computing approach to GIS is presented and architecture for GIS Cloud System is proposed.

Keywords: Geographical Information Systems \cdot Cloud GIS \cdot Cloud computing \cdot Microsoft's Windows Azure \cdot Traccar Web UI

1 Introduction

Geographical Information Systems (GIS) play an essential role in wide range of areas of interest and they are widely used by Governments, Businesses, Scientists, Environmental Organizations, and Researchers etc. GIS is a collection of tools that collects, stores, manages, analyzes and presents data that are related to geographical locations. With great success, SIG has a long history of adapting to new technologies, applications, client types and business models. Nowadays, every technical innovation cycle has improved GIS [1–3].

With the arrival of smartphones, there is an increase in the demand for location-based services. It has then led to the need for more accurate positioning services which is a combination of GIS data, internet technology and software application [2].

With the explosion of wireless devices and social collaboration technologies, data has become much more complex than ever. All this data can't be stored and processed in traditional systems. Traditional analytic platforms can't handle the variety of Data, which brings along challenges to us for trying to deal with the speed at which the data is flowing that requires that we perform analytics against their volume and variety of data. Cloud computing can be seen as an emerging trend to deploy and maintain software and has been adopted by many industries such as Google, Amazon, IBM and Microsoft. Cloud Computing can be a key computing platform for sharing resources including infrastructure, software, applications and business processes [4, 5].

Cloud Computing can be defined as one or more unified computing resources based on service-level agreements and has been presented as a distributed and parallel system consisting of a pool of interconnected and virtualized computers that are dynamically provisioned [5].

Cloud computing structures typically use a Web client to serve as a user interface to perform certain functions such as accessing applications and exploring data stored on off-site servers. In other words, the GIS user may be able to collect, access and analyze data from a smartphone or tablet while staying in the field or out of the office [6, 7].

This paper is organized as follows. Section 2 discusses the Geographic Information System (GIS). Section 3 gives an overview Cloud computing. In Sect. 4, the need for GIS cloud is discussed, then, the proposed architecture is presented in Sect. 5. The paper is concluded in Sect. 6.

2 Geographic Information Systems

A geographic information system (GIS) is a computer system designed to capture, store, manipulate, manage, analyze and display data related to positions on Earth's surface and show many different kinds of data on one map. This enables people to more easily see, analyze, and understand patterns and relationships [2, 3].

GIS can use any information that includes location. The location could be presented in several different ways, such as latitude and longitude, address, etc. The system can include data about people, such as population, income, traffic, or education level. It can contain information about the land, as well as the location of streams, different kinds of vegetation and soil. It can contain information about the sites of factories, farms and schools, or storm drains, roads, and electric power lines [2, 7].

By using GIS, users can compare the locations of different sites for determine how they relate to each other. For example, the same map can include sites that produce pollution, like gas stations, and sites that are sensitive to pollution, like wetlands. Such a map would help people determine which wetlands are most at risk [3, 5, 7].

GIS is more than just software. Methods are combined with geospatial software and tools, to enable spatial analysis, management large datasets, and the display of information in a map/graphical form [2, 7].

3 Cloud Computing

Cloud computing refers to the recent advancement of distributed computing by providing 'computing as a service' for end users in a 'pay-as-you-go' mode; like a mechanism had been a long-held dream of distributed computing and has now become a reality [1, 8].

The National Institute of Standards and Technology (NIST) defines cloud computing as '... a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [8].

3.1 Cloud Service Models

Cloud computing is provided through at least four types of services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Data as a Service (DaaS) [1, 4, 6].

3.1.1 Infrastructure as a Service (IaaS)

Iaas provides many competences to the clients such as provision processing, storage, networks, and other fundamental computing resources where the client is capable to deploy and run arbitrary software, which can include operating systems and applications. The user does not manage or control the underlying cloud physical infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components.

3.1.2 Software as a Service (SaaS)

Saas allows the user, via several capabilities, to use the provider's applications running on a cloud infrastructure. The applications are accessible from various devices through a client interface such as a web browser. The user does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of provider-defined user-specific application configuration settings.

3.1.3 Platform as a Service (PaaS)

The capability provided to the user is to organize onto the cloud infrastructure usercreated or acquired applications created using programming languages and tools supported by the provider. The user has control over the deployed applications and possibly application hosting environment configurations, but does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage.

4 The Need for Gis Cloud

GIS Cloud offers reliable tools that can help various companies, particularly, when optimization and cost reduction are critical. Some basic principles that characterize the GIS Cloud to be accepted as the serious challenger for next generation GIS computing prototype are [5–7]:

- Providing Application Infrastructure
- Support Technology Infrastructure
- Implementation Cost Benefit
- Reducing Support and Maintenance
- Leveraging Data Command
- Data Conversion and Presentation
- Location Independent Resource- Pooling

4.1 Providing Applications Infrastructure

GIS Cloud provides the framework for systems and geo- location of business data. For companies that have already invested in GIS, Cloud GIS resources can be leveraged to grow the support, making the organizations business and geographic data easier to be analyzed, authored, and managed. GIS Cloud provides Web services and application hosting for organizations to facilitate access, publication and consumption of the organizational geographic data.

4.2 Support Technology Infrastructure

GIS Cloud as a computing prototype for geographical data delivers subscribers' leverage of virtualized sophisticated hardware and full access to data creation and software resources, visualization, analysis and editing. Simple collaborative utilities further improve the spread of GIS through an office or through the globe.

4.3 Implementation Cost Benefit

GIS Cloud has a great capability to provide to users the advanced geo-technology infrastructure, the services and the geo-spatial data without massive initial investment in cost and time, or partial maintenance. This is very important because the geographic information system implementation cost can be relatively great.

4.4 Reducing Support and Maintenance

Implementing GIS Cloud within an organization eliminates the need for in-house GIS capability for basic geo-information access capabilities. For societies that already have a GIS competence, it will be complimentary for highly skilled in-house staff from having to take care of basic information requirements, and letting them deal with more complex
responsibilities and services. For customers, that said no greater investment in implementation and significant reductions in their in-house support and maintenance.

4.5 Leveraging Data Command

Provide Imagery and Topographic mapping, which acts as a basis on which other spatial data are encrusted, is the aim of GIS. For GIS application providers, to obtain and process from spatial data source, it is too expensive. The GIS Cloud has abilities to deliver the basic data as component of the core services made available through standard Internetenabled devices. The rapid elastic nature of GIS Cloud makes it certain that users can increase or decrease capacity according to their needs. GIS Cloud delivers advanced services for Storage and Management Spatial Information prove to be favorable to users.

4.6 Data Conversion and Presentation

A data conversion service indicates the transformation and importing from one format into a new database. For any GIS, this service is utmost importance and requires dedicated in-house skilled potential and technical resources which include infrastructure, software services. GIS Cloud offers to users the spatial data conversion services on demand and without any requirement of in-house resource capabilities. The advanced features like 3D presentation of spatial information in GIS Cloud came to substitute the traditional presentations that flatten all of the interesting details into force-fitted plane geometry.

4.7 Location Independent Resource- Pooling

GIS Cloud has the wonderful ability of providing location independent resource pooling. That mean, the processing and storage demands are balanced across a shared infrastructure with no particular resource assigned to any individual user. The pay-per-use GIS cloud property offers the leverage that consumers are charged based on their usage of a combination of computing power, bandwidth use and/or storage.

4.8 Cloud GIS Platforms

Cloud-based virtual machines are primarily seen as Infrastructure as a service (Iaas). In other words, the instances of VM provided should be perceived (nominally, of course) as a real computer with its CPU, memory, network, storage etc. [10].

There are many VM cloud infrastructures that support cloud technology and deployment of a GIS application. When searching, there are 3 largest cloud systems available: Amazon Elastic Cloud 2, Google Compute Engine and Microsoft Azure Virtual Machines. Because each cloud provider has unique features and attractive offers with advantages and disadvantages, we have chosen in our work for Windows Azure that supports the following criteria presented in Table 1 [10, 12].

Challenge	Amazon EC2	Google CE	Microsoft Azure VM
Number of instance templates available	39	18	40
GPU acceleration	Yes	No	No
Custom instance creation feature	No	Yes	No
CPU limits	1–40	1 Shared – 32 dedicated CPU	1–32 CPU
Memory limits	0,5–244 GB	0,6–208 GB	0,75–448 GB
Temporary storage limits	Up to 48 TB (Multiple Disks)	3 TB	2 TB
Number of OS supported	11	9	9
Number of databases supported	5+	3	3
Autoscaling	Yes, clone building	Yes, clone building	Yes, presettable group
Size change	Available	Available	Available
Service Level Agreement (SLA Terms)	Credit for 1+minutes downtime, max monthly credit: 30%, uptime SLA: 99.95%	Credit for 5+consecutive minutes downtime, max monthly credit: 50%, uptime SLA: 99.95%	Credit for 1+minutes downtime, max monthly credit: 25%, uptime SLA: 99.95%
Cloudberry support	Yes	In progress	Yes

Table 1. Comparison between amazon EC2, Google CE and Microsoft azure VM

Microsoft Azure has faster growth than other solutions and it is clear visionary in the cloud market and has much better footsteps on enterprise world because of Microsoft. And it is a great choice for windows shops also for many open source shops. Also, it is auto-scaling functions via Availability groups, in which VMs can be added. Then we can set the number of instances to scale up and down and choose the criteria (CPU workload or a queue size). All instances can be easily resized using either Azure web-interface [11–13].

Microsoft Azure SLA are the same as Amazon's, with distinction in maximum credit (only 25%). And can be additionally performed by its Azure Backup tool and Recovery Services. They cope well with basic functions.

5 Cloud Computing Architecture

5.1 Architecture Basic

Figure 1 describes the layers of the proposed GIS Cloud architecture. The objective is to provide the power of desktop GIS solutions on web-based platform. So, we have built the system using Microsoft Azure as cloud infrastructure and Traccar Web UI as GIS platform. The given system will be capable of providing flexible solution, heterogeneous

and infrastructure, secure and personalized environment, extensive business intelligent system and elastic platform to the GIS users [10, 11, 14].



Fig. 1. Proposed GIS cloud architecture

5.2 Components and Layers

The proposed GIS Cloud architecture can be divided into five key components which are:

5.2.1 Client Layer (CL)

The Client Layer the interface directly oriented to users. Like client layer in distributed-GIS, in cloud GIS system, CL plays same roles like enabling user to manipulate spatial data and manage information and authorization.

5.2.2 The Application Service Layer (ASL)

ASL contains management applications which users can organize their data in order to form an Information System (IS) and produce decision making.

5.2.3 GIS Platform Layer (GPL)

In this layer Traccar Web UI is used to support the seamless functioning and optimization of the Cloud GIS System as a whole focused GIS utilities for mapping, address lookup, routing, reverse geocoding, and navigation.

5.2.4 GIS Cloud Storage Layer (CSL)

The Windows Azure SQL Database is used the utilization to deal with different database providers such as MySQL server and SQL Server 2012 on Azure Cloud SQL Database.

322 A. Ziani and A. Medouri

5.2.5 Infrastructure Layer (IL)

In this model, Microsoft Azure VM is proposed to be used as cloud infrastructure offers complete control of computing resources by using virtual machines physical storage, operating system.

5.3 Implementation of Proposed System

The implementation of the proposed, see Fig. 2, model can be described as the presentation of client web interface cloud based for mobility management which locates data from different devices about its recourses. So the power of infrastructure as a service is exploited by determining the suitable resources.



Fig. 2. Implementation of GIS application with Microsoft Azure and Traccar Web UI technology

6 Conclusion

Using cloud computing technology provides a plethora of benefits for GIS applications as compared to the traditional approaches. This paper discussed cloud computing technology in the GIS application, first we introduced the concepts of GIS and we presented the cloud computing; Then the paper focused on how to apply the cloud computing techniques in the GIS spatial data storage and processing; we proposed a model in which the process of integrating GIS application into cloud platform. Finally we implemented this model using Microsoft Azure VM as cloud computing platform, power of processing and storage of GIS applications have been done to enhance management GIS system.

There are still difficulties that needed to be enhanced in this paper such as applying, improving the proposed model to support raster data of GIS system. Furthermore using full computing services in the cloud computing infrastructure to host large volume of data as possible using Microsoft Azure storage services. Traccar Web UI as GIS platform is one of the Successful experiences can be used with Microsoft Azure and it is not very costly to be deployed.

References

- 1. Murugesan, S., Bojanova, I.: Encyclopedia of Cloud Computing. Wiley, New York (2016)
- 2. Pourabbas, E.: Geographical Information Systems: Trends and Technologies. CRC Press, Boca Raton (2014)
- 3. Zhu, X.: GIS for Environmental Applications: A Practical Approach. Routledge, London (2016)
- 4. Rafaels, R.: Cloud Computing: From Beginning to End. CreateSpace Independent Publishing Platform, Charleston (2015)
- 5. Zheng, D.: Future Intelligent Information Systems, vol. 1. Springer, New York (2011)
- Furht, B., Escalante, A.: Handbook of Cloud Computing. Springer Science & Business Media, New York (2010)
- 7. Vinodkumar, T.M.: Geographic Information System for Smart Cities. Copal Publishing Group, Ghaziabad (2016)
- Mell, P., Grance, T.: The NIST definition of cloud computing. In: Vance, T.C., Merati, N., Yang, C., Yuan, M., (eds.) Cloud Computing in Ocean and Atmospheric Sciences. Elsevier, NIST, USA (2011)
- 9. Yang, C., Huang, Q.: Spatial Cloud Computing: A Practical Approach. CRC Press, Boca Raton (2013)
- 10. Chappell, D.: Introducing Windows Azure. Microsoft Corporation, Redmond (2009)
- 11. Oracle Corporation, Database rolling upgrade using Data Guard SQL Apply, Maximum Availability Architecture White Paper (2008)
- Sleit, A., Misk, N., Badwan, F., Khali, T.: Cloud computing challenges with emphasis on Amazon EC2 and windows azure. Int. J. Comput. Netw. Commun. (IJCNC) 5(5), 35–40 (2013)
- 13. Ivan, I., Singleton, A., Horák, J., Inspektor, T.: The Rise of Big Spatial Data. Springer, Ostrava (2016)
- Bhat, M.A., Shah, R.M., Ahmad, B., Bhat, I.R.: Cloud computing: a solution to information support systems (ISS). Int. J. Comput. Appl. 11(5), 5 (2010). (0975–8887)

New Real Time Cloud Telemedicine Using Digital Signature Algorithm on Elliptic Curves

Asma Jebrane^{$1(\boxtimes)$}, Naima Meddah², Ahmed Toumanari², and Mohamed Bousseta¹

¹ Faculty of Sciences, Ibn Zohr University, B.P 8106, 80000 Agadir, Morocco jebrane.asma@gmail.com
² National School of Applied Science, Ibn Zohr University, 80000 Agadir, Morocco

Abstract. In order to help doctors to communicate in real time with patients, to better diagnose their problems and protect the confidentiality of medical information. We propose a new real-time cloud telemedicine based on voice over IP protocol (VoIP). A prototype based on elliptic curve and digital signature has been developed to confirm a secure encryption scheme.

Keywords: Cloud voice over IP \cdot Elliptic curve \cdot Digital signature \cdot Telemedicine

1 Introduction

Cloud Computing is a promising paradigm that provide everything as service (XaaS), like SaaS (software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service), CaaS (Communication as A service) [3]. Medical information requires huge amount of data storage. Thus, the cloud computing based healthcare service is able not only to provide more economical and secure data storage, but to extend global medical data sharing. The data in the cloud are then accessed by doctors, nurses, care-takers and also by other hospitals, by this way patients can have better care at low cost. Voice over IP protocol provide to cloud telemedicine more efficiency, that achieves to doctors a communication in real time with the patient to better diagnose their problems and to provide better aid.

The patients and authorized personnel could access, retrieve and decrypt the encrypted data via the Internet. To confirm a secure encryption scheme, a new prototype based on elliptic curve and digital signature has been developed. The proposed scheme comprises four phases: initialization phase, registration phase, authentication phase, and the password change phase.

The paper is structured as follow. The next section presents some basic preliminaries and notations used in this paper. Section 3 provides details of framework of our proposed approach. The conclusion and future work are given in Sect. 4.

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_29

2 Preliminaries

In this section, we define real-time cloud and the Benefits of Cloud based Telemedicine. Then, we briefly review the fundamentals of Elliptic curve ECC and digital signature algorithm on ECC (ECDSA).

2.1 Real Time Cloud

Cloud computing has become one of the fastest growing paradigms in the computer science, By adopting Cloud services, companies and simple users are enabled to externalize their hardware resources, services, applications and their IT functions. Cloud VoIP (Voice over IP Protocol) is bundled with traditional VoIP and data where the user can directly communicate in real time with the other user. The user can even communicate with other people from the real-time cloud via E-mails, Texts, Voice and Video Calls. Figure 1 shows the general VoIP architecture.



Fig. 1. Cloud VoIP architecture [1]

The elements of communication infrastructure connect phones remotely through the Internet. The servers use software to emulate a telephone exchange.

The voice nodes are virtual machines that execute a variety of services (call transfer, voice mail, voice, etc.). This solution reduces cost and adds new features.

2.2 Benefits of Cloud Based Telemedicine

Telemedicine enables patients to be remotely monitored or taken care of at a distance by the usage of information communication technologies. Telemedicine connects patient and specialized doctors remotely and also allows them to share the sensitive medical records. Cloud based telemedicine comes with lots of benefits such as: [2]

- Affordability: Patient need not travel over a long distance.
- Accountability: Cloud stores all the medical records which can be accessed from anywhere and anytime.
- Availability: Patient and doctor always need not be available simultaneously. In the absence of doctor, an attendant can record the patients details.
- Accessibility: Since the application is developed in regional language, rural people will not hesitate to use.
- Assistance: Doctors can be assisted by logging the health issues, way by which the patient treated and recovered. Similar health problems can be solved by them easily.
- Assured Quality: Patients from rural place also have the chance to obtain second opinion by consulting with the experts.

There are two major methods of telemedicine; real-time and store-and-forward. The two methods are applied to different purposes depending on the patients disease conditions. Real-time telemedicine allows simultaneous both way communications between the patient and healthcare provider. The patient sends medical information to healthcare provider almost instantaneously. In contrast, store-and-forward allows non-simultaneous communications between patient and healthcare provider [4].

Real time mode add more advantages and functionality to Cloud Telemedicine as:

- Patients at a remote site can be diagnosed by the healthcare provider in real time basis.
- Healthcare provider can make immediate decisions and requests at the time of the diagnosis session with patients.
- More effective in terms of consultation and patient satisfaction than storeand-forward telemedicine.

2.3 Elliptic Curve Cryptography (ECC)

An elliptic curve is a cubic equation of the form: $E: y^2 + axy + by = x^3 + cx^2 + dx + e$, where a, b, c, and e are real numbers. Let F_p denote the finite of points where p is a large prime number and containing x, y. We focus on the finite field of ECC, the mathematical equation of ECC to be of form:

$$y^2 = (x^3 + ax + b)modp$$

Where $a,b \in F_p$ satisfying $(4a^3 + 27b^2)modp \neq 0$.

The arithmetic of elliptic curve discrete logarithm problem (ECDLP) is given points Q and P. Where Q, $P \in F_p$, and compute $Q = \alpha \times P$ it is hard to determine α given Q and P. In view of shortness, we omit the details and refer to [5, 6].

A key exchange can be accomplished as follows:

Algorithm 1. Generation keys with Elliptic Discreet Logarithm

Inputs : p, E, P, n Generate public parameters **Outputs** : public key (Q) and private key d **Begin** :

- 1. Select a random d in the interval [1, n-1]
- 2. Compute Q = dP
- 3. Return (Q, d)

2.4 Digital Signature Algorithm Using Elliptic Curve (ECDSA)

The procedure for generating signature using the ECDSA is presented below, we refer to [7]. To sign a message m an entity A with domain parameters (E, P, n, a, b, h) and associated key pair (d, Q) does the following:

Algorithm 2. Digital signature using elliptic curve

Inputs : (p, E, P, n), d and m public parameters, private key and message m **Outputs** : (r, s) signature

Begin :

- 1. Select $k \in [1; n 1]$.
- 2. Compute kP of coordinate (x_1 ; y_1), Let $r = x_1$. if rmodn = 0 return to step 1
- 3. Compute k^{-1} modn
- 4. Compute h = H(m)
- 5. Compute $\mathbf{s} = \mathbf{k}^{-1}(\mathbf{h} + d\mathbf{r})$. If $\mathbf{s} = 0$, return to step 1
- 6. Return (r, s)

3 Framework of the Proposed Approach

In this section, we propose a new real-time cloud using a secure encryption scheme based on ECDSA [8].

The public key is computed directly from the signature of third trust party (TTP) on the user's identity. The notations adopted through this paper are summarized in Table 1.

_	
Notation	Definition
IDi	Identity of the user
PWi	Password of user
ТА	Trust authority
$(s_T; PK_T)$	Key pair of TA
Р	A generator point with order n over $E_p(a; b)$
h(.)	Secure one way hash function
←	Concatenation operation

Table 1. Notations used in this paper

3.1 Proposed Scheme

The proposed architecture of system based on voice over IP protocol (VOIP) using Asterisk servers. As shown in the Figure we define the following entities:

- **TA**: Trusted Authority: organization expected to be responsible for supervising exchange among HCP.
- Registered user (Patient): Patient who is registered to the Trusted Authority.
- **Cloud VoIP**: it provides Voice and Video Calls, E-mails, Texts, after a proper authentication.
- Data Access Requester: general users of health information, like a doctors, nurses, physician

In this approach, the Cloud is divided in two part one for data storage and the other for communication in real time. The voice nodes are operated by servers Asterisk, each



Fig. 2. The proposed scheme

node has Asterisk running processes, and they are grouped in different data centers. This helps the doctors to communicate directly with the patients to better diagnose their problems. Even in the absence of doctor, patient can record his details, the recorded details can be stored at the cloud and the doctor can listen his patients details (Fig. 2).

3.2 Initialization Phase

We define a trusted authority (TA) to issue the private keys to the entities or users. All entities have agreed upon a high elliptic curve E defined over a finite field which is used with a base point generator P of prime order n. TA selects a random number $s_T \in Z^*$ as his private key, and then computes his public key PK_T ($PK_T = s_T \times P$). Then TA keeps s_T and publishes the system parameters (PK_T , P, n, G, h).

Algorithm 3. General procedure of public and secret key generation

Inputs : ID, PW, K **Outputs** : public key P_{key} and private key S_{key} **Begin** :

- 1. User U_i selects a random k_i
- 2. Computes $K_i = k_i \times P$
- 3. Sends (ID_i, K_i) to TA
- 4. TA chooses a random number t
- 5. Compute $x = t \times P$
- 6. Compute the signature parameters (r_i, s_i) as follows:
 - a) $r_i = xmodn$
 - b) $\mathbf{s}_i = \mathbf{t}^{-1}(\mathbf{h}(\mathbf{I}\mathbf{D}_i \parallel \mathbf{R}_i) + \mathbf{s}_T \cdot \mathbf{r}_i)$
 - c) $R_i = r_i \times K_i$
 - d) $S_i = s_i \times x$
- 7. TA sends $(\mathbf{R}_i, \mathbf{S}_i)$ to the user
- 8. User computes his secret key $S_{key} = S_i \times k_i$
- 9. public key $P_{key} = S_{key} \times P$

3.3 New Patient Registration Phase

When a new patient (Bob) wants to access to the cloud, he performs the following process with TA to complete the registration process:

Algorithm 4. New patient Registration phase

Inputs : ID_B , PW_B , K_B

Outputs : (PK_B) , s_B public and secret key

Begin :

- 1. Bob selects her identity ID_B , and her password PW_B freely
- 2. Bob chooses a random number k
- 3. Compute $K_B = k \times P$
- 4. Sends $\{ID_B, PW_B, K_B\}$ to TA over a secure channel
- 5. TA generates a random value t
- 6. Computes $x = t \times P$
- 7. Computes the signature parameters (r_B, s_B)
- 8. Computes the following parameters using her secret key s and the received message from Bob:
 - a) $\overline{s}_B = s_B \times x$
 - b) $R_B = r_B \times K_B$
 - c) $c_1 = h(PW_B || R_B)$
 - d) $m_B = \overline{s}_B + c_1$
 - e) $PK_B = [h(ID_B||R_B).P.K_B + PK_T.R_B]$
- 9. TA personalizes a smart card with the secret parameters (*m_B*, *R_B*, *PK_B*, *K_B*)
- 10. Delivers this smart card to Bob through a secure channel
- 11. Upon receiving the smart card, Bob inputs (ID_B , PW_B , k) and the smart card computes:
 - a) $\overline{s_B} = m_B c_1$
 - b) $s_B = \overline{s_B} \cdot k$
 - c) Check if $(s_B \times P)$ is equal to PK_B
 - d) if yes update $m_{\rm B}$ to the new value $m_{\rm B}^{\rm MW} = m_{\rm i} \times k$

3.4 New Doctor Registration Phase

When a new doctor (nurse, physician...) wants to access on the cloud, he must authenticate and granted access permissions. The doctor first selects his identity ID_D , and his password PW_D and performs the seem process with TA as the patient, to complete the registration process. In the end of registration phase the doctors return his public key and compute his secret key. The TA publishes all doctors' public keys. The corrupted TA cannot obtain the long-term private keys of user clients (doctors and patients).

3.5 Authentication Phase

Assume that Bob wants to communicate with a doctor he must enter his user-name ID_B and his password PW_B .

The patient encrypts the PHI data and the pair (R_B, K_B) by using the public key of the specific doctors, nurse or physician. After encryption the patient sends the data to the specific doctors. The doctors upload the message and decrypt encrypted data using his secret key.

Assume that the doctors want to response Bob, he can use the pair (R_B , K_B) included in the message of Bob and compute the public key of Bob (PK_B) with the following equation: $PK_B = [h(ID_B \parallel RB) \cdot P \cdot K_B + PK_T \cdot R_B]$ after that he can encrypt data using the public key of Bob.

3.6 Password Changing Phase

During the password changing phase, users U_i (patient, doctors) can change his PW_i freely and securely, without any interaction with TA. The smart card can change password itself after performing the following steps:

Algorithm 5. Password changing phase

Inputs : *ID_i*, *PW_i* original password and identity
Outputs : *PW^{new}_i* new password
Begin :

Compute s_i = m_i - h(*PW_i* || *K_i*)
Check whether (*s_i* × P) is equal to *PK_i*If not, the request is rejected. If yes next step

- 4. User enter PW_i^{new}
- 5. Compute $m_i^{new} = s_i + h(PW_i^{new} \parallel K_i)$
- 6. Password is updated successfully
- 7. Return(Q, d)

4 Conclusion and Future Work

In this paper, we proposed a new cloud telemedicine using Voice over IP protocol to ensure a fine access control in real-time telemedicine. We designed and presented a prototype of a secure healthcare system by using digital signature algorithm and elliptic curve. Future work, we will look into security analysis and computational cost of the proposed scheme.

References

- 1. Corts-Mendoza, J.M., Tchernykh, A., Drozdov, A., Bouvry, P., Simionovici, A.M., Kliazovich, D., Avetisyan, A.: Distributed adaptive voip load balancing in hybrid clouds (2015)
- 2. Jeyanthi, N., Thandeeswaran, R., Mcheick, H.: SCT: secured cloud based telemedicine, pp. 1–4. IEEE (2014)
- 3. Mahmood, Z.: Cloud computing: characteristics and deployment approaches. In: 2011 IEEE 11th International Conference on Computer and Information Technology (CIT), pp. 121–126. IEEE (2011)
- Tan, Y.-L., Goi, B.-M., Komiya, R.: Real-time/store-and-forward telemedicine with patients data protection by KP-ABE encryption. In: The International Conference on E-Technologies and Business on the Web (EBW2013), pp. 79–84. The Society of Digital Information and Wireless Communication (2013)
- 5. Koblitz, N.: Elliptic curve cryptosystems. Math. Comput. 48, 203-209 (1987)
- Arshad, N.I.R.: Elliptic curve cryptography based mutual authentication scheme for session initiation protocol. Multimed. Tools Appl. 66, 165–178 (2013)
- 7. Johnson, S.V.D., Menezes, A.: The elliptic curve digital signature algorithm (ECDSA). Certicom Research, Canada (1999)
- Jebrane, A., Toumanari, A., Meddah, N., Bousseta, M.: A new efficient authenticated and key agreement scheme for SIP using digital signature algorithm on elliptic curves. J. Telecommun. Inf. Technol. 2, 62–68 (2017)

Scalable Lightweight ABAC Scheme for Secure Sharing PHR in Cloud Computing

Naima Meddah^{1(云)}, Asma Jebrane², and Ahmed Toumanari¹

¹ National School of Applied Sciences, ENSA, Laboratory of Engineering, Sciences and Technology Information, LISTI, Ibno ZOHR University, 80000 Agadir, Morocco n.meddah@gmail.com
² LETSMP, Faculty of Sciences, Ibn Zohr, 80000 Agadir, Morocco

Abstract. Data access control is of critical importance in cloud computing, in particular for e-health systems, where a patient Personal Health Records (PHR) data, have a serious privacy concerns about outsourcing to the cloud servers. Presently Key policy attribute based encryption (KP-ABE) is promising advanced cryptographic system for fine-grained access control in cloud computing systems. Yet, Existing access control schemes based on attribute based encryption (ABE), are no longer applicable due to the heavy cryptographic computation and communication overhead of key management. Existing ABE schemes are based on expensive bilinear pairing that make its not scalable and not suitable for cloud e-health systems. In this paper we propose a new Scalable lightweight LKP-ABE scheme based on elliptic curve integrated encryption scheme (ECIES), The best known encryption scheme based on Elliptic Curve Cryptography, applied in e-health system, in order to ensure fine grained access control and data confidentiality of personal health records, and present an advanced secure and scalable encryption/decryption system based on Key Policy Attribute Based Encryption (KP-ABE) for PHR's. our scheme provide semantic security against chosen cipher-text attacks (CCAs), guaranteed resistance collusion and provide hight level data confidentiality of sharing PHR, by using elliptic curve integrated encryption scheme (ECIES) that has much stronger bit security than RSA as well as other exponential-based public key algorithm and the advanced attribute based encryption KP-ABE. The proof security, performance comparison among LKP-ABE and related schemes is given to prove the performance, low cost communication and execution efficiency of LKP-ABE.

Keywords: Cloud computing · Access control · Attribute based encryption (ABE) · Personal health records (PHR) · Elliptic curve cryptography (ECC) · ECIES · Chosen Cipher-text attacks (CCAs)

1 Introduction

The term of personal health record (PHR) has undergone substantial changes along with the emergence of cloud computing. The ultimate goals of e-health system's is to deliver best possible health services by using Body Sensor Network (BSN) [1]. The

[©] Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_30

body sensor network has brought forth to the advancement of today's medical industry, called as e-health systems [2]. Many research works have been carried out to counteract the security issues in personal health records and ensure fine-grained access control for PHR's [3–6]. Cloud computing is an emerging paradigm that provide technology and computer resources as a service, which has potential to bring great benefits to information and communication technology (ICT)for companies, organizations and even countries [7]. Promising as it is, this paradigm also brings forth new challenges issues for data security and privacy when users outsource sensitive data for sharing on cloud servers.

Due to cloud storage services requirement, fine-grained access control is needed to provide users permission management when dealing with complex users hierarchy and large amount of data [8]. Beside access control challenge in cloud-based services, we face another challenge, which is the surge in attacks such as chosen cipher-text attacks (CCAs). Then the confidentiality of the data becomes a problem. At present, encryption is the primary mechanism to implement data protection [14]. However, traditional encryption schemes are not suitable in such a situation for their lack of ability in access control and huge overhead of key management. Aiming at providing fine-grained access control over encrypted data, Sahai and Waters (2005) made some initial steps to solving this problem by introducing the concept of attributed-based encryption (ABE) [9]. In KP-ABE scheme there is an authority that is responsible for attribute management and key distribution [6]. The authority can be social security administration, the health service provider (HSP)... [2] the data owner encrypt the data under a set of attributes and each user defines the access policies, each user can decrypt the cipher-text if and only if its attributes satisfy the access policy. The feature of ABE that any user can decrypt the cipher-text as long as its meet the required attribute, which makes it very suitable for fine-grained access control. However it very difficult to implement the existing ABE schemes in cloud computing, because they are all based on the expensive bilinear pairing operation [10]. In order to ensure fine-grained access control based on ABE and keep data privacy and confidentiality in cloud computing by reducing communication overhead, a lightweight attribute based encryption is crucial. In this paper, considering the fact that ECC algorithm has much stronger bit security than RSA as well as exponential-based public key cryptography algorithm [11], we propose a first KP-ABE based on ECIES scheme, to address data privacy and security issues, and fine-grained access control in cloud e-health. The new KPL-ABE scheme based on no-pairing ECC, replaces the expensive bilinear pairing operation with point scalar multiplication on elliptic curve, it can address the lightweight attribute access control in cloud Computing e-health systems. The main contributions in this work is summarized as follow:

– To the best of our knowledge, the first lightweight LKP-ABE based on ECIES that can be applied in cloud e-health, is p roposed to ensure fine-grained access control, data privacy and privacy issues for PHR's in cloud e-health.

- The proposed scheme ensure security semantic against CCAs (Chosen Cipher-Text Attacks) and CCP (Chosen Plain-Text Attacks).
- The security proof is performed in attribute based selective-set model.
- The new LKP-ABE scheme's security depends on the ECDDH (Elliptic Curve Decisional Diffie-Hellman) problem instead of bilinear Diffie-Hellman assumption, which can diminish the computation and communication overhead.
- The lightweight feature is proved by comparing it with the existing ABE schemes, which indicate that can be applied in cloud e-health.

The rest of this paper is organized as follows: In Sect. 2, we examine the related works on attribute based access control in Cloud computing. In Sect. 3 we give the necessary technique preliminaries pertaining to our proposed scheme. Section 4 we present the description of the LKP-ABE scheme for cloud computing. The security proof is made in Sect. 5, and the performance analysis is discussed in Sect. 6. Finally draws a conclusion.

2 Related Work

This paper is primarily related to works in cryptographically ensuring fine-grained access control for outsourced data and attribute based encryption. Attribute based encryption (ABE) is a promising technique that designed for fine-grained access control for encrypted data. The notion of attribute based encryption was first introduced by sahai and waters [12]. It's an extension of Identity-based Encryption (IBE), which can ensure fuzzy identity by combining the user's identity with a series of attributes and achieve data preserving [1, 13]. The ABE based access control has attracted many researchers attention and a myriad of data access control based on ABE [7, 10, 15–17], have been proposed to construct a secure fine-grained access control and data privacy in semi-trusted cloud computing. Yet, based on the Delov-Yao model [18], security goals such as chosen cipher-text attacks (CCAs) resistance, data confidentiality and low cost overhead computation and execution efficiently, of the most solution designs are not perfectly guaranteed. Furthermore to face challenges in expressive access control policy, many variant of access control based on ABE have been proposed to achieve an expressive access control by constructing constant size cipher-text [6, 10, 15, 19] to limit the size of the cipher-text. Though some achievements have been made at the expense of increasing overheads, they do not meet the lightweight requirement of Cloud Computing yet. The research community is mostly focused on the design of fine-grained access control for encrypted data which can applicable and suitable for cloud environments without trying to minimize the high computational complexity of attribute based encryption systems [20]. Our contribution is to construct a new lightweight LKP-ABE based on ECC, for fine-grained access control with chosen cipher-text security and resistance collusion with minimal cryptographic computation that can be applied in cloud e-health systems.

3 Preliminaries

Our scheme is based on elliptic curve cryptosystem and its related primitives, key-policy attribute based encryption, secret sharing shamir scheme SSSS [9] and access structure in ABE. The main notation used in our paper are summarize in Table 1.

Notation	Notations meaning
q	A large prime, the finite field with p elements F_{-p}
\mathbb{E}	An elliptic curve over the finite field F_{-p} , of large prime order q
q	A large prime, used to denote the order of G in \mathbb{E} over F_{-p}
\mathbb{Z}_q	A finite integer field, whose elements set is $\{0, 1,, q - 1\}$
G	A base point on the elliptic curve \mathbb{E}
\mathbb{G}_E	A subgroup of \mathbb{E} with the order of q
0	The zero element of an elliptic curve group
MAC()	Message Authentication Code
Hash()	A hash function, used within the KDF and the MAC functions
MSK	The master private key of the ABE scheme
MPK	The master public key of the ABE scheme
τ	The access policy
PPrm	The public parameters of the ABE scheme
$\mathbb{E}NC$	A symmetric encryption algorithm, using KDF
DEC	A symmetric decryption algorithm
k	The number of the attributes used to encrypt data
n	The number of the attributes in a system
P_S	One point scalar multiplication

Table 1. Notations used in our paper

3.1 Elliptic Curve Cryptography

Elliptic curve cryptographic ECC schemes are public-key mechanisms that provide encryption, digital signature and key exchange capabilities, it was introduced in 1985 by Victor Miller and Neal Kobliz [11],whose security relies on the elliptic curve discrete logarithm problem (ECDLP) problem. The base of ECC operations is finite field (Galois Field) algebra with focus on prime Galois Fields F_{-p} or binary extension Galois Fields $F_2^m.Z_p$ is a prime Galois Field F_{-p} , and an Elliptic curve over Z_p is defined by a cubic equation $y^2 = x^3 + a.x + b$ and can be described by a set of parameters (q, a, b, G, p), where (a, b) $\in Z_p$, and $4a^3 + 27b^2 \#(modp)$ The operations in ECC consist of basic prime field operations and point operation. The former operations are simple, the latter operation refers to point scalar multiplication, which can be further refined by point add and point double operations. The point scalar multiplication is the fundamental and most time-consuming operation in ECC [11, 20]. ECC security is based on ECDH and actually no algorithm is known that solves the ECDLP in an efficient way. Moreover, the ECDLP is regarded as the hardest of these three problems. From this fact derives one of the most important benefits of ECC: the key size. Keys in ECC are significantly shorter than in other cryptosystems such as RSA [11]. A shorter key implies easier data management, lower hardware requirements (in terms of buffers, memory, data storage, etc.), less bandwidth when transmitting the keys over a network [11, 21].

3.2 Elliptic Curve Integrated Encryption Scheme

The most extended encryption and decryption scheme based on ECC is the Elliptic Curve Integrated Encryption Scheme (ECIES) [11]. This scheme is used to ensure data confidentiality and data integrity for each users, which uses the following functions [11]:

- Key Agreement (KA): Function used for the generation of a shared secret by two parties.
- Key Derivation Function (KDF): Mechanism that produces a set of keys from keying material and some optional parameters.
- Encryption (ENC): Symmetric encryption algorithm.
- Message Authentication Code (MAC): Data used in order to authenticate messages.
- Hash (HASH): Digest function, used within the KDF and the MAC functions.

The encryption and decryption algorithm of ECIES is described in Figs. 1 and 2.



Fig. 1. ECIES encryption diagram

Fig. 2. ECIES decryption diagram

3.3 Access Structure

The access structure defines the access policy in the attribute based encryption scheme the access structure can be defined as Definition 1 [17].

Definition 1. Let $\{P_1, P_2, \ldots, P_n\}$ be a set of parties or attributes. A collection $\mathbb{A} = 2^{\{P_1, P_2, \ldots, P_n\}}$ is a monotone if $\forall B, C$: if $B \in \mathbb{A}$ and $B \subseteq C$ then $C \in \mathbb{A}$. An access structure (respectively, monotone access structure) is a collection (respectively, monotone collection) \mathbb{A} of non-empty subsets of $\{P_1, P_2, \ldots, P_n\}$, i.e., $\mathbb{A} \subseteq 2^{\{P_1, P_2, \ldots, P_n\}} |\{\emptyset\}$. The sets in \mathbb{A} are called the authorized sets, and the sets not in \mathbb{A} are called the unauthorized sets.

3.4 Secret Sharing Scheme

The secret sharing refers to distributing a secret amongst a group of parties. The first secret sharing schemes were proposed by Shamir [8] who employs the fact that an n degree polynomial can be determined by $P(x)_n = \sum_{k=1}^n f(x)_k y_k$ in the case that (n + 1) points are given, where f(x) is called Lagrange coefficient. Assume that there are (n+1) points (x_i,y_i) for $i = \{0, ..., (n-1)\}, f(x)_k$, can be computed by $f(x)_k = \prod_{j=0, j \# k}^n \frac{X-X_i}{X_k-X_j}$. The method to construct the polynomial is called Lagrange interpolation, the polynomial $P(x)_n$ is called Lagrange interpolation polynomial. In order to share the secret a_0 among n parties and reconstruct it at least by r parties, a(r-1) degree polynomial $P_{(r-1)}(x) = \sum_{i=0}^{r-1} a_i x_i$ should be defined, where, the coefficients $\{a_1, a_2, ..., a(r-1)\}$ are assigned as random values in a finite field. The value $P(ID_i)$ is a share of the secret and given to user i. When t two-tuples $(ID_i, P(ID_i))$ are put together, the polynomial $P_{(r-1)}(x)$ can be reconstructed by Lagrange interpolation, and the secret a_0 can be accordingly determined [8].

4 Proposed Scheme

Our proposed lightweight ABE is KP-ABE single authority, which implies a single certificate authority that is responsible for key generation for attributes and key management. In this section, we describe the system architecture model, system overview and highlight, and algorithm of LKP-ABE.

4.1 Key-Policy Attribute Based Encryption Scheme

KP-ABE encryption scheme is composed of the following four algorithms, namely, Setup, Encrypt, KeyGen, and Decrypt [22].

- Setup: The Setup algorithm is a randomized algorithm, which is run by the CA and outputs the master public key MPK and the master secret key MSK. The master public key and set of parameters are published and the master key secret is kept secret by the certificate authority. Also CA define the attribute universe.
- Encrypt (M, μ , MPK): The Encrypt algorithm is also a randomized algorithm, which is run by the sender and outputs cipher-text CM by taking the message M to be encrypted, the attributes set μ that the data user should satisfy, and the master public key MPK as input.
- **KeyGen** (τ , **MSK**): The Key-Generation algorithm is a randomized algorithm too, which is run by the CA and takes an access structure τ and the master secret key MK as input. It outputs the decryption key D corresponding to the access structure.
- **Decrypt (CT, D, MPK):** The Decrypt algorithm is run by the receiver, which takes the cipher-text CT encrypted under the attributes set μ , the decryption key D for access control structure τ , along with the master public key MPK as input. If $\tau(\mu) = 1$, it decrypt the cipher-text CT and outputs message M.

4.2 System Model

Our system model is defined as following:

Health Care Staffs (HCS): are the data users in e-Health system. Each data user has a set of attributes, such as affiliation, department and type of health-care staff, and is authorized to search on encrypted PHRs based on his set of attributes or access policy.

CA: Certificate authority which is the fully trusted authority in the system. It sets up the system by defining the attribute universe and validate the registration of all users and HCS's. For each registered user in the system, the CA assigns a global unique identity and generates a pair master secret key MSK and master public key MPK and also a set of public parameters PPrm.

Data owner (Patient): define the access policy τ , and encrypt data under the access policy and upload the cipher-text to the cloud servers if he use the CP-ABE, or he use the set of attribute to encrypt her vitals signs by using KP-ABE.

CSP: Cloud Service Provider, cloud servers store the patient's data (PHR) and provide access service to users by contacting CA and validate the global unique identity of all user issuing by CA.

Health insurance and **Health–care** are added to complete the whole system, and ensure the general communication and sharing the patient PHR's between the different parties of e-health systems (Fig. 3).



Fig. 3. System model of lightweight ABAC for secure sharing PHR in Cloud Computing.

4.3 The Proposed Lightweight KP-ABE Scheme

The constructed lightweight KP-ABE is based on elliptic curve cryptography. The ECC scheme is defined by set of parameters (q, a, b, G, p), it assumed that all parameters are secure enough to satisfy the requirement of the applications. The framework of lightweight KP-ABE based on ECC mainly consists of five phases: CA-Setup, User registration, Encryption, key-generation and Decryption-Alg. The whole system can be defined with following steps:

CA-Setup: The certificate authority initialize the system by running the CA-algorithm: Define the attribute universe: $U = \{1, 2, \dots, n\}$. Choose a random $r_i \in \mathbb{Z}_q^*$ and assign for each attribute $i \in U$. the public key of each attribute is $Pk_i = r_i.G$ where $G \in \mathbb{E}$ (elliptic curve). Also choose a random $r \in \mathbb{Z}_q^*$ to define the: Master secret key, MSK = r Master public key, MPK = r.G, the public parameters are: $PPrm = \{MPK, Pk_1, Pk_2, \dots, Pk_U\}$.

Registration: Each user must send the set of legal information (set of attribute) to the Certificate authority, the CA assign the global unique user identity (G_{uid}) for each legal user and a pair of secret and public key.

Encryption: Here we have two possibilities, we can encrypt the message M by using the lightweight KP-ABE, or we can use ECIES to encrypt the message M, and use the LKP-ABE to encrypt the secret sharing.

For more lightness and more security we use the method 2.

The owner use ECIEC to encrypt the Message M, see the Fig. 1

Encrypt the secret sharing by LKP-ABE, following the steps:

Specify the set of attributes μ

Choose a random $k\in\mathbb{Z}_q^*$ and compute $C_i=k.PK_i/PK_i:$ public key for each attribute, $I^in\mu$

For $k \in \mathbb{Z}_q^*$ compute C' = k.MPK

Use KDF function we have $C' = k.MPK = (K_x, K_y)$

Let $K_{\boldsymbol{x}}$ be the encryption key and $K_{\boldsymbol{y}}$ the integrity key for the message, here we have M= secret sharing

$$C_{SP} = ENC(SP, K_x), MAC_{SP} = Hash(SP, K_y).$$

$$\text{Output } CT_{SP/KP-ABE} = (\mu, C_SP, MAC_{SP}, C_i, i \in \mu) \tag{1}$$

Finally output the cipher-text

$$CT_{LKP-ABE} CT_{LKP-ABE} = (CM_{ECIES}, CT_{SP/KP-ABE}) CT_{LKP-ABE} = (CM_{ECIES}, \mu, C_{SP}, MAC_{SP}, C_{i}, i \in \mu)$$
(2)

Key-Generation (τ , **MSK**): This algorithm is used by the certificate authority to generate the decryption key of message under set of attribute μ if and only if $c(\mu)=1$ This algorithm output a private key (decryption key) embedded with an access structure

 τ . The decryption key must satisfy $\tau(\mu) = 1$. For this purpose we should define an access structure by following steps:

Each non-leaf node is defined as a Shamir Threshold Scheme. Set the degree of secret polynomial $P_u(x)$ to be $d_u = k_u - 1$.

For root node r, set $P_r(0) = s$. And randomly choose d_r element in \mathbb{Z}_q^* to completely define P_r . For any other non-leaf node u, set its secret to be one secret share of its parent node, that is $P_u(0) = P_{parent(u)}(index(u))$. And randomly choose d_u element in \mathbb{Z}_q^* to completely define P_u .

- For each leaf node u, assign the following value $D_u = P_u(0)/r_i$, i = attrib(u), $r_i \in \mathbb{Z}_q^*$ is randomly chosen in CA-Setup. The decryption key can denoted by $D = \{\tau, D_u = P_u(0)/r_i, i = attrib(u), i \in \mu\}.$

Decryption-Alg: To decrypt the Message, we first begin by decrypting the secret sharing encrypted under LKP-ABE, and use the secret sharing to decrypt the cipher-text under ECIES. Decrypt $CT_{SP/LKP-ABE}$ Similar to all ABE schemes:

 $\begin{array}{l} - & \mbox{For each non-leaf node } u, \mbox{ all nodes } v, \mbox{ are children of } u. \mbox{ Let } \mu_u \mbox{ be an arbitrary } d_u - \\ & \mbox{sized set of child nodes } u \mbox{ such that DecryptNode} \{\mbox{CTSP/LKP-ABE,D}, u) \mbox{ a } \mu_u, \mbox{ DecryptNode } \mbox{CTSP/LKP-ABE,D}, u) = \bot, \mbox{ in other, DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = \sum_{v \in u} .(\Delta_{i,u'}(0)). \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, v) \mbox{ where } i = \mbox{ index}(v), \mbox{ } U'_u = \mbox{ index}(v), v \in U_u \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = \sum_{v \mid inu} .(\Delta_{i,u'}(0)).\mbox{P}_u(0).\mbox{ k.G DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = \sum_{v \mid u} .(\Delta'(0)).\mbox{P}_p \mbox{ are } (v)).\mbox{ k.G DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = \sum_{v \in u} .(\Delta'(0)).\mbox{P}_p \mbox{ are } (v)).\mbox{ k.G DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } u) = \sum_{v \in u} .(\Delta_{i,u'}(0)).\mbox{P}(u)(i).\mbox{ k.G DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ becryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ becryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ DecryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ becryptNode } (\mbox{CT}_{SP/LKP-ABE}, D, u) = P(u) \\ mbox{ } (0).\mbox{ k.G } \mbox{ becryptNode } (\mbox{ conde }$

Consequently, for the root node R of access tree, there should be DecryptNode $(CT_{SP/LKP-ABE}, D, r) = P(r)(0).k.G = r.k.G = (K'_x, K'_y)$. Here, K_x is considered to be the decryption key for the message CT and K_y is the integrity key for the message. The encrypted Message M' can be decrypted by $M' = DEC(C, K'_x)$. if Hash $(M', K'_y) = MAC(M)$, it indicates that message M is correctly decrypted and is not tampered. That is to say that the correctness, integrity and authenticity are all verified by MAC(M).

- for a leaf node u,

- DecryptNode(CT_{SP/LKP-ABE},D,u) =
$$\begin{cases} D_u.C_i.r_i^{-1}.P_i\\P_u(0).r_i^{-1}.k.r_i.G\\P_u(0).k.G(ieu)\\ \perp Otherwise \end{cases}$$

N.B: the output of DecryptNode $(CT_{SP/LKP-ABE}, D, u)$ is an element in elliptic curve group \mathbb{G}_E or

Decrypt CM_{ECIES}: after decrypting the $CT_{SP/LKP-ABE}$, the user can use the secret sharing to decrypt the cipher-text by using ECIES.

5 Security Analysis

5.1 Security Model

The selective-Set Security game for a KP-ABE scheme as defined in [9] is used to prove the security of all ABE scheme's CP-ABE/KP-ABE. Key Policy Attribute-based Encryption game captures the indistinguishability of messages under Chosen Plain-text Attacks and the collision resistance of user secret keys, namely, the attackers cannot generate a new secret key by combining their secret keys. To capture the collision-resistance, the multiple secret key queries can be issued by an adversary A after the challenge phase. the game is described as indicated [9].

5.2 Security Analysis of LKP-ABE

The formal security analysis of LKP-ABE, is given to prove that our LKP-ABE can acheive, security Chosen Cipher-text Attacks (CCAs), Chosen Plain-text Attacks (CPAs), Data confidentiality under attribute security selective-set model with minimal cryptographic computation.

Theorem 1. The LKP-ABE scheme is secure under attribute selective security model. If there is an adversary who breaks our scheme in selective-set security model in polynomial-time, a simulator Λ can be constructed to solve the ECDDH problem with a non-negligible advantage.

Suppose we have an adversary A with non-negligible advantage $\xi/2$ in the selective security game, a simulator Λ can be constructed to gain the ECDDH game with advantage $\xi/2$. The constructed simulator is described as indicated in [20].

Init. The simulator Λ get the attributes set Φ , which is the adversary Θ wishing to attack upon. It is assumed that the access structure corresponding to the attributes set Θ is Γ Setup The simulator Λ runs the Setup algorithm of the proposed scheme to set the system public key parameters and sends them to the adversary Θ . According to the DDHA, Λ sets the system parameter Y = A = c.G, and sets $Y_i, \forall i \in U$ and, $c \in \mathbb{Z}_q$:

If $i \in \Theta$, Λ randomly chooses $si \in \mathbb{Z}_q$ and sets $Y_i = s_i.G$ and $Y_i = s_i$.

If $i \in (U - \theta)$, A randomly chooses $\rho_i \in \mathbb{Z}_q$, and sets $Y_i = \rho_i . B = d.\rho_i . G$ and $Y_i = d.\rho_i$

The simulator sends the system public key parameters $Y, Y_i, i \in U$ to Θ . It is obvious that Y, Y_i corresponds to PK, P_i of the proposed LKP-ABE scheme.

Phase 1. The adversary Θ can make many queries for the decryption key corresponding to any access structure, where the challenge set ρ does not satisfy γ , that is $\gamma(\rho) = 0$. The simulator Λ needs to assign a polynomial P_u with degree of d_u to every node u in the access tree Γ . The polynomial P_u for each node u in Γ is defined as $P_u(x)$ to be $p_u(x)$ in the proposed scheme. The shared secret is set to be the constant of the

root's polynomial, that is $P_u(0) = c$. The secret key corresponding to each leaf node x in Γ is given by its polynomial as follows:

$$D_x = P_x(0)/s_i$$
$$D_x = \begin{cases} \frac{p_x(0)}{r_i}, & \text{if } (attr(x) \in \theta) \\ \frac{p_x(0)}{\delta_i.d} & \text{if } (attr(x) \notin \theta) \end{cases} \text{for } i = attr(x).$$

The secret key for access structure is $(D_x, x \in \Gamma)$, which is distributed to the adversary.

Challenge: like the challenge steps in Selective-set game of ABE [17], the adversary will submit two challenge messages M_0 and M_1 the simulator Λ , who flips a fair binary coin u, and compute the cipher-text M_0 . To encrypt the message M_0 , in order to encrypt the message M_u the simulator execute the Encryption algorithm as indicated in Sect. 4.3. The cipher-text is $CT(\rho, C, MACMu_0, C_i, i \in \rho)$ and sent to the adversary more details in [18].

Phase 2. Both of the simulator and adversary Execute precisely as they did in phase 1. prediction: the

If $\mu = 0$, then Z = c.d.G. If k is set to be d, there should be, C' = k.Y = d.c.G = Z, and Ci = k.Yi = d.Yi = d. r i.G = r i.B, where, $i \in \rho$.

Chosen Cipher-text Attacks Security

Elliptic Curve Integrated Encryption Standard (ECIES), provides semantic security against an adversary who's allowed to use "chosen plain-text attack CPA" and "chosen cipher-text attacks CCA".

Theorem 2. ECIES is secure against CCA adversaries.

Proof. Indeed, let A be an adversary that computes the Diffie-Hellman function after n queries to an ECDDH oracle, with probability ξ . Then from any ρ – ECDDH oracle we can build a ρ' – ECDDH oracle, achieving $\rho' \leq \xi/2$. Therefore, if one simulates the perfect oracle, called by Adversary, using this ρ – ECDDH simulator, then A succeeds in solving the computational Diffie-Hellman problem with probability (ξ – n) $\rho' \geq \xi/2$, Still, even if bot problems are polynomially equivalent, the computational cost of the above reduction may be huge, depending on the original value of ρ [21].

6 Performance Comparison of Our Scheme

in this section, we compare the performance of our LKP-ABE scheme with the related existing schemes. All the existing ABE scheme are based on bilinear pairing group $\mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$. Since the basic operation of \mathbb{G}_1 and \mathbb{G}_2 are modular exponentiation similar to RSA system, which make the ABE scheme based on bilinear pairing group very complex with high computational overhead of encryption and decryption key. According to the survey of ECC algorithm in [11], we have one bilinear pairing is

about 20 point scalar multiplication, and one modular exponential operation is 2 point scalar multiplication.

Consequently point scalar multiplication can be token as the unit of computation overhead in our LKP-ABE scheme. based on above assumption, let δ_1 be the security level, the size of a point on elliptic curve of the 160-bit is 2 δ_1 , the size of its private key is δ_1 and the size of its public key is 2 δ_1 . Accordingly, both the public and private key size of the 1024-bit RSA are 6,4 δ_1 , and the size of an element in \mathbb{G}_1 of a RSA based ABE scheme is 6,4 δ_1 and the size of element in \mathbb{G}_2 of it is 12,8 δ_1 [20].

In our LKP-ABE scheme for securing and sharing PHR we have:

the cipher-text $CT_{LKP-ABE} = (CM_{ECIES}, \mu, C_{SP}, MAC_{SP}, C_i, i \in \mu)$, compute the length of $CM_{ECIES}, \mu, C_{SP}, MAC_{SP}, C_i, i \in \mu$:

$$C_i = k \cdot P \cdot K_i, i \in \mathbb{E}, C_i \in \mathbb{E} \Rightarrow lenght(C_i) = 2\delta_l$$

length (C_{SP}) = $\delta_l + \delta_l + k \cdot 2\delta_l \Rightarrow \text{length}(C_{SP}) = \delta_l(2k + \delta_l)$

 $\begin{array}{l} \text{length } (\text{MPK}, P_{k_1}, P_{k_2}, P_{k_3}, \cdots, P_{k_U}), \ U: \text{universofattribtes}, P_{k_i} \in \mathbb{E} \Rightarrow \text{length}(P_{k_i}) \\ = 2\delta_l \Rightarrow \text{length}(\text{PPrm}) = (2n+1)\delta_l \end{array}$

length of decryption key, $D = \gamma$, $D_u = P_u(0)/r_i$, i = attrib(u), $i \in \mu$, where γ is an access structure and u is user attributes. $\Rightarrow \text{length}(D) = k.\delta_l$

The computational overhead is measured by the costs for bilinear mapping, the public key based encryption and decryption operation .in our LKP-ABE scheme no bilinear mapping is involved. The overhead and efficiency of the LKP-ABE scheme in cloud e-health is described in the graphs Figs. 4, 5, 6 and 7.



Fig. 4. LKP-ABE computation overhead



Fig. 5. Size of LKP-ABE Public key



Fig. 6. Size of LKP-ABE decryption key



Fig. 7. Size of LKP-ABE cipher-text

7 Conclusion

In this work we proposed a new and effective lightweight data access control scheme for secure sharing personal health records PHR in cloud computing based on no-pairing attribute based encryption, which ensure fine-grained access control, Data confidentiality and secure data transfer in cloud computing . As the best of our knowledge, Our scheme is more suitable and applicable in the cloud environment - because it provide low cost communication overhead and execution efficiency- than other attribute based access control scheme applied in e-health cloud computing. The proposed scheme provide semantic security against chosen cipher- text attacks (CCAs), and resistance collusion attacks security, by using a combination of elliptic curve integrated encryption scheme (ECIES) which has much stronger bit security than RSA as well as other exponential-based public key algorithm, and key policy attribute based encryption KP- ABE which is regarded as one of the most suitable technologies for fine-grained access control in cloud environments, because it gives the data user more direct control access policies. As indicated in Sect. 7 the proposed scheme is lightweight KP-ABE for one trusted authority (CA) and very convenient for cloud computing e-health, our algorithm is highly scalable, secure and adapted to the cloud e-health system's and can be generalized for the multi-authority cloud e-health.

References

- Li, J., Chen, X., Li, J., Jia, C., Ma, J., Lou, W.: Fine-grained access control system based on outsourced attribute-based encryption. In: European Symposium on Research in Computer Security, pp. 592–609. Springer (2013)
- 2. Kobrinskii, B.A.: E-health and telemedicine: current state and future steps. E-Health Telecommun. Syst. Netw. **3**(4), 50–56 (2014)
- Li, M., Yu, S., Ren, K., Lou, W.: Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings. In: SecureComm 2010, pp. 89–106. Springer (2010)
- Li, M., Yu, S., Zheng, Y., Ren, K., Lou, W.: Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Trans. Parall. Distrib. Syst. 24(1), 131–143 (2013)
- Tan, Y.-L., Goi, B.-M., Komiya, R.: Real-time/store-and-forward telemedicine with patients data protection by KP-ABE encryption. In: The International Conference on E-Technologies and Business on the Web (EBW2013), pp. 79–84. The Society of Digital Information and Wireless Communication (2013)
- 6. Yang, Y., Liu, X., Deng, R.H., Li, Y.: Lightweight sharable and traceable secure mobile health system. IEEE Trans. Dependable Secure Comput., 11 (2017)
- Meddah, N., Toumanari, A.: Reinforce cloud computing access control with key policy attribute-based anonymous proxy reencryption. Int. J. Cloud Comput. 5(3), 187–197 (2016)
- 8. Rivest, R., Shamir, A.: How to share a secret. Commun. ACM 22(11), 612–613 (1979)
- Sahai, A., Waters, B., et al.: Fuzzy identity-based encryption. In: Eurocrypt, vol. 3494, pp. 457–473. Springer (2005)
- Odelu, V., Das, A.K., Goswami, A.: An Efficient CP-ABE with Constant Size Secret Keys using ECC for Lightweight Devices. IACR Cryptology ePrint Archive 2015, 841 (2015)

- 11. Martnez, V.G., Hernndez Encinas, L., Snchez Ávila, C.: A survey of the elliptic curve integrated encryption scheme. Ratio **80**(1024), 160–223 (2010)
- Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-policy attribute-based encryption. In: Proceedings of the 2007 IEEE Symposium on Security and Privacy, SP07, pp. 321–334. IEEE Computer Society, Washington, DC, USA (2007)
- Lewko, A., Okamoto, T., Sahai, A., Takashima, K., Waters, B.: Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption. In: Advances in Cryptology-EUROCRYPT 2010. LNCS, vol. 6110, pp. 62–91. Springer, Heidelberg (2010)
- Singh, S., Jeong, Y.-S., Park, J.H.: A survey on cloud computing security: Issues, threats and solutions. J. Netw. Comput. Appl. 75, 200–222 (2016). Elsevier
- 15. Attrapadung, N., Liber, B., de Panafieu, E.: Expressive key-policy attribute-based encryption with constant-size ciphertexts. In: PKC 2011. LNCS, vol. 6571, pp. 90–108 (2011)
- Rafath, N., Ghouri, W., Raziuddin, S.: Security in cloud using ciphertext policy attribute-based encryption with checkability. Int. J. Innov. Res. Comput. Commun. Eng. 3 (5), 4427–4434 (2015)
- 17. Zhen, Y.: Privacy-preserving personal health record system using attribute-based encryption. Worcester Polytechnic Institute (2011)
- Delov, D., Yao, A.C.: On the security of public key protocols. IEEE Trans. Inf. Theory 29 (2), 98–208 (1983)
- Odelu, V., Das, A.K., Sreenivasa Rao, Y., Kumari, S., Khan, M.K., Choo, K.-K.R.: Pairing-based CP-ABE with constant-size ciphertexts and secret keys for cloud environment. Comput. Stand. Interfaces (2016). doi:10.1016/j.csi.2016.05.002
- 20. Yao, X., Chen, Z., Tian, Y.: A lightweight attribute-based encryption scheme for the internet of things. Future Gener. Comput. Syst. **49**, 104–112 (2015). Elsevier
- Muthurajan, V., Narayanasamy, B.: An elliptic curve based schnorr cloud security model in distributed environment. Sci. World J. 2016, 18 (2016)
- Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute-based encryption for finegrained access control of encrypted data. In: Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS06, pp. 89–98. ACM, New York (2006)

IR, Big Data, Business Intelligence, and Knowledge Management

Arabic Stemming Techniques as Feature Extraction Applied in Arabic Text Classification

Samir Boukil¹, Fatiha El Adnani¹, Abd Elmajid El Moutaouakkil¹, Loubna Cherrat¹, and Mostafa Ezziyyani^{2(⊠)}

¹ Department of Computer, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco boukilsamir@yahoo.fr, ftheladnani@gmail.com, cherratloubna2@gmail.com, elmsn@hotmail.com
² Faculty of Sciences and Technologies, Abdelmalek Essaadi University,

Tangier, Morocco

ezziyyani@gmail.com

Abstract. In this paper, we conduct a comparative study about the impact of stemming algorithms, as feature extraction systems, on the task of classification of Arabic text documents. Stemming is forceful and fierce as in reducing words to their three-letters roots. Which may influence the semantics, as various words with divers implications may share the same root. Light stemming, by examination, expels oftentimes utilized prefixes and suffixes in Arabic words. Light stemming doesn't extract the root and thus doesn't influence the semantics of words. However, the result of the light stemming is not necessarily a word. For the evaluation, we used corpus contains 5,070 records that fall into six classes. A several tests were done utilizing two separate illustrations of the same corpus. The K-Nearest Neighbors (KNN) classifier was utilized for the classification task. The recall measure is used to evaluate the performance of these methods.

Keywords: Text mining \cdot Automatic language processing \cdot Classification \cdot Feature extraction \cdot Arabic language \cdot Stemming \cdot Light-Stemming \cdot K-Nearest neighbors

1 Introduction

Arabic Language is one of the widest spread languages. It is the seventh most used language in the internet [1] and the fourth in the word. It is used by 6.6% of the world's inhabitants [2] (more than 442 million Arabic speakers: 295 million as first language and 246 million as second language [3]).

The volume of information that is accessible on the Internet is expanding persistently. Thus, the complexity level of applications processing these immense amount of information is increasing. Organizing these information resources into classes are required to help the applications better do its tasks [4]. Text classification is a task of assigning one or more predefined classes to the analyzed document, based on its content. Several classification algorithms have been tested on Arabic text classification,

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_31

for instance, the Naive Bayes probabilistic classifiers [5], Decision Tree classifiers [6], Neural Networks [7], K-Nearest Neighbor classifier (KNN) [8]. and Support Vector Machines [9]. The high dimensionality of the feature space is a notable issue for some machine-learning algorithms. Since the complexity of many learning algorithms increments with increased data dimension, algorithms that can enhance the classification efficiency, by decreasing the data into smaller dimensional space, are profoundly wanted and preferred. These algorithms make the learning task of classification and information retrieval systems faster, more proficient and spare more space [10].

While classifying text document, the features does not represent semantically the document in a similar way. Some of these features might be excess and add nothing to the significance of the document, others may be synonymous and consequently choosing one of them is sufficient to increase the semantic for classification purposes. Thus, the effective determination of feature words is a critical task in text classification, this operation is called: feature extraction. Feature extraction extracts a subset of features that contain solid and informative information about the original dataset, while expelling unessential or excess features [11].

This paper present a description and a comparison of two morphologic feature extraction techniques for Arabic text classification. To be specific it's about stemming and light-stemming techniques. The KNN classifier was tested on Arabic dataset. The dataset includes 5,070 Arabic documents. The documents are physically grouped into six classes: Sports, Business, Entertainment, Middle_East, Scitech and World. The efficiency of the previously mentioned methods was measured regarding vector sizes, time to run the classifier; and the recall of the classifier.

This paper is organized as follows: Introduction has been presented in Sect. 1. Section 2 describes a state of art about feature extraction techniques and classification of Arabic text document. Section 3 resumes the structure in which the feature extraction techniques were utilized. Section 4 presents the outcomes and Sect. 5 outlines the conclusions and future work.

2 Related Works

Feature extraction techniques uses numerous systems to locate and extract the optimal and ideal subset of features. Eliminating stop words from the records is one way; calculating document frequency, information gain and other statistical characteristics is another way [12]. Other methods are based on similarity and relations between the words, such as, stemming systems that extracts the word's root and light stemming that removes the suffixes and prefixes and keep the resulting word.

All those methods of extracting information in Arabic could be categorized into two categories as in Fig. 1, based on linguistic dependencies:

- Language-independent approaches, called statistical approaches, can be categorized per similarity and dissimilarity coefficients.
- Language-dependent approaches, known as stemming/morphological approaches, are classified into two major approaches:
 - Root-based stemming approach.
 - Light stemming approach.



Fig. 1. Methods of extracting information from Arabic texts

To understand the mechanism of stemming approaches, we must understand the structure of arabic word.

The structure of an arabic word is broken down into five components: the proclitic, the prefix, the stem, the suffix, and the enclitic.

In Table 1, the agglutinated word $u_{k,\ell}$ is framed by the proclitic u (for), the prefix aigenrightarrow and the suffix <math>
u (which indicate the third person of plural), the enclitic $aigenrightarrow and the stem <math>
u_{k,\ell}$ (which). The stem (without prefix and suffix), presents the kernel vocabulary, possibly surrounded by extensions [13].

Table 1.	Word	ليراقبوهم	decortication
----------	------	-----------	---------------

Proclitic	Prefix	Stem	Suffix	Enclitic
J	ي	راقب	و	هم
For	(they)	Observ	(they)	Them

2.1 Light-Stemming

Light Stemming implies on pulling out the light-stem from the word. It is to expel prefixes, suffixes and extensions as illustrating in Tables 2 and 3:

Table 2.	Light	stemming	example	1
----------	-------	----------	---------	---

Word	Un-agglutinated	Stem
يراقبونهم	ي + راقب + ون + هم	راقب
They observe them	Them $+ (e) + observ + (they)$	Observe

Table 3.	Light	stemming	example	2
----------	-------	----------	---------	---

Word	Un-agglutinated	Stem
المواصلات	ال + مواصل + ات	مواصل
The communication	s + communicator + the	Communicator

In Table 2, after dropping affixes and extensions from the verb بيراقبونهم (they watch them), we get the stem راقب (watch), and in Table 3, the word المواصلات (the communicators) is decreased to مواصل (communicator).

2.2 Stemming

Stemming is the final procedure that comprises on reducing the word to its root as shown in Table 4.

 Table 4.
 Stemming example

Word	Stem	Root
يراقبونهم	راقب	رقب
They observe them	observe	See

In Table 4, after light-stemming the word براقبونهم (they observe them) to راقب (observe), the stemming gives us رقب (see) which is the root of the verb رقب.

2.3 Light-Stemming/Stemming Comparison

The procedure of stemming denies a word from its augmentations and form. By examining 'derived nouns' "الأسماء المشتقة" [14, 15], we can say that the significance of a word is learned from its pattern "زازن". In this way, getting the word out to its root may rise error rate when the work is on the semantics. We give as examples:

We give as examples: (Well enough \rightarrow Good) \rightarrow Good) \rightarrow Cood) \rightarrow Cood) \rightarrow Cood) \rightarrow Cood) \rightarrow Cood) The light-stemming main goal is to drop only the affixes from the word as shown in the next example: \rightarrow Cood) The light-stemming main goal is to drop only the affixes from the word as \rightarrow Lattice \rightarrow Better) \rightarrow Lattice \rightarrow Hetter) \rightarrow Hetter) \rightarrow Lattice \rightarrow Lattice \rightarrow Hetter) \rightarrow Lattice \rightarrow

2.4 Stemming Models

Main Stemming Models. Stemming algorithms can be organized in two categories: those that extract the stem by comparing the word's pattern with a predefined lexicon and those that extract the stem directly.

Among the first category of stemming algorithms [14], the most famous is the *Khoja* stemmer [15]. Its standard is to eliminate the longest prefix and the longest suffix, and after that comparing the result with some predefined base roots to extract the correct root.

By processing the word "المواصلات", *Khoja* stemmer begins by removing the prefix "لل and the suffix "ار", the stem that results "مواصل" has the pattern "أر". For this pattern, it drops the first and the third letters to get the stem "وصل" which exists in *Khoja* dictionary.

Second category of stemming algorithms doesn't use any dictionary; they reduce the word depending on statistical approach. The most famous is the *AI-Shalabi* extraction algorithm that give the root of a word from its letters, each letter is attributed to, and characterized by, a weight, and its rank relies on its position in the word [17, 18]. Hence, the letters of affixes (سألتمونييها) are instantly recognized by their weight (see Table 5).

In Table 5, letters "ب", "ق" and "ب" have the smallest products in the word "يراقبونهم". They constitute the root

At the point when the root doesn't contain any letter from (سالتمونيها), this model gives great results as appeared in the case above. However, when the root incorporates these letters, it can create troubles like the word المواصلات (the communications) which is decreased to "روسل" despite "روسل".

Word		يراقبونهم							
Letters	م	٥	ن	و	Ļ	ق	١	ر	ي
Weight	2	1	2	3	0	0	5	0	3.5
Rank	8.5	6.5	5.5	4.5	3.5	4	7	8	9
Product	17	6.5	11	13.5	0	0	35	0	31.5
Root	ر قب								

Table 5. Example of al shalaby root extractor

Models Comparison. Sawalha and Atwell looked at the principle stemming algorithms from both categories on four measures of precision [19]. They credited to *Khoja* stemming model the most elevated precision on the extraction of three letters root, then the morphological analyzer of *Tim Buckwalter*, and the Tri-literal root algorithm and the Voting algorithm calculation (realizing that 80–85% of arabic words are derived from roots of three letters).

In this paper, we have utilized *Khoja* stemming algorithm and light-stemmer10 algorithm [20] techniques as feature extraction methods. These methods were compared in a text classification test.

Feature extraction has several advantages for text classification such as:

- 1. It lessens the running time of the classification operation. It wipes out the repetitive and needless features, so that the classifier uses a feature vector relevant and smaller than the original ones, this will diminish the running time of the classification operation.
- 2. It yields more precise outcomes. Feature extraction enhances the outcomes' precision by eliminating the unimportant and pointless features that doesn't help in the classification, and keeping the important ones that helps exploring the semantic of the text documents.
- 3. It limits the obliged memory to deal with the documents' vectors by decreasing and lowering the quantity of the features described by the vectors.

2.5 Arabic Text Classification

A large number of studies have been done to find a solution of the text classification problem. Most of its concerns the English and French texts [21], yet couple of ones have been applied to Arabic content.

Mesleh has used the Support Vector Machines (SVM) algorithm with the chi-square as a feature selection technique [22]. He assessed the execution of his classifier by a corpus gathered from online Arabic newspaper archives, including Aljazeera, Al Nahar, Al Hayat, Al Ahram and AlDostor and a couple of other websites. The gathered corpus contains 1445 records with various length. These records are classified into nine classes. *Mesleh* has reasoned that the SVM algorithm combined the chi-square technique surpass Naïve Bayesian and the KNN classifier in terms of F-measure.

Al-Harbi et al. have measure the performance of two famous classification algorithms: C5.0 decision tree algorithm and SVM algorithm, in classification of Arabic text. They used seven Arabic corpora: Saudi News Agency, Saudi News Paper, website, journalists, Discussions, Islamic subject and Arabic Poems [23]. The results demonstrated that the C5.0 algorithm has outrun SVM classifier as far as precision by around 10%, the SVM the average precision for SVM is 68.65%, whereas the average precision for the C5.0 is 78.42%.

El-Kourdi et al. used the Naïve Bayes algorithm for Arabic text classification [24]. They employed a corpus of 1,500 records gathered from Al Jazeera website categorized in 5 classes: Sport, Business, Culture and Art, Science and Health, every class contains 300 records. All words in the records were transformed to their roots. The outcomes demonstrated that the average precision was around 68.78% in cross validation and 62% in evaluation set experiments. The best precision by class was 92.8%.

Houssien et al. used three classification algorithms on Arabic text documents: Naïve Bayesian (NB), Sequential Minimal Optimization (SMO) and J48(C4.5) employing Weka [24]. The gathered corpus contained 2,363 records that fall into six classes: Sport, Economic, medication, politic, religion and science. The authors removed the stop words and decrease the number of features extracted from the records by using normalization approach. The precision, recall and error rate were deployed for comparing the accuracy of those classifiers. The outcomes demonstrate that SMO classifier accomplishes the most astounding accuracy and the least error rate, trailed by J48 (C4.5), and the NB classifier; while J48 classifier took the longest time, trailed by NB classifier then SMO classifier.

3 KNN with Multiple Feature Selection Techniques

In this work, text classification was tested using the KNN classifier on the CNN-Arabic corpus. Considering that the aim of our work is to study the impact of using the morphological feature extraction methods on Arabic text classification, we tested this classification operation in two cases: the first case is using the stemmer approach and the second one uses the light stemming approach.

Figure 2 presents the principle modules in our system; the next sections depict each of these modules.



Fig. 2. System architecture

Preprocessing: its goal is to present a modified copy of the documents that the classifier can understand and manipulate with ease. Basic functions for preprocessors include: converting the document, removing stop words, removing foreign characters, removing punctuation, removing articles and numbers.
Root Stemming: The *Khoja* algorithm was followed here as a feature extraction method. The *Khoja* algorithm finds the three-letter roots for Arabic words by following this procedure:

- 1. Remove diacritics
- 2. Remove stop words, punctuation and numbers
- 3. Remove definite article (الر, وال, فال, كال, بال)
- 4. Remove conjunction ()
- 5. Remove suffixes
- 6. Remove prefixes
- 7. Compare result against a list of patterns. If a match is found extract the root.
- 8. Check the match with the predefined root based.
- و by ي, و, / Replace: و by
- / by کرو : 10. Replace
- 11. If the root contains only two letters, check if they should contain a double character

Figure 3 describes an example of the *Khoja* algorithm. Note that several words such as (*اللاعب الملعب اللحية*) which mean "The player", "playground" and "the game" respectively are reduced to one stem (*لعب*).



Fig. 3. An example of preprocessing with root-stemming algorithm (Khoja algorithm).

Light Stemming: it's main goal is to improve the efficiency of classification while holding the words' semantics. It maintains the word's signification untouchable. In this stage, we applied the light-stemmer10 algorithm as a feature extraction technique. The principal of this algorithm lies on many runs that try to find and take off the most recurrent prefixes and suffixes from the word. Figure 4 represents an example of



Fig. 4. An example of preprocessing with light-stemming algorithm.

preprocessing with light-stemming algorithm. Here we mention that light stemming keeps up the distinction between (*اللاعبون اللعبة*) which means "players" and "the game"; their light stems are (*اللاعب اللعبة*) which means the player and the game.

Feature Vector: In this step, the bag of words (BOW) method was used:

- 1. We collect all the words contained in all the documents, then we eliminate the doubling, we call the result "global document".
- 2. For each document, its stems (roots) compared with the global document and their frequencies are recorded. The vector containing the stems with their frequencies is the characteristic vector.

Classification: we applied the KNN classifier. It is one of the simplest classification algorithm. Even with such simplicity, it can give highly interesting results. Moreover, KNN presents the following advantages:

- 1. Easy to interpret output
- 2. Low Calculation time
- 3. Hight Predictive Power

4 Experimentation and Result Analysis

4.1 Dataset Description

The dataset is assembled from the CNN-Arabic site and contains 5,070 textual records belonging to six classes: Middle East News 1,462, World News 1,010, Economics 836, Sport 762, Entertainment 474 and SciTech 526. It contains 2,241,348 words. This dataset was split into two sections: training and testing. The training dataset represents 70% of each class, while the testing dataset represents the remaining 30%. Records in the training dataset were stored as vectors, where every vector comprises of its unique words and their frequencies. Those Vectors were set up in two adaptations: Stem-level vectors and Light stem-level vectors.

Table 6 presents the characteristics of the two adaptations of the document vectors in addition to the original version of dataset. The significance of this Table is to demonstrate that the feature extraction techniques decrease the size of the dataset in addition to minimizing the required memory to deal with the dataset. As it's described in the Table 6, the stem-level form presented the most reduced space (11 MB). This is normal, as stemming decreases many words to one stem. On the other hand, the light stem-level vectors devoured (18 MB). This is kind of higher than the stem-level vectors.

Table 6.	The	properties	of	the	dataset
----------	-----	------------	----	-----	---------

Dataset version	Size in MB
Original version	25 MB
Stemmed version	11 MB
Light-stemmed version	18 MB

4.2 Experiments

Light-Stem Level. Applying text classification after using the light-stemming technique as a feature extraction technique. First of all, the preprocessing step was applied for every document, and the light-stems of the words was extracted. From that point forward, it is represented as a vector of light-stems with their relating weights. At last, these vectors are passed to the KNN classifier.

Stem Level. Applying text classification after using the stemming technique as a feature extraction technique. For this situation, the preprocessing step was applied for every document, and the stems of the words was extracted. Lastly, stem-level vector was created by storing the stemmed words with their relating weights and passed to the KNN classifier.

The experiments have been done on an Intel Core i7 CPU 2.20 GHz, with a RAM of size 8 GB. Table 7 presents the time passed by pre-handling and classification to all testing datasets for both stem and light-stem techniques.

Experiment	Preprocessing time	Classification time	Total time
Stem level	418	3,258	3,676

 Table 7. The elapsed classification time (seconds)

Table 7 demonstrates that the most minimal preprocessing time was accomplished in relation to stemming. And the raison of that is the smaller size of the vectors compared with light-stemming. Additionally, the used stemming algorithm needs to scan every given word just once to derive its stem.

Classification time mentions the time needed to classify the testing dataset by the KNN classifier. This time, stemming overrule light stemming. Note that the classification time is specifically corresponding to vector sizes. To whole up, stemming needs less preprocessing and classification times by comparison with light stemming.

4.3 Result Analysis

Figure 5 shows the recall of the Arabic text classification using the two stemming approaches on the testing dataset.



Fig. 5. Recall measure of classified testing dataset with different number of features

The results were obtained from the testing dataset. Those results demonstrate that the lightstemmer10 algorithm surpass the *Khoja* algorithm in almost all categories.

Besides comparing the two algorithms, shifting the number of the chosen features helped us analyzing the behavior of the two algorithms. The observational outcomes demonstrated that the recall measure decreases when the number of features is high or low, which can be translated by the way that some chosen features are unimportant for the primary case or they aren't sufficiently representative for the second case.

5 Conclusions and Future Work

In this work, we have tested feature extraction techniques for Arabic text classification. The dataset was grouped physically into six classes to be specific Middle East News, World News, Economics, Sport, Entertainment, Science and Technology and Arts And culture. The dataset was split into two sections: training dataset and testing dataset. The testing dataset present 30% of every class, while the training dataset represent the remaining 70%. The chosen approaches for feature extraction are the stemming [15] and light stemming [16]. The Stemming approach finds the three-letter roots for Arabic words, while light stemming drops the prefixes and suffixes from the Arabic words and maintains the resulting word. The K-NN was utilized for the classification task. The practices were as follows, the KNN classifier was run twice on the dataset: once with stemming (called stem-level), and once with light-stemming (called light stem-level).

The experiments showed that the best values of recall were achieved when light-stemming is used as a feature extraction method. Light stemming outperformed the stemming approach, and the main reason of that is that the stemming affects the words meanings.

However, the both algorithms suffer from several problems, such as; the results of light-stemmer10 algorithm does not always represent significant words, and that's because the algorithm is not based on any dictionary or linguistic rules. And the *Khoja* algorithm have difficulties when it concerns roots of four or five letters and irregular plurals. For those raisons, our future work will concentrate on improving those algorithms.

References

- Al-Arabizi: Why do Google fight it more than Arabic? http://www.bbc.com/arabic/mobile/ scienceandtech/2012/12/121220_arabic_language_internet_arab_days.shtml. Accessed 20 Mar 2017
- 2. Sorting the languages of the world in terms of proliferation. https://arabic.rt.com/news/786982- ترتيب لمعات العالم-الانتشار. Accessed 20 Mar 2017
- List of languages by total number of speakers. https://ar.wikipedia.org/wiki/قائمة_اللغات_حسب العدد_الكلي للمتحدثين/Accessed 20 Mar 2017
- 4. Correa, R.F., Ludermir, T.B.: Automatic text categorization: case study. In: the 7th Brazilian Symposium on Neural Networks, Pernambuco, Brazil, November 2002
- Eyheramendy, S., Lewis, D., Madiagn, D.: On the naive Bayes model for text categorization. In: The Artificial Intelligence and Statistics Conference, Key West, Florida, January 2003

- Peter, B.: Active learning of SVM and Decision tree classifiers for text categorization. In: The 4th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence, Herlany, Slovakia, January 2006
- Wang, P., et al.: Semantic Clustering and convolutional neural network for short text categorization. In: 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, pp. 352–357, 26–31 July 2015
- Gongde, G., et al.: An kNN model-based approach and its application in text categorization. In: The International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), Seoul, Korea, February 2004
- Basu, A., Walters, C., Shepherd, M.: Support vector machines for text categorization. In: 36th Annual Hawaii International Conference, Los Alamitos, California, USA, January 2003
- Jun, Y., et al.: OCFS: optimal orthogonal centroid feature selection for text categorization. In: 28th Annual International ACM SIGIR Conference, Salvador, Brazil, August 2005
- Seo, Y., Ankolekar, A., Sycara, K.: Feature selection for extracting semantically rich words. Technical report CMU-RI-TR-04-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, March 2004
- Yang, Y., Pedersen, J.: A comparative study on feature selection in text categorization. In: 40th International Conference on Machine Learning, Nashville, Tennessee, USA, July 1997
- 13. Tuerlinckx, L.: La lemmatisation de l'arabe non classique. In: 7th International Days of Textual Data Statistical Analysis (JADT), Louvainla- Neuve, Belgium (2004)
- Anjali, G.J., et al.: A comparative study of stemming algorithms. Int. J. Comput. Technol. Appl. (IJCTA) 2(6), 1930–1938 (2011)
- 15. Khoja, S., Garside, R.: Stemming arabic text. Computer Science Department, Lancaster University, Lancaster, UK (1999)
- Al-Shalabi, R., Kanaan, G., Al-Serhan, H.: New approach for extracting Arabic roots. In: The International Arab Conference on Information Technology, Alexandria, Egypt, pp. 42– 59 (2003)
- 17. Evens, M.: A computational morphology system for Arabic. In: The Workshop on Computational Approaches to Semitic Languages, Montreal, Quebec, Canada, August 1998
- Sawalha, M., Atwell, E.: Comparative evaluation of arabic language morphological analysers and stemmers. COLING: companion volume – Posters and Demonstrations, Manchester, UK, pp. 107–110 (2008)
- Connell, M.: Improving stemming for arabic information retrieval: light stemming and cooccurrence analysis. In: 25th Annual International Conference on Research and Development in Information Retrieval, Tampere, Finland, pp. 275–282, August 2002
- Joachims, T.: A statistical learning model of text classification for support vector machines. In: 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, Louisiana, USA (2001)
- 21. Moh'd, A., Mesleh, A.: Chi square feature extraction based svms arabic language text categorization system. J. Comput. Sci. **3**, 6 (2007)
- Al-Harbi, S., Almuhareb, A., Al-Thubaity, A., Khorsheed, M.S., Al-Rajeh, A.: Automatic Arabic text classification. In: 9th International Conference on the Statistical Analysis of Textual Data, Lyon, France, 12 –14 March 2008
- Kourdi, M.E., Bensaid, A., Rachidi, T.: Automatic Arabic document categorization based on the Naive Bayes algorithm. In: The Workshop on Computational Approaches to Arabic Script-based Languages, Geneva, Switzerland (2004)
- Majed, F.O., Hussien, I.: AL-dwan, M., Shamsan, A.: Arabic text classification using smo, Naïve Bayesian, J48 Algorithms. Int. J. Res. Rev. Appl. Sci. 9, 10 (2011)

A Comparative Study of the Four Well-Known Classification Algorithms in Data Mining

Safae Sossi Alaoui^(IZI), Yousef Farhaoui, and Brahim Aksasse

ASIA Team, M2I Laboratory, Department of Computer Science, Faculty of Sciences and Technics, Moulay Ismail University, BP 509 Boutalamine, 52000 Errachidia, Morocco sossialaouisafae@gmail.com, youseffarhaoui@gmail.com, baksasse@yahoo.com

Abstract. Data mining is about extracting useful knowledge from data. It has various techniques and algorithms. Yet, the most widely used are classification algorithms which deal with the problem of affecting new data element to one of predefined classes. There are a wide range of classification algorithms such as decision trees, neural networks, K-NN, Bayes, support vector machines (SVM); and so on. This study focuses on four algorithms; Naive Bayes, Multi-Layer Perceptron (MLP), SVM and C4.5; all of them are based on mathematical calculations but in different ways. In this paper, we aim to make a comparison between the four algorithms in terms of well-chosen criteria like classification accuracy and execution time. Moreover, we implement these algorithms with the same dataset; relative to diabetes on women; in order to present the different results by using Waikato environment for knowledge analysis (Weka).

Keywords: Data mining \cdot Classification algorithms \cdot Decision tree \cdot Naive Bayes \cdot SVM \cdot MLP \cdot C4.5

1 Introduction

Diabetes is a set of diseases in which human body has high levels of blood sugar. This issue can risk for many health problems such as skin infections, eye problems, nerve damage... etc. Diabetes is too danger for women because it can affect both mothers and their babies during pregnancy [1]. Therefore, the need for predicting this disease from data mining techniques becomes extremely crucial.

Data mining refers to the process of discovering interesting knowledge from a large amount of data. It has its great application in diverse areas such as telecommunication industry, marketing, business, biological data analysis and other scientific applications.

When dealing with data mining, there exist several types like classification, clustering, association rules learning and regression. Each type gives a different impact or result depending on the problem that we try to solve. For our case, we focus on classification algorithms which arrange the data into predefined classes. It provides many algorithms and techniques, but the common widely used are Bayes, neural network; support vector machines (SVM) and decision tree.

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_32 The rest of the paper is organized as follows; in the second section we present the different works related to data mining especially the latest comparative studies of classification algorithms. In the third section, we describe the methodology followed and the tool used. In the fourth section, we present the different results obtained. In the last section, we conclude our paper.

2 Related Works

Today, data mining community is interested to make comparative studies for several classification algorithms.

For example, Dogan and Tanrikulu [2] have made a comparison of Fourteen classification algorithms representing the different types of classification models (the AIRS2P, C4.5, CART, CSCA, Ex-CHAID, IBk, Logistics, LogitBoost, MLP, MLVQ, NaiveBayesian, QUEST, RSES, and SVM) for various datasets; exactly ten; and their implementation techniques. WEKA, SPSS, and Rosetta, were the main components used to run the selected algorithms. In addition, this study was based on multiple criteria such us algorithm accuracies, speed (CPU time consumed) and robustness.

Moreover, researchers like Lim et al. [3] have made a comparison of prediction accuracy, complexity, and training time of thirty-three old and new Classification Algorithms namely twenty-two decision tree, nine statistical, and two neural network algorithms by using thirty-two datasets. In this work, experiments generally showed that the mean error rates of several algorithms are sufficiently similar that their differences are statistically insignificant. Unlike error rates, there are huge differences between the training times of these algorithms.

Also, Rashid et al. [4] searched on decision support system for Diabetes Mellitus through Machine Learning Techniques. SMO, C4.5 and ANN are three classification models designed for classification, prediction, and description purposes to offer complete knowledge about Diabetes Mellitus patients.

In addition, Gupta et al. [5] focused on the following classification algorithms; BayesNet, NaiveBayes, J48, JRip, OneR and Decision Table; in order to analyze crime and accident datasets from Denver City, USA during 2011 to 2015. The criteria used in this study, are correct classification, incorrect classification, True Positive Rate (TP), False Positive Rate (FP), Precision (P), Recall (R) and F-measure (F). They have also used two different test methods: k-fold cross-validation and percentage split.

3 Methodology

For achieving the goal of this research, we have based on Fig. 1 which demonstrates the methodology followed in this study. Firstly, it consists on one dataset related to women diabetes.

Secondly, it specifies the software used among the best four open source tools available in the market. Thirdly, it compares a set of selecting classification algorithms in data mining. Finally, it shows the different results obtained.



Fig. 1. Methodology of study

3.1 Dataset Description

In this paper, we have used one dataset related to women diabetes; at least 21 years old of Pima Indian heritage; taken from the UC Irvine Machine Learning Repository which is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms [6] (Tables 1 and 2).

Name of dataset	Pima Indians diabetes database	
Original owners	National Institute of Diabetes and Digestive and Kidney Diseases	
Date received	9 May 1990	
Number of instances	768	
Number of attributes	9	
Type of attributes	Numeric	
Number of classes	2	

Table 1. Description of the dataset

Table 2.	Description	of	attributes
----------	-------------	----	------------

Name of attribute	Details of attribute
preg	Number of times pregnant
plas	Plasma glucose concentration a 2 h in an oral glucose tolerance test
pres	Diastolic blood pressure (mm Hg)
skin	Triceps skin fold thickness (mm)
insu	2-Hour serum insulin (mu U/ml)
mass	Body mass index (weight in kg/(height in m)^2)
pedi	Diabetes pedigree function
age	Age (years)
class	Class variable (tested positive, tested negative)

3.2 Tool Used

Plenty of open source tools are available for data mining in the market, but the most powerful tools are Weka, KNIME, Orange, and RapidMiner. Bellow descriptions of these data mining toolkits:

Weka [7] (Waikato Environment for Knowledge Analysis) is a widely used suite of machine learning software written in Java and developed at the University of Waikato in New Zealand. It is very sophisticated and it supports many tasks including data preprocessing, clustering, classification, regression, visualization, and feature selection. Weka is free under the GNU General Public License and it provides access to SQL databases. It can be accessible from both the command line and the user interface named the Explorer.

KNIME [8] (Konstanz Information Miner) is a user friendly open source data analytics, reporting and integration platform written in java and based on Eclipse. It gives users the ability to add plugins which provide additional functionalities.

Orange [9] is a Python-based toolkit for data mining. It is very powerful and can be used for both novices and experts. It has a complete set of components for data preprocessing, feature scoring and filtering, modeling, model evaluation, and exploration techniques.

RapidMiner [10] is an environment for machine learning and data mining experiments written in the java programming language. It provides advanced analytics through template-based frameworks. It can be utilized for both research and real-world data mining tasks such as data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment.

According to a recent paper [11], which has conducted a comparison study between four data mining toolkits namely; Weka, Orange, Tanagra, and KNIME over some classification methods. They proved that Weka was the best tool in terms of the ability to run the selected classifier. Also, it has achieved the highest performance improvements when moving from the Percentage Split test mode to the Cross Validation test. Therefore, thanks to its various advantages we had chosen the open source software Weka to evaluate our classification algorithms.

3.3 Classification Algorithms

Overall, Classification in data mining is an outstanding task that can be used to portion the data into classes that are already predefined. Numerous classification algorithms are available. Yet, in this research, we had selected four well known classification algorithms; Naive Bayes, Multi-Layer Perceptron (MLP), SVM and C4.5, which are outlined in Table 3.

Algorithm name	Acronym	Years	Introduced by	Category
Naive Bayesian	Naive	1995	John and	Bayesian classification
classification	Bayes		Langley [12]	
C4.5 Decision tree	C4.5	1993	Quinlan [13]	Decision tree
revision 8				classification
Support vector	SVM	1992	Boser et al. [14]	Support vector
machines				classification
Multi-layered	MLP	1986	Rumelhart et al.	MLP Neural network
perceptron			[15]	classification

Table 3. Description of selected algorithms

3.3.1 Naive Bayes

Naive Bayes is statistical method for classification based on Bayes' Theorem:

$$\mathbf{P}(\mathbf{H}/\mathbf{X}) = \mathbf{P}(\mathbf{X}/\mathbf{H})\mathbf{P}(\mathbf{H})/\mathbf{P}(\mathbf{X}). \tag{1}$$

Where X is data attribute and H is some hypothesis.

It can solve problems involving both categorical and continuous valued attributes.

Moreover, the name Naive comes from the fact that this classifier considers that the effect of the value of a predictor on a given class is independent of the values of other predictors.

Naive assumption of «class conditional independence» :

$$P(X/Ci) = P(x1,x2,...,xn/Ci) = P(x1/Ci) * P(x2/Ci) * ... * P(xn/Ci).$$
(2)

Where X: (x1,x2,x3,...,xn) is the different data attributes and C:C1,C2,C3,...,Cm is a set of classes.

A naive Bayes classifier predicts class membership probabilities such as the probability that a given feature belongs to a particular class which maximizes this probability P(X|Ci)*P(Ci).

3.3.2 Decision Tree C4.5

C4.5 is a classification algorithm witch improves the ID3 algorithm by managing both continuous and discrete properties, missing values and pruning trees after construction. It is available in Weka tool as J48 and it is used to generate a decision tree based on Shannon entropy in order to pick features with the greatest information gain as nodes.

Entropy at a given node t:

$$Entropy(t) = -\Sigma p(j \mid t) \log 2 p(j \mid t).$$
(3)

 $p(j \mid t)$ is the relative frequency of class j at node t.

Information Gain:

$$GAINsplit = Entropy(t) - \Sigma(ni/n \ Entropy(i)). \tag{4}$$

Parent node t with n records is split into k partitions; ni is number of records in partition (node) i.

Gain Ratio:

$$GainRATIOsplit = GAINsplit/SplitINFO.$$
(5)

SplitINFO =
$$-\Sigma(ni/n\log ni/n)$$
. (6)

Parent node p is split into k partitions, ni is the number of records in partition i. The decision tree C4.5 is described on the following algorithm:

Algorithm C4.5 [16]:

```
Input: an attribute-valued dataset D
 Tree = \{\}
 if D is "pure" OR other stopping criteria met then
 terminate
 end if
 for all attribute a \in D do
 Compute information-theoretic criteria if we split on a
 end for
 abest = Best attribute according to above computed
criteria
 Tree = Create a decision node that tests abest in the
root.
Dv = Induced sub-datasets from D based on abest
 for all Dv do
 Treev = C4.5(Dv)
Attach Treev to the corresponding branch of Tree
 end for
 return Tree
```

3.3.3 Support Vector Machines (SVM)

Support vector machines is among the most robust and accurate methods in all well-known data mining algorithms. It is a supervised learning model its goal is to find a hyperplane that can separate two classes of given samples with a maximal margin. Intuitively, a *margin* refers to the amount of space, or separation, between the two classes as defined by a hyperplane. Geometrically, the margin corresponds to the



Fig. 2. SVM algorithm

shortest distance between the closest data points to any point on the hyperplane. Figure 2 illustrates the optimal hyperplane which can be defined as:

$$\mathbf{w}^T \mathbf{x} + \mathbf{b} = \mathbf{0}.\tag{7}$$

Where w and b denote the weight vector and bias respectively.

3.3.4 Neural Network (MLP)

Multi-Layer perceptron (MLP) is a feedforward neural network model which is an emulation of biological neural system. MLP consists of one or more layers between input and output layer. The term "feedforward" significates that data flows are in one direction; from input to output layer. MLP is trained with a supervised learning technique called backpropagation learning algorithm. MLP is an amelioration of the standard linear perceptron. MLPs are widely used for pattern classification, recognition,



Fig. 3. Multi-layer perceptron (neural network algorithm)

prediction and approximation. Multi-Layer Perceptron can solve problems which are not linearly separable.

An MLP can be represented graphically as follows: (Fig. 3)

3.4 Criteria of Comparison

It is very important to evaluate classifiers in order to know which methods are the best. Indeed, there are several metrics that can be utilized to measure the performance of a classifier or predictor; among these criteria: accuracy, speed, robustness, scalability, interpretability. In our work, we utilize the first two criteria in order to compare selected algorithms. Bellow a general definition of each metric:

- Accuracy: refers to the ability of the model to correctly predict the class label of new or previously unseen data.
- **Speed**: means the computation costs involved in generating and using the model. It depends on the algorithm complexity.
- **Robustness**: refers to the ability of the model to make correct predictions given noisy data or data with missing values.
- Scalability: refers to the ability to construct the model efficiently given large amount of data.
- **Interpretability**: refers to the level of understanding and insight that is provided by the model [17].

4 Results

Empirical studies have proved that Cross validation is the most elaborate method for dataset with less than 1000 instances. It consists on a number of folds n which must be specified. The dataset is randomly reordered and then split into k folds of equal size. In each iteration, one fold is used for testing and the other k-1 folds are used for training the classifier. For our case, we have chosen the default option: 10-fold cross-validation.

It is quite clear; that we can consider an algorithm is a perfect classifier if at least on the training data, all instances were classified correctly as well as all errors are zero. However, it is not the case in reality. Therefore we can admit that a best classifier is the algorithm with the maximum of Correctly Classified Instances or the minimum of Incorrectly Classified Instances (Fig. 4).

SVM classifier has the highest accuracy with 76.8%, followed by Naive Bayes having correct classification rate of 75.75%, then MLP with 74.75% and finally Decision tree C4.5 has determined least correct instances with 74.49% (Fig. 5).

Best classifier is the algorithm with the minimum of Execution time. Naive Bayes is the fast classifier that took 0.03 s then Decision tree C4.5 with 0.05 s. While SVM time to build the model was 0.08 s, and MLP was the slowest classifier with 1.9 s (Fig. 6).

A more detailed performance description via; precision, recall, true- and false positive rate are useful measures for comparing classifiers. All these values are based on the confusion matrix and can be computed from it (Table 4).



Fig. 4. Correctly and incorrectly classified instances for each algorithm



Fig. 5. Classifiers accuracy



Fig. 6. Execution time of selected algorithms

Algorithm	Correctly classified instances	Incorrectly classified instances
Naive Bayes	586	182
MLP	579	189
SMO	594	174
C4.5 (J48)	567	201

Table 4. Correctly and incorrectly classified instances for each algorithm



Fig. 7. Confusion Matrix

Table 5. Execution time of selected algorithms

Execution time (second)
0,03
1,9
0,08
0,05

The confusion matrix [18] named also contingency table in the case of two classes, the confusion matrix is 2×2 . The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified. Figure 7 shows a representation of confusion matrix, where: TP = true positive, TN = true negative, FP = false positive, FN = false negative (Table 5).

The True Positive (TP) rate [18] refers to the proportion of examples which were classified as class x, among all examples which truly have class x. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row. It is equivalent to Recall.

True positive rate =
$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}).$$
 (8)

The False Positive (FP) rate [18] is defined as the proportion of examples which were classified as class x, but belong to a different class, among all examples which are

Algorithm	TP rate	FP rate	Precision	Recall	F-Measure
Naive Bayes	0,763	0,307	0,759	0,763	0,76
MLP	0,754	0,314	0,75	0,754	0,751
SMO	0,773	0,334	0,769	0,773	0,763
C4.5 (J48)	0,738	0,327	0,735	0,738	0,736

Table 6. Detailed performance description

not of class x. In the matrix, this is the column sum of class x minus the diagonal element, divided by the rows sums of all other classes.

True negative rate =
$$TN/(TP + FN)$$
. (9)

The Precision [18] means the proportion of the examples which truly have class x among all those which were classified as class x. In the matrix, this is the diagonal element divided by the sum over the relevant column.

$$Precision = TP/(TP + FP).$$
(10)

The F-Measure [18] is simply a combined measure for precision and recall:

$$2 * Precision * Recall/(Precision + Recall).$$
 (11)

Bellow Table 6 shows the TP and FP rate of each classifier, the weighted average of Precision, Recall and F-Measure, obtained by using the 10-fold cross-validation approach.

SVM has the highest TP Rate and Recall with the value 0.773, it has also the greatest Precision, followed by Naive Bayes having TP rate and recall value of 0.763. Then MLP with 0.754 and finally decision tree C4.5 with 0.738.

5 Conclusion

This paper has concentrated on using Data mining techniques in order to make prediction for health problem, especially diabetes in women; a disease in which the body cannot control the level of sugar in the blood. Therefore, we have chosen the most four well-known classifiers; Naive Bayes, Multi-Layer Perceptron (MLP), SVM and C4.5; in order to pick the best algorithm for our dataset. Also, we have made a comparison between selected classification algorithms using the tool Weka. Indeed, according to the results obtained, our analysis indicates that support vector machines classifier has the highest accuracy followed by Naive Bayes, then multi-layer perceptron MLP and finally Decision tree C4.5. Moreover, in terms of execution time; Naive Bayes is the fast classifier then Decision tree C4.5 followed by SVM and MLP was the slowest classifier.

For the future work, we try to use a dataset with a huge number of instances. Certainly, it can lead us to manage Big data mining technologies with the aim of achieving the precise and perfect predictions.

References

- 1. How Diabetes Affects Women: Symptoms, Risks, and More. http://www.healthline.com/ health/diabetes/symptoms-in-women
- Dogan, N., Tanrikulu, Z.: A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness. Inf. Technol. Manag. 14, 105–124 (2013)
- Lim, T.-S., Loh, W.-Y., Shih, Y.-S.: A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Mach. Learn. 40, 203–228 (2000)
- 4. Rashid, T.A., Abdulla, S.M., Abdulla, R.M.: Decision support system for diabetes mellitus through machine learning techniques. Database **7** (2016)
- Gupta, A., Mohammad, A., Syed, A., Halgamuge, M.N.: A comparative study of classification algorithms using data mining: crime and accidents in Denver City the USA. Education 7 (2016)
- 6. UCI Machine Learning Repository: Pima Indians Diabetes Data Set. https://archive.ics.uci. edu/ml/datasets/Pima+Indians+Diabetes
- Weka 3 Data Mining with Open Source Machine Learning Software in Java. http://www. cs.waikato.ac.nz/ml/weka/
- 8. KNIME | Open for Innovation. https://www.knime.org/
- 9. Orange Data Mining Fruitful & Fun. https://orange.biolab.si/
- 10. Data Science Platform | Machine Learning. https://rapidminer.com/
- Wahbeh, A.H., Al-Radaideh, Q.A., Al-Kabi, M.N., Al-Shawakfa, E.M.: A comparison study between data mining tools over some classification methods. IJACSA. Int. J. Adv. Comput. Sci. Appl. 8, 18–26 (2011). Spec. Issue Artif. Intell.
- 12. John, G., Langley, P.: Estimating continuous distributions in Bayesian classifiers (1995)
- 13. Salzberg, S.L.: C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. Mach. Learn. 16, 235–240 (1994)
- Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 144– 152. ACM (1992)
- 15. Rumelhart, G., Hinton, G., Williams, R.: Learning internal representations by error propagation. Presented at the (1986)
- Xindong, W., Vipin, K.: The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, New York (2009)
- 17. Stefanowski, J.: Data Mining Evaluation of Classifiers (2008). http://www.cs.put.poznan. pl/jstefanowski/sed/DM-4-evaluatingclassifiersnew.pdf
- Kirkby, R., Frank, E., Reutemann, P.: Weka Explorer User Guide for Version 3-5-8. Univ. Waikato, Hamilton (2007)

Advanced SQL-to-SPARQL Query Transformation Approach

Nassima Soussi^(⊠) and Mohamed Bahaj

Department of Mathematics and Computer Science, Faculty of Science and Technologies, Hassan 1st University, Settat, Morocco nassima.soussi@gmail.com,mohamedbahaj@gmail.com

Abstract. No one can deny the emergence of semantic web technologies with their considerable performance in data management offering a better cooperation between people and computers, but unfortunately, relational databases are still the most used; therefore establishing a connection between them becomes an active topic aiming to bridge the gap between the both heterogeneous systems. Regarding the interoperability between their query languages, more precisely, SQL-to-SPARQL query transformation direction, some solutions have been explored to ensure this conversion, but all these approaches have the same gap because they start the mapping process before analyzing and optimizing the input SQL query; this weakness has motivated us to add a pretreatment phase aiming to optimize some SQL components with a specific focus on Left and Right Outer Join command(s) generating the most complex ity of the output SPARQL query.

Keywords: SQL-to-SPARQL \cdot Query transformation \cdot SQL optimization \cdot Outer join optimizer

1 Introduction

The semantic web [10] is a W3C recommendation; it allows data to be shared and reused across multiple applications, businesses, and user groups aiming to manage intelligently the large amount of data in the web and offering a better cooperation between computers and people. In order to realize these goals, the semantic web uses several technologies: RDF (Resource Description Framework) [4] as the standard model for representing data on the web, OWL (Web Ontology Language) [2] for creating structured ontology and SPARQL language [6] for querying RDF data. On the other hand, the majority of web information is stored in relational databases, which makes the development of methods and tools to contribute in the interoperability between the both systems a relevant need.

In this light, some approaches have been developed regarding the query transformation from SQL to SPARQL in order to facilitate for relational users the data extraction by querying RDF stores with SQL language, but unfortunately, all these approaches have the same weakness in their proposed systems since they convert directly the input SQL query to its equivalent SPARQL one without any pre-processing

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_33 phase enabling the optimization of this raw SQL query filled by users before starting the conversion process. This problem has motivated us to write this paper so as to remedy this gap and establish an intermediate component aiming to optimize some SQL statements in order to minimize the complexity of the output query and contribute in the improvement of existing SQL-to-SPARQL conversion systems with a specific focus on the one-sided outer joins (Left and Right outer join) optimization. In fact, the main source of the high complexity in SPARQL queries is due to Optional operator that makes alone the query evaluation PSPACE-hard [3]. On the other hand, in the SQL-to-SPARQL conversion systems, this special operator is obtained from an outer join operation; hence, we deduce that it is very helpful to rewrite outer joins as inner joins, whenever possible, so as to avoid having SPARQL equivalent queries with optional operators.

The remainder of this paper is structured as follows: Sect. 2 presents a brief description of some related works to the current topic, and then Sect. 3 introduces our approach by giving some theoretical background. Section 4 presents the architecture and the proposed algorithm that aims to rewrite left and right outer joins before starting the SQL-to-SPARQL mapping conversion. Finally, Sect. 5 concludes this work and suggests some future extensions of this topic.

2 Related Works

Recently, the semantic web has become a buzzword in the web world due to its great considerable performance in data management. In fact, several researches have been developed aiming to establish an efficient connection between the semantic web and relational world (considered as the most used database management system), more precisely, to ensure the query interoperability between SQL and SPARQL languages. Regarding the translation direction SQL-to-SPARQL [1, 5, 7, 8], there are not so much works related to this later, and the few existing approaches have several weaknesses.

The RETRO method [5] provide interoperability between RDF Stores and RDBMS by firstly deriving a relational schema from the RDF store, and secondly providing a means to translate an SQL query to a semantically equivalent SPARQL query. In addition, SQL2SPARQL solution [1] aims also to transform a classic SQL query into an SPARQL query using some dynamic mappings and transformation rules based on the combination of ideas already presented in other works. Furthermore, the same mapping direction is treated via R2D method presented in [8]; it proposes a mechanism which enables reusability of relational tools on RDF data including SQL-to-SPARQL translation; this work converts SQL queries, with pattern matching and aggregation, into the SPARQL equivalent queries. Similarly, the authors in [7] propose an algorithm for querying RDF data using SQL by translating SQL queries (simple and complex ones containing UNION, INTERSECT or EXCEPT expressions) into an equivalent SPARQL queries.

All the previous approaches convert directly the input SQL query to its equivalent SPARQL one without any pre-processing phase enabling the optimization of one-sided

outer joins. To the best of our knowledge, this paper is the first work developed in this topic treating the optimization of the input SQL queries before starting the transformation process.

3 Overview

This section introduces our work by giving some theoretical background of the current paper. Firstly, we detail the different operations used in this optimization approach (grammar and examples), secondly we describe the problematic and the motivation of the current work, and finally we quote some optimization and rewriting rules used to attain our aim.

3.1 SQL Outer Join

Outer join is a relational operation used via SQL language aiming to combine information from two or more tables by returning all rows from at least one of these tables mentioned in FROM clause. We distingue three types of outer joins: Left, Right and Full outer join. In our study, we will focus on Left and Right outer join. In fact, the Left outer join returns all rows from the left table and the matched rows from the right table; similarly, the Right outer join returns all rows from the right table and the matched rows from the left table (as illustrated in Figs. 1 and 2). The outer join operation can be used to combine information via two-sides from different companies work groups using independent databases; in addition, this operation is very helpful to preserve objects without children and make hierarchical views. The syntax of this operation is described via the follow grammar:

```
SQLquery = SelectClause FromClause WhereClause
FromClause = 'FROM' JoinClause
JoinClause = LeftRelation JoinOperator RightRelation OnCondition
JoinOperator = 'LEFT OUTER JOIN' | 'RIGHT OUTER JOIN'
LeftRelation = LeftRelationName | JoinedQuery
RightRelation = RightRelationName | JoinedQuery
```

SELECT * FROM left_relation LEFT OUTER JOIN right_relation ON left_relation.id = right_relation.id



Fig. 1. Left outer join description

SELECT * FROM left_relation RIGHT OUTER JOIN right_relation ON left_relation.id = right_relation.id



Fig. 2. Right outer join description

OnCondition = ON '(' Expressions ')' Expressions = Expression (LogicOperator Expression)?* Expression = Attribute Op Value | Attribute 'IS NOT NULL' LogicOperator = 'AND' | 'OR' Op = '<' | '>' | '=' | '>=' | '<=' | '<>'

At database level, we consider two tables, Researcher and Supervisor. We aim to make a list of all researchers and their supervisors by joining the both previous tables with a left outer join operator in order to return the researchers who have no supervisor. The valid SQL query is given below:

SELECT Researcher.last_name, Supervisor.spv_name FROM Researcher LEFT OUTER JOIN Supervisor ON Researcher.rsch_id = Supervisor.rsch_id

3.2 SPARQL Optional Operator

Optional graph pattern is a combination of a pair of graph patterns [6]; if the optional match fails, the whole query does not fail, and in this case, a NULL value will be returned for unbound variables. A syntaxical description of this operation is given via the grammar bellow:

```
SPARQLQuery = SelectClause 'WHERE{' GraphPattern OptionalClause '}'
GraphPattern = TriplePattern ('.' TriplePattern)?*
OptionalClause = 'OPTIONAL{' GraphPattern (OptionalClause)?* '}'
TriplePattern = Subject Predicate Object
```

An example of SPARQL query with an Optional pattern is given below:

SELECT ?last_name ?spv_name
WHERE {
?rsch_id :Reseacher ?last_name.
OPTIONAL { ?rsch_id :Supervisor ?spv_name }
}

3.3 Problematic and Motivation

The notions defined above allow us to deduce that the basic equivalent semantic of Left/Right outer join is Optional pattern represented via the semantic relation presented below:

$$[[P_1 \text{ OPT } P2]]_D = [[P_1]]_D = \bowtie \ [[P_2]]_D = [[P_2]]_D \bowtie = [[P_1]]_D.$$
(1)

Where OPT denotes an Optional operator, $=\bowtie$ a Left outer join operator, $=\bowtie$ a Right outer join operator, $[[.]]_D$ the evaluation of a graph pattern (detailed in [9]), D an RDF document, P₁ and P₂ graph patterns. Therefore, and remaining within our context of SQL-to-SPARQL query transformation, the Optional operator is generated from the one-sided outer joins; and since it is the responsible for the high complexity in SPARQL queries (PSPACE-hard), so a rewriting or an optimization of the outer joins is recommended before starting the mapping process.

3.4 Optimization Rules

In order to realize the previous goal, we have used the algebraic equivalence rules defined bellow so as to simplify the one sided outer joins. The Left outer join operator can be replaced by an inner join if and only if the condition is null-rejected (False or Unknown for the generated null row), else we use the second rule to guarantee this equivalence. On the other hand, the right outer join simplification is obtained by permuting the joined relation so as to have a left outer join expression and continue to use these two rules.

$$\mathbf{R}_1 = \overset{c}{\bowtie} \mathbf{R}_2 := \mathbf{R}_1 \overset{c}{\bowtie} \mathbf{R}_2. \tag{2}$$

$$\mathbf{R}_1 = \stackrel{c}{\bowtie} \mathbf{R}_2 := \left(\mathbf{R}_1 \stackrel{c}{\bowtie} \mathbf{R}_2\right) \cup (\mathbf{R}_1 \setminus \mathbf{R}_2). \tag{3}$$

4 Architecture and Algorithms

4.1 Architecture Design

Regarding SQL-to-SPARQL conversion system, all existing approaches transform directly the received SQL query to an equivalent SPARQL query (see Fig. 3). However, our proposed system bridges this gap by adding a pretreatment layer before starting the mapping process, more precisely, one-sided outer join optimization (see Fig. 4).

The proposed architecture of our optimization component, as schematized in Fig. 5, illustrates the different steps that guarantee this aim. The system is composed of four components: *Query Analyser & Corrector, One-sided Outer Join Optimizer, Right to Left Outer Join Converter* and *Is Null Rejected*.



Fig. 3. Architecture of existing SQL-to-SPARQL translation systems



Fig. 4. Mechanism of existing SQL-to-SPARQL translation systems



Fig. 5. Global architecture of one-sided outer join optimizer

4.2 Transformation Algorithm

This section details the different components composing our optimization layer designed bellow:

Query Analyser & Corrector. Intended to analyze and scan SQL query in order to check syntactic errors and correct them before starting the optimization process.

One-sided Outer Join Optimizer. This is the main component in our system; it takes as input an SQL query q_{in} containing one-sided outer join command(s) so as to return at the end an SQL optimized query. Firstly, we parse q_{in} to a binary tree in order to extract

each component of SQL query separately, and then we parse the FROM clause to a binary tree named T. Secondly, we glance through the tree T (containing the outer joins operators) from top to down and we check if the current operator is null-rejected using the sub component *IsNullRejected* described subsequently, then the Left or Right outer join operator is changed to an inner join and the ON condition is translated to the Where clause. Else, we verify via the sub components *IsNullRejected* and *IsRightOJ* if the operator is not null-rejected and it is a Right outer join one respectively, then we call the *Right2LeftOJConverter* aiming to convert the Right outer join expression to a Left outer join one before starting the optimization process using the third rule in the previous section. At the end, we concatenate all the previous result so as to form the optimizer output SQL query.

Algorithm of the One-sided Outer Join optimizer component

```
Input: SQL query with one-sided outer join(s), qin
Output: Optimized SQL query, q<sub>out</sub>
Begin
Tree tree = parse(q_{in})
q_{in}^{SELECT} = tree.getSelectClause()
q_{in}^{FROM} = tree.getFromClause()
q_{in}^{WHERE} = tree.getWhereClause()
Tree T = parse(q_{in}^{FROM})
    For (i=1; i<= T.size(); i++)</pre>
           Op_i = T[i]
       If (Op<sub>i</sub>.IsNullRejected() = True) then
Replacing the one-sided outer join operator with the
inner join one (Rule 1)
Adding the ON condition to the where clause (q_{in}^{WHERE})
       Else
           If (IsRightOJ(Op_i) = True) then
                 Op_i = Right2Left0JConverter(Op_i)
           End if
Replace the one-sided outer join operator with the in-
ner join one union with minus (Rule 2)
Adding The ON condition to the where clause (q_{in}^{WHERE})
       End if
  End for
q_{out} = q_{in}^{SELECT} + q_{in}^{FROM} + q_{in}^{WHERE}
Return q<sub>out</sub>
End Algorithm
```

Is Null Rejected. This sub component takes as input an outer join's condition in order to check if it is null-rejected or not and return a Boolean value. We present bellow the proposed algorithm.

Algorithm of Is Null Rejected component

```
Input: Outer join's condition, C
Output: Boolean value
Begin
If (C is like a joined tables attribute with 'IS NOT
  NULL' condition
   OR
   C is evaluated to False or Unknown for the generated
  null tuples
  OR
   C is a conjunction containing a null-rejected condi-
tion
   OR
   C is a disjunction of null-rejected conditions) then
Return True
End if
Return False
End Algorithm
```

Right to Left Outer Join Converter. Takes as input a Right outer join expression Exp_{ROJ} so as to convert it to an equivalent Left outer join one by swapping the joined tables ($R_L Right Outer Join T_R := T_R Left Outer Join T_L$). In this algorithm, we have used the procedures *getLeftRelation*, *getRightRelation* and *getOJOperator* aiming to extract the different components of Exp_{ROJ} (left relation, right relation and the outer join operator respectively). The description of this sub component is described subsequently.

Algorithm of Right to Left Outer Join Converter component

```
Input: Right outer join expression, Exp_{ROJ}
Output: Equivalent Left outer join expression, Exp_{LOJ}
Begin
R_L = ExpROJ.getLeftRelation()
R_R = ExpROJ.getRightRelation()
Op = ExpROJ.getOJOperator()
Exp_{LOJ} += R_R + Op + R_L
Return Exp_{LOJ}
End Algorithm
```

5 Conclusion

In this paper, we have contributed in the reinforcement and the amelioration of interoperability's tools between SQL and SPARQL languages, more precisely, we have proposed an enhancement of SQL-to-SPARQL translation approaches since there are not so much works related to this later, and all existing ones convert directly SQL queries without a pretreatment and optimizer phase. We have developed an optimizer for one-sided outer join components in SQL query in order to avoid the generation of optional pattern(s) in the output SPARQL query.

One obvious extension of our approach would be the implementation of our algorithm in order to prove its performance on a real data and make our strategy easily and effortless exploitable by the target audience. Another promising direction for our future research works regarding the enhancement of SQL-to-SPARQL transformation is to add more optimization rules to our optimizer component.

References

- 1. Antal, M., Anechitei D., Cuza, A.I.: SQL2SPARQL (2012)
- Bechhofer, S.: OWL: Web ontology language. In: Encyclopedia of Database Systems, pp. 2008–2009. Springer US, (2009)
- 3. Schmidt, M.: Foundations of SPARQL Query Optimization (Doctoral dissertation, Institut für Informatik) (2009)
- 4. World Wide Web Consortium: RDF 1.1 Concepts and Abstract Syntax (2014)
- Rachapalli, J., Khadilkar, V., Kantarcioglu, M., Thuraisingham, B.: RETRO: a framework for semantics preserving SQL-to-SPARQL translation. In: EvoDyn Workshop (2011)
- Harris, S., Seaborne, A., Prud'hommeaux, E.: SPARQL 1.1 query language. W3C Recommendation, 21 (2013)
- Alaoui, L., Abatal, A., Alaoui, K., Bahaj, M., Cherti, I.: SQL to SPARQL mapping for RDF querying based on a new efficient schema conversion technique. Int. J. Eng. Res. Technol. IJERT 4(10), 57–61 (2015)
- Ramanujam, S., Gupta, A., Khan, L., Seida, S., Thuraisingham, B.: R2D: A bridge between the semantic web and relational visualization tools. In: IEEE International Conference on Semantic Computing, ICSC 2009, pp. 303–311. IEEE, September 2009
- Pérez, J., Arenas, M., Gutierrez, C.: Semantics and complexity of SPARQL. In: International Semantic Web Conference, pp. 30–43. Springer, Berlin, November 2006
- Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intell. Syst. 21(3), 96–101 (2006)

Migration from Relational Databases to HBase: A Feasibility Assessment

Zakaria Bousalem¹(^[M]), Ilias Cherti¹, and Gansen Zhao²

Faculty of Science and Technologies, Hassan 1st University, Settat, Morocco zakaria.bousalem@gmail.com, iliaschertilO@gmail.com ² School of Computer Science, South China Normal University, Guangzhou, China gzhao@m.scnu.edu.cn

Abstract. Relational Databases are currently at the heart of information system of the companies. In recent years, the relational model has become de facto standard thanks to its maturity and efficiency. However, the fact that the data of some companies or institutions have become too large, new systems has appeared namely NoSQL which belongs to the Big Data era. Big Data comes due to the emergence of new online services on which customers have become increasingly connected, which creates a large digital data unbearable by the traditional management technical tools, which raise new challenges for companies especially to access, store and analyse data. In this paper we will propose a feasibility study of migration from relational databases to NoSQL databases specifically HBase database, by applying the operations of the relational algebra in HBase data model and explore the implementation of these operations on HBase by using the native functions of this DBMS and also by using the MapReduce Framework.

Keywords: Migration · Relational database · Relational algebra · HBase · Column-oriented · NoSQL · Big data · MapReduce · Feasibility assessment

1 Introduction

Relational database management systems (RDBMS) are the most common solution in many applications for storing and retrieving data due to its maturity and reliability. Relational databases are based on the Codd model (relational) [1] which has privileged a system of relations based solely on the values of the data, and a manipulation of these data using a high level language called SQL [3], implementing a new mathematical theory similar to the set theory proposed by Codd called "relational algebra" [1]. Relational algebra defines operations that can be applied on relations. Relational operations allow to create a new relation (table) from elementary operations on other tables namely union, intersection, selection, and join.

However, these systems cannot support the explosion of digital data that modern Web applications have introduced. This explosion of digital data forces new ways of seeing and analyzing the world. These applications must support a large number of simultaneous users (tens of thousands or even millions), ensure the scalability generated by storage of large data capacities, be always available, manage semi-structured data and non-structured data and adapt quickly to changing needs with frequent updates and new features.

To address this problem, many solutions have turned to non-relational databases, commonly known as NoSQL databases, to enable massively parallel and geographically distributed database systems to support the internet applications such as facebook, ebay, twitter, Sears and Amazon [6].

The "NoSQL" databases are not usually a replacement, but rather a complementary addition to RDBMS and SQL. Consequently, developing a mapping tool between relational and NoSQL will be very much requested.

In our paper we will work on the HBase database [4, 8] thanks to its popularity. Indeed HBase is a Hadoop subproject, it is a distributed non-relational database management system, Written in Java, with structured storage for large tables. Zhao et al. [11] has carried out a comparison between MongoDB and relational algebra to investigating the feasibility of migrating relational databases to MongoDB, but there is no one for HBase. Migration requires feasibility assessment of the potential performance for new systems. For this purpose we will study the feasibility of applying the operations of the relational algebra in HBase data model and then explore the implementation of these operations on HBase by using the native functions of this DBMS and also byusing the MapReduce Framework [5]. The rest of the paper is structured as follows: In Sect. 2 we will introduce the basic definitions which we will begin with an introduction to the NoSQL databases, after we will present the HBase database, so we will see what MapReduce is. In Sect. 3 we will investigate the application of the relational algebra in HBase. Finally, Sect. 5 concludes our paper.

2 Basic Definitions

2.1 NoSQL Databases

A NoSQL database does not mean that no more queries are made, NoSQL simply means Not Only SQL. It is not a new query language for dialoging with the DBMS; it's a new approach for data storage. The term NoSQL refers to a category of massively parallel and geographically distributed databases management systems (DBMSs), most of them are designed to process large datasets within acceptable response time of user queries. They thus enrich the panel of traditional storage engines.

There are different categories of NoSQL DBMS [2]:

- **Key/Value:** The simplest NoSQL DBMS. It is in fact a huge hashmap with millions of entries. E.g. Redis, Riak, Voldemort (LinkedIn).
- **Document Oriented:** DBMS of key/value type with a document in value. The principle is to associate with a key a document regrouping different values. These documents are often represented by JSON or XML files. E.g. CouchDB, MongoDB
- **Column-oriented:** DBMS most resembling to the relational DBMSs, however Column-oriented DBMS allow missing values (unlike the relational model). These DBMSs are based on a notion of pair {key, value}. The column name can be seen

as the key. The column-oriented model, it's the model used in Hadoop. E.g. Cassandra, HBase, BigTable (Google)

• **Graph**: The goal is to represent the information in the form of nodes connected by edge (oriented or not). A node or edge can have attributes. This kind of DBMS stores effectively the relationships between data points, it's very useful for fraud detection, Real-Time recommendation engines and network and IT operations [19]. E.g. Neo4j, FlockDB

2.2 HBase

HBase is a distributed, column-oriented DBMS based on Hadoop. HBase provides random access and consistency for large amounts of unstructured and semi-structured data in a schema-less database organized by column families [9]. HBase uses HDFS as the file system for data storage and supports both queries and MapReduce. It was designed from the Google DBMS "BigTable" [10]. It's capable to storing a very large data (billions of rows/columns).

As show in Fig. 1, the HBase data model is based on six concepts [12], which are:

- Table: In HBase the data is organized in tables. Tables' names are strings.
- **Row:** In each table, the data is organized in rows. A row is identified by a unique key (RowKey).



Fig. 1. HBase model [20]

- **Column Family:** Data within a row is grouped by "Column Families ". Each row of the table has the same "Column Families", which can be populated or not. The "Column Family" is set when the table is created in HBase. The names of "Column Family" are strings.
- **Column Qualifier:** Access to data within a "Column Family" is done via the "column qualifier" or column. It is not specified at the creation of the table but earlier at the insertion of the data.
- **Cell:** Stores the values of this cell. The combination of the "RowKey", the "Column Family" and the "Column Qualifier" uniquely identifies a cell.
- Version: The values within a cell are versioned. The versions are identified by their timestamp.

2.3 MapReduce

MapReduce [5, 7] is a framework for parallel distributed computing over large amounts of data. Distributed computing is done via a cluster of machines. MapReduce fully manages the cluster and load balancing. This allows handling distributed computing without any knowledge of the underlying infrastructure. MapReduce is based on the notion of job. A job is split in a set of tasks. There are two types of tasks: **Map task** and **Reduce Task**.

3 Relational Algebra in Hbase

3.1 Union Set

Union it's a basic operation in relational algebra which requires two union-compatible operands.

Given two HBase tables T1 and T2, the definition of the union set is: $T1 \cup T2 = \{x | x \in T1 \text{ or } x \in T2\}$ where T1 and T2 are union-compatible.

3.2 Cartesian Product

A Cartesian product is a binary operation that combines two relations R1 and R2 and builds a third relationship exclusively containing all the possible combinations of occurrences of R1 and R2 relations, we note R1 \times R2.

The number of occurrences of the resulting relationship of the Cartesian product is the number of occurrences of R1 multiplied by the number of occurrences of R2. In HBase we can define the Cartesian product as follows:

$$T1 \times T2 = \{(r, s) : r \in T1 \text{ and } s \in T2\}$$

Where T1 and T2 are two tables in HBase, r and s are rows in these tables.

3.3 Intersection

The intersection is an operation that holds in two relations R1 and R2 with the same pattern and building a third relation that contains all rows of R1 also belong to R2, but no other rows. In HBase we can define the intersection as follows:

$$T1 \cap T2 = \{r : r \in T1 \text{ and } r \in T2\}$$

Where T1 and T2 are two tables in HBase and r is a row in these tables.

3.4 Selection

A selection is a unary operation that extracts certain row (or rows) from an HBase table where the selection condition P is satisfied.

- Notation: $\sigma_p(T)$
- Parameter: Table T and P is a propositional formula formed of a combination of comparisons and logical operators.
- Result: $\sigma_p(T) = \{r \in T: r \text{ satisfies the conditions given by } P\}$

3.5 Projection

The projection of an HBase table T1 is the HBase table T2 obtained by selecting the rows with columns in the C set and eliminating duplicate rows.

- Notation: π_{c_1,\ldots,c_n}*T*1
- Parameter: T1 is an HBase Table and C is a set of columns of selecting rows
- Result: T2 is an HBase Table with only the columns specified in C.

3.6 θ-Join and Equijoin

 θ -join (theta join) is a binary operation that consists of all combinations of rows in two HBase tables T1 and T2 that satisfy a condition.

• Notation:

$$T1_{r1} \bowtie_v T2 \text{ OR } T1_{r1} \bowtie_{r2} T2$$

- Parameter: T1 and T2 are two HBase Tables. r1 and r2 are two columns qualifier, θ is binary relational operator that can be :>, \geq , =, < or \leq , v is a constant value.
- Result: a subset of Cartesian product where the condition is satisfied.

We call this operation equijoin where the θ operator contains equality.

3.7 Natural Join

The natural join is a binary operation; it's the result of the Cartesian product of two HBase tables with the condition that must be at least one common attribute with the same name and the same value. If this condition is omitted, and the two HBase tables have no common attributes, the natural join becomes simply the Cartesian product. It's defined as follows:

$$T1 \bowtie T2 = \pi_{C \cup D} \sigma_{(T1.a1 = T2.a1)^{\wedge}(T1.a2 = T2.a2)^{\wedge} \dots \wedge (T1.an = T2.an)}(T1 \times T2)$$

Where C is the set of the column names of the HBase Table T1 and D is the set of the column names of the HBase Table T2.

3.8 Division

The division is a binary operation; it's a very powerful and useful operation, it's written as follows:

$$T1(k) \div T2(c)$$

Where T1and T2 are two HBase tables, k and c are the sets of column names of these HBase tables:

 $K = \{k_1, \ldots, k_m, c_1, \ldots, c_n\}, c = \{c_1, \ldots, c_n\}$ Where $c \subset k$ The result of this operation consists of all rows r(x) in T1 that appear in T1 in

combination with every tuple from T2, where x = k - c

The division it can be defined as follows:

$$T1(k) \div T2(c) = \{t | t \in \pi_{k-c}(T1) \text{ and } \forall u \in T2(t \times u \in T1)\}$$

4 Operations of Hbase

In this section we will model the HBase query capability by using the relational algebra.

4.1 Get

The "get" command is used to read data from an HBase table. This command returns a single line according to the row ID parameter. It's the equivalent of the SELECT command in SQL. Its syntax is as follows:

get '', '<row Id>'

by using this command we can read the data from an HBase table, but only one record, where the row Id of the row equals the second parameter of the command '<row Id>'. We can also specify the column showing in the result by using the following svntax:

get'', '<row Id>', {COLUMN ⇒ '<column family>:<column name>'}

This command can be modeled with relational algebra as follows:

get 'T', 'r1', {COLUMN
$$\Rightarrow$$
['cf1:c1', 'cf1:c2',...,' cf1:cn', cfm:c1',
'cf2:c2',...,'cf2:cn']}
=
 $\pi_{cfi:cj,i=1,2...m,j=1,2...n} \sigma_{rowkey='r1'}(T)$

T is an HBase table,

4.2 Group by

Group by is often used in aggregate functions, it allows grouping tuples by the value of an attribute and applies an aggregate function for each group.

By default HBase does not support group by and aggregate functions, but it's possible to perform these tasks on data by using the MapReduce framework.

In the "Shuffle and Sort" phase [7], MapReduce performs sorting and grouping by key to ensure that the input parameter of a reducer is a set of tuples t = (k, [v]) where [v] is the collection of all the values associated with the key k. A reducer can call the aggregate functions on this list of grouped values.

Group by is an additional relational algebra operation [13]. We can present the MapReduce function that performs the Group by operation by this algebra calculation:

Algorithm 1.
selectedColumn $\leftarrow \pi_{c1,,cn}$ T
for each unique k in π_{c_G} selectedColumn
for all Row $r \in selectedColumn$
v[] ← $\pi c_A \sigma c_{G=k}$ r
return reducer(k,v[])

Where T is an HBase table, c is a set of selected column, c_A is the column on which will be applied an aggregate function, c_G is the column on which the grouping will be performed and v[] is a set of values of c_A column grouped by c_G .

4.3 Aggregate Function

Aggregate functions: are functions that will group the values of multiples rows. They have applied on a numeric column and return a single result for all selected rows or for each group of rows. Also for aggregate functions, HBase does not support these functions, but by using the MapReduce framework or HBase Coprocessors EndPoints we can implement them. Common aggregate functions include: count, sum, avg, max, and min. We can present these functions in relational algebra as follows:

 $G_1,\ldots..,G_kg_{F1(C1),\ldots..,Fn(Cn)}(T)$

Where G is a set of grouping column, each C is one of the columns qualifiers of the HBase table T. each aggregate function F will be applied for each group according to G_1, \ldots, G_k .

In HBase these functions can be handled by using the MapReduce framework or HBase Coprocessors EndPoints. We will treat three functions: count, sum, and avg. In this paper we will use MapReduce.

Count

The goal is to enumerate all the distinct value of a column in an HBase table, with the number of times that they are present within the table for each of them

Algorithm 2 . Driver class for Count operation

```
class driver
method main(...)
scan ← HBaseTable.scan()
scan.addColumn(columnFamilyName, columnName)
jobStart
```

Algorithm 3 . Mapper class for Count operation

```
class Mapper
  method Map(key, cellValue)
   Emit(cellValue, 1)
```

```
Algorithm 4 . Reducer class for Count operation
```

```
class Reducer
  method Reduce(cellValue, integers [1,1,1,1,1,....])
    count ← 0
    for each integer i ∈ integers [i<sub>1</sub>,i<sub>2</sub>,....] do
        count ← count + i
    Emit(cellValue, integer count)
```

The driver class, which runs on a client machine, is responsible for scanning the HBase table, selecting the grouping column, configuring the job and submitting it for execution.

The Mapper class will produce a list of pairs (key, value) $[(k_1; v_1)]$. Before being sent to the reducer class, the file is automatically sorted by key by Hadoop in the "shuffle & sort" phase.

The Reducer class; it will receive a group of pairs (key, values) $(k_1; [v_1, v_1, ...])$ as input. Its role will be kept the unique key, calculates the sum of the values of all the pairs (key, values) received as input, and to generate a single pair (key, value) $[(k_2; v_2)]$ as output, composed of the unique key and the obtained total.

The Count function can be handled also by using the HBase commands "count" or "get_counter" but without the "group by" feature.

Sum

It calculates the sum of a column in an HBase table containing numeric values.

Algorithm 5. Driver class for Sum operation

```
class driver
method main(...)
scan ← HBaseTable.scan()
scan.addColumn(columnFamilyName1,columnName1)
scan.addColumn(columnFamilyNamen,columnNamem)
jobStart
```

Algorithm 6. Mapper class for Sum operation

```
class Mapper
  method Map(key , row)
  Emit(row.groupingColumn, row.sumingColumn)
```
Algorithm 7 . Reducer class for Sum operation

```
class Reducer
  method Reduce(string groupingColumn, values
  [v1,v2,...])
  sum ← 0
  for each value v ∈ values [v1,v2,....] do
    sum ← sum + v
  Emit(groupingColumn , sum)
```

The function "sum" has the same principle of the function "count" the only difference is in the Map phase; that performs for each row associates for the grouping column the value of the summing column instead of the value of 1.

Avg

It allows calculating an average value of a column in an HBase table containing numeric values. Avg function has the same driver class of the sum function. In the mapper class, the map function sends a series of pairs (key, value) composed by the grouping column as a key and the value of the column on which the average will be calculated as value. Then, in "shuffle & sort" phase, Hadoop performs the sorting by key. Therefore, the Reducer can sum the values then calculate the average by dividing the sum by the number of items in the set of the values.

Algorithm 8 . Mapper class for Avg operation

```
class Mapper
 method Map(key , row)
 Emit(row.groupingColumn, row.avgColumn)
```

Algorithm 9. Reducer class for Avg operation

```
class Reducer
method Reduce(string groupingColumn, values
[v1,v2,....])
sum→0
count← 0
for each value v € values [v1,v2,....] do
sum ← sum + v
count ← count+1
avg ← sum / count
Emit(groupingColumn , avg)
```

4.4 Join

Join in relational database allows associating several tables in the same query. It allows exploiting the power of relational databases to get results that combine efficiently data from multiple tables. HBase doesn't support the join operation, but it can be handled by using the declarative query languages that built on top of Hadoop like Pig [17] or Apache Hive [16] that launches implicitly MapReduce jobs for joining two HBase tables.

There are various join processing algorithms for MapReduce [21] environment like Repartition Join [15], Broadcast Join [15], and Trojan Join [18]. Following is the pseudo code of the Repartition Join, the most commonly used join algorithm.

Algorithm 10 . Map and Reduce functions for Repartition Join algorithm

```
Map (K: null, V : a record from a split of either R or
L)
  join_key ← extract the join column from V
  tagged_record ← add a tag of either R or L to V
  emit (join key, tagged record)
```

Algorithm 11 . Reducer class for Repartition Join

Repartition Join [15]

In our case, to joining two HBase tables T1 and T2; the input parameter is a row of either T1 or T2 as a value. The first step in Map phase is to extract the join column from the row then adding a tag for each row to identify its originating table in Reduce phase using the secondary sorting [14], and finally emitting the pairs (k,v) where k is the join key and v is the tagged row.

Then the MapReduce framework is the responsible for the partitioning, sorting and merging tasks. In these tasks the framework sorts by key and sends all the rows with same join key to the same reducer.

For the Reduce phase, the input parameter is a pair of $(k_1; [r_1, r_2,...])$ where k_1 is the join key and $[r_1, r_2,...]$ is a list of tagged rows associated to the k_1 key. The joining

operation is performed by splitting and buffering the tagged rows in two sets according to the table tag and handles the cross-product of the two sets.

5 Conclusion and Future Work

In this paper we proposed a feasibility study of migrating relational databases to HBase databases by applying the operations of the relational algebra in HBase data model and explore the implementation of these operations on HBase by using the native functions of this DBMS and also byusing the MapReduce Framework. Based on the above sections we can deduce that is theoretically the migration between relational databases and HBase databases can be handled efficiently. In perspective, we envisage to compare the performance of execution of the commons relational operations (bulk load, select, update, delete, join, group by, and aggregate functions) over a large database in relational and in HBase.

References

- 1. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM 13(6), 377–387 (1970)
- Moniruzzaman, A.B.M., Hossain, S.A.: Nosql database: new era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191 (2013)
- 3. Codd, E.F.: The significance of the SQL/data system announcement. Computerworld **15**(7), 27–30 (1981)
- 4. George, L.: HBase: the Definitive Guide. O'Reilly Media Inc., Sebastopol (2011)
- Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM 51(1), 107–113 (2008)
- Abadi, D.J.: Data management in the cloud: limitations and opportunities. IEEE Data Eng. Bull. 32(1), 3–12 (2009)
- Yang, H.C., Dasdan, A., Hsiao, R.L., Parker, D.S.: Map-reduce-merge: simplified relational data processing on large clusters. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, pp. 1029–1040. ACM (2007)
- 8. Apache HBase Databases. http://hbase.apache.org/
- 9. Dimiduk, N., Khurana, A.: HBase in Action. Manning, Shelter Island (2013)
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.E.: Bigtable: a distributed storage system for structured data. ACM Trans. Comput. Syst. (TOCS) 26(2), 4 (2008)
- Zhao, G., Huang, W., Liang, S., Tang, Y.: Modeling MongoDB with relational model. In: 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), pp. 115–121. IEEE (2013)
- 12. Khurana, A.: Introduction to HBase schema design. White Paper, Cloudera (2012)
- 13. Ceri, S., Gottlob, G.: Translating SQL into relational algebra: optimization, semantics, and equivalence of SQL queries. IEEE Trans. Softw. Eng. 4, 324–345 (1985)
- Lin, J., Dyer, C.: Data-intensive text processing with MapReduce. Synth. Lect. Hum. Lang. Technol. 3(1), 1–177 (2010)

- Blanas, S., Patel, J.M., Ercegovac, V., Rao, J., Shekita, E.J., Tian, Y.: A comparison of join algorithms for log processing in mapreduce. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, pp. 975–986. ACM (2010)
- 16. Apache Hive TM. http://hive.apache.org/
- Olston, C., Reed, B., Srivastava, U., Kumar, R., Tomkins, A.: Pig latin: a not-so-foreign language for data processing. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, pp. 1099–1110. ACM (2008)
- Dittrich, J., Quiané-Ruiz, J.A., Jindal, A., Kargin, Y., Setty, V., Schad, J.: Hadoop++: making a yellow elephant run like a cheetah (without it even noticing). Proc. VLDB Endow. 3(1–2), 515–529 (2010)
- 19. Webber, J., Robinson, I.: The Top 5 Use Cases of Graph Databases, Neo Technology (2015)
- 20. LarbretB.: Hadoop HBase Introduction (2015). https://www.slideshare.net/larbret/ hadoop-hbase
- Shaikh, A., Jindal, R.: Join query processing in mapreduce environment. In: Advances in Communication, Network, and Computing: Third International Conference, CNC 2012, Chennai, India, February 24–25, 2012, Revised Selected Papers, vol. 108, p. 275. Springer (2012)

Big Data and IoT: A Prime Opportunity for Banking Industry

Abdeljalil Boumlik^(☉) and Mohamed Bahaj

Faculty of Sciences and Technologies of Settat, LITEN Laboratory, Hassan 1st University, Settat, Morocco boumlik.abdeljalil@gmail.com, mohamedbahaj@gmail.com

Abstract. Banking industry is one of the most complex and sensitive industries that experience enormous changes in daily basis. Likemany others businesses, Big data is a serious problematic, data management and real time monitoring fraud issues also are even bigger challenges in this sector, due to the huge quantity of data, coming swiftly and rapidly from different devices in structured and unstructured formats, waiting for instantaneously treatments and decisions. Most financial institutions and banks try to innovate and diversify payment processes to make it more challenging and secure to improve their digital skills. Understand customer's behaviors also become a successful key factor in the market at the same time, that's why Internet of Things (IoT) can be the best solution to solve the issue of collecting and sharing data via internet among different "things", as devices and objects (Sensors, ATMs, POS, Smartphones, Computers, payment gateways (ecommerce), notebooks, etc.). The architectural and technical sides remain a problem, since conventional database management system and existing banking systems are not capable anymore to handle, store and process this massive volume of data with sufficient real time. This paper, discuss Hadoop Distributed File System and MapReduce, as an architecture for storing and retrieving information from massive volumes of datasets that we can collect via Internet from different objects based on the advantage and potential of Internet of things.

Keywords: Big data \cdot Internet of thing \cdot IoT \cdot Hadoop \cdot HDFS \cdot MapReduce \cdot Fraud monitoring

1 Introduction

In recent years, data is increasing at dreadful frequency day after day, therefore manage, explore and visualize this big data is becoming a new challenge. This challenge includes analysis, capture, search, sharing, storage, transfer, querying and information confidentiality, not just about being huge in size, but the aim is how to profit and manipulate the big data we have. Traditional data processing applications are no more capable to deal with it. There is an unlimited interest and benefit to use big data technologies to increase productivity and economy of all business sectors to build efficient investment strategies in different private and public domains. These technologies have a high capability to expect conclusions, with low cost consumption in materiel perspective, increase

efficiency and improve decision-making in various areas like banking, finance, fraud control, real time monitoring and transactions processing. Banking and financial sectors consider as one of the most complex and sensitive worlds, who's suffering from scalability and inefficiency problems when processing or analyzing this data, also the performance issues become worse due to huge amount of data that are not be handled anymore by conventional database management system or processed by traditional systems architectures. Fraudulent transactions in real time are increasing each day, as an example BGFI Bank was losing 1.9 billion CFA [10] on 2017 due to credit card fraud. Recognizing large-scale patterns across several transactions and identifying strange behaviors in the banking systems from an individual or multiple users can change, and prevent bank loses as online fraudulent transaction. Therefore, banks specialists adopt important attention to new projects that combine between performance and analytics to prevent such type of major problems that cause financial loss to the banks in real time. From other side, the number of transaction made by cardholder's peer second was extremely increased, and become unmanageable by current traditional systems, thus it causes a delay during the treatment of requests, that become very significant and impact banks business strategies.

Recently, the biggest banks and financial institutions realize that their business is not just around processing transactions, but it's about engaging and distributing value through the transaction lifecycle by offering multiple services via diverse payment channels and several devices like Mobiles, Point of Sale, Contactless, Near Field Communications (NFC), wearables, Tokenization, Biometric chip, Smart Kiosk, Smart ATM in order to get have opportunity to increase their volume of transactions and generate more revenue. For this reason, the Internet of Things (IoT) emerge and offers this potential advantages to benefit from the market dynamics and trends.

This paper is mainly implemented for banking domain, in which we present bank systems limitations and emergence needs for collecting data via different devices through internet and have also a fast access to such huge databases that requires an effective computing model. Based on that, we proposed also, a more sophisticated architecture that integrate Hadoop framework to get a complete system that deals with banks data and online fraud monitoring. Next, improve the architectures scalability and efficiency using big data environment, which is implemented on Hadoop (HDFS) distributed file system and MapReduce model as one of the most generally used parallel computing platforms for processing, stores and retrieve data from massive volume of datasets across multiple devices and sensors. The rest of this paper is organized as follows. Section 2, discusses related works and approaches that cover this common area. Section 3 presents a general view on Hadoop framework, MapReduce model and Internet of Things (IoT), then we highlight the advantages of this technologies Sect. 4, describes briefly the proposed solution from Transaction flow till Fraud detection systems. Implementation and evaluation is presented in Sect. 5. Finally, we conclude this paper, and discuss the future steps of our work.

2 Related Work

We can found several approaches that deal with fraud system in banking, finance, insurance, medical domains [1, 4, 5, 7, 8] depend on the needs of each institutions, but most of them are very simple and cover only limited scenarios or single parties of fraud, hence there is no enfranchisement and new challenges that get evoked. lasselare et al. presented a credit card fraud detection system that uses several intrinsic features to perform the detection process [1]. This technique is based on the buying behavior of the customer. The major parameters utilized are regency, frequency and monetary levels. These properties are used to predict frauds. However, the author was limited only on credit card fraud type and not a global system. A rule based fraud detection method that provided huge improvements in the detection process is described in [6]. This proposes to be a real-time system that has been implemented in a Turkish Bank. A cost sensitive credit card fraud detection method that uses Bayes Minimum Risk classifier is presented in [7]. The author claims to provide realistic views of the monetary gains and losses occurring due to fraud detection. A method that concentrates on providing effective fraud detection using imbalanced data is presented in [8], this solution is considers an extremely sparse and imbalanced data environment for performing the fraud detection process. An evaluation of accuracy provided by the Hadoop MapReduce environment on the Credit Card Fraud detection data is presented in [9], the Negative Selection algorithm is parallelized in the Hadoop environment for determining the accuracy. Authors Mahmoudi et al. presented a Modified Fischer Discriminant Analysis based anomaly detection method [2]. This technique uses Fischer Discriminant function to identify anomalies and Fraud. Halvaiee et al. presented an Artificial Immune System (AIS) based fraud detection model in paper [3]. This technique utilizes AIS to identify legitimate transactions from the fraudulent transactions it also limited to one type of fraud rules. Paper [4] present an ANN (Artificial Neural Networks) technique in the fraud domain based on the machine learning technique. Authors of paper [5] use a technique that utilize several classifiers, groups their results to identify fraudulent transactions.

From above results confirmed that recent solutions that deal with fraud systems, doesn't involve Big Data technologies as it should, none of them include Internet of Things (IoT) or discuss the way that they collect data. There are few articles that use Hadoop and MapReduce and only some of them involved the concept of IoT. We also conclude that these solutions have at least few defects or limitations. Moreover, we take the above as a source of motivation to provide efficient solutions based on IoT and Hadoop from architectural point of view and performance perspective.

3 Internet of Things and Hadoop

3.1 Internet of Things for Banking Domain

The Internet of Things (IoT) represents new opportunity for financial institutions and banks to find out more about what their customers require based on the information revolution, which offers an extraordinary level of data and data-driven customer services. Financial institutions realize that the customer experience becomes a key differentiator to identify new ways to distinguish themselves in the marketplace, and to deep existing relationship and increase the customers based on today's competitive environment. By applying this technology, banks will provide exceptional services, and adaptable financial solutions and advices, that closely associate with day to day events in customer's lives that will impact positively the bank's revenues and gain many competitive advantages. It is a network of billion devices connected through internet, by doing so, this become an intelligent system of systems. These devices present in Fig. 1 can collect data that allows banks to provide a complete view of customer's finance status in real time. Consequently, banks can anticipate customer's needs through data collected and analyzed, then provides solutions that can helps customers take sound and smart financial decisions.



Fig. 1. Different source of data for financial institutions.

3.2 Hadoop Framework to Process Big Data

Hadoop is an open source distributed processing framework from Apache. Developed in Java language, Hadoop used for storing a very large amount of data sets in different huge number of computer clusters. Hadoop's key advantages are related to his ability and flexibility during processing of large scale data, manage and control hardware failures or fault tolerance in the software level, cost effectiveness, scalability and robustness in real time processing. The Hadoop framework core consists of three major components that are Hadoop Distributed File system (HDFS), NameNodes and DataNodes. The Hadoop distributed file splits data and store it in large blocks to different nodes in the cluster system. Therefore, performance, access, operations and visualization of big data will be executed in parallel throw Hadoop HDFS layer, which means data will be processed faster and more efficiently than it would be in any other most conventional super-computer architecture. Hadoop makes replication of data automatically, due to Master/Slave architecture in which the Master called NameNode and Slave called Data-Node. NameNode is the responsible part of managing file system and mapping of files to their considered blocks which knows which data node stores which blocks, manage block replication, store all metadata in the RAM...etc. Otherwise, DataNode is a slave node consist to stores and reads blocks of files on top of native host (file system), also is responsible to forward a stored block to another server on another frame and replicate to a third server. Below we present the HDFS Architecture (Fig. 2).



Fig. 2. Hadoop HDFS architecture

3.3 MapReduce Model

In addition to the HDFS, MapReduce is core part of Hadoop. It is a programing model, which allows parallel processing of large volume of data. The MapReduce concept is simple and easy to understand, below is a graphical representation for all logical data flow with key functions (Fig. 3).



Fig. 3. MapReduce logical flow

The MapReduce jobs contains two major tasks, Map and Reduce are prepared to dividing whole workload into number of tasks and distributing them over different machines in Hadoop cluster. The Map task refers to a job that perform filtering and sorting. It takes input data, create a key/value pairs, and prepare them in a queue, as results and it will be sorted and sent to the Reduce task. In Fig. 4, we illustrate the graphical representation of the Map job flow.

3.3.1 Mapping Phase

See Fig. 4



Fig. 4. Map job architecture

3.3.2 Reduce Phase

After all the Map tasks, have completed successfully, the master controller combine and aggregate the results from each Map task and process them to be as a sequence of keyvalue pairs. At that time, the Reduce job correspond to this request take as input the returned Key and linked values to it as illustrate in the Fig. 5.



Fig. 5. Reduce job architecture

Also, reduce job, combine the values of the input key by reducing the list of retrieved values corresponding to the initial occurrence. Accordingly, the reduce job generate the output as a set of key-value pairs.

4 Research Design and Methodology

4.1 Overview

Recently, banks and financial institutions reconsidering how they model their enterprises. The statistical modeling and analytics insights become a key role in the industry to improve optimization, forecasting and operation decision process. Therefore, Banks and financial institution continue to focus on revenue progress and higher borders through operational efficiency, and especially better risk management, and improved customer intimacy. All these above and others factors force the innovation and invent solution and software architectures to deal with these challenges with necessity of taking into consideration duration and costs front of advantages to lead in the markets. The banking systems also have the problem of Volume, Velocity and Variety and it's considers as exciting factor which mean one of the most reason behind Big Data innovation to capitalize this data for strategic advantages depends on the niche (Fig. 6).



Fig. 6. Big data's important 3 axes.

4.2 Risk Management and Fraud Systems

In this paper, we focus only on fraud monitoring systems as a first step, we describe how things works with this type of systems, the current architecture used to build a full Fraud system, and the issues occurred on this type of architectures. Using Big Data technologies in this kind of services can make enormous changes on the current version of processing in performance perspectives, which is considered as a key factor and most important aspects in real time fraud monitoring logic, because every millisecond become very important to prevent financial impact on the banking institutions.

4.2.1 Existing Fraud Systems and Limitations

Today's financial institutions need a real-time automated system to detect fraud through multiple channels and masses of transactions a day. Current architecture used in Banks does not reach yet the level of sensitivity required in this type of systems, either, most of uncomfortable performance systems are mostly fraud systems due to the large size of data that need to be processed and compared with defined banking rules before triggering a financial decision. All the above, should be done in few seconds, even milliseconds to improve effectiveness and impact of a risk.

4.2.1.1 Limitations

Existing architecture will never help financial institution to become more efficient to detect fraud, due to many reasons that we cite below with current architecture too:

- None of the system use the IoT technology to collect data.
- Needed use of ETLs for Extract, transform and load functions.
- Lose original raw data for any new information after processing.
- Limitation in term of processing common pool of storage data.

- Considerable amount of time during refreshment of data through Real time dashboards.
- Many applications that act on the data stored on relational databases and obtain required information.
- The collection layer (big data layer) give raw data from different sources, which lead to a delay on the treatment before reaching back the requested system due to the diverse data structured and unstructured format.

4.2.1.2 Existing System's Designs

In the below Fig. 7, we try to describe the main workflow of the banking systems, without including all parties. As you can see, we have many data sources and some of them are linked with online applications, which means that we have an activity for 24/7 nonstop that should be manageable. All data are stocked in different databases with structured



Fig. 7. Example of existing fraud systems' architecture

and unstructured format due to difference in channels and communication's level. The access to such data require many development skills, controls and ETLs that should be done on the application layer to represent data correctly, also, the way used to access to different databases still traditional (First in, first out). All the above impact consumed timing during the execution of queries, retrieve information, and trigger corresponding actions.

4.2.2 Proposed Solution and Design

In this article, we propose a system's design using powerful big data technologies like Hadoop and MapReduce. The proposed design works on providing a more efficient and more exact fraud detection system. It consists to integrate Hadoop framework inside current system's architecture and use MapReduce algorithms to get a direct impact in



Fig. 8. Presentation of our proposed solution

term of performance frequency, and information distribution that is ignored in many previous solutions and in the current systems too.

The main objective is to identify fraudulent claims before it happens or during the processing of any online transaction that meet any fraud rule defined by the banks risk department in the database. In our architecture, we focused on MapReduce, one of the crucial enabling approaches for meeting increased Fraud systems demands by using high parallel processing, data storage, analytics and online processing on a large number of commodity nodes.

The suggested solution and architecture is presented in Fig. 8, in which we describe each element and the powerful role of Hadoop and MapReduce during the treatment to achieve the best effectiveness and efficient for detecting fraud and reduce the execution and cost.

In the above architecture, we will focus only on the role of Hadoop and his impact on the processing. In this architecture, the source of data is presented in the Top of the Fig. 7. This, include many data collected using Internet of Thing (IoT) technology to collect data via different channels, sources devices and actors in both mode Online (real time) and offline (processed by machines or human) before it came to the staging table to be processed with the core banking system. In the below section, we will explain all parties, elements and in the architecture, with their key roles.

- HDFS: As you know, HDFS provides Hadoop's efficient scale-out storage layer. It is used to serve and allowing wide variety of data access methods to operate on data stored in Hadoop. The required data by Fraud module will be kept for both Online and Batch processing in real time in Hadoop's Distributed File System. The main of our solution in this architecture is to use Hadoop sub programs like MapReduce and Machine learning to understand fraud patterns and trigger the rules matching.
- MapReduce: is used to be able to connect powerful large clusters of computers. In our proposed design, MapReduce is applied to large batch and online orientation processing of transactions, and, organize and reduce the result of the map job from each node into unified response to a query. In addition, this is the reason why MapReduce program is designed for applications such as monitoring and stream processing.

From initial testing perspective, we conclude that the integration of Hadoop will solve the performance issue in term of response-time during online transactions. Which means, that the full core banking system will be impacted positively in time consumption propose and more efficient during the generation of alert process, thanks to the powerful MapReduce algorithm that detect and much fraudulent rules with fraudulent activities and respond back to separate Fraud system in order to decline the processing of the concerned transaction. The parallel processing provided by MapReduce on large number of commodity nodes is a very good adventure in this type of systems in which time matters and quick action make difference.

5 Conclusion

In conclusion, our paper introduces some preliminary knowledge of Fraud systems its fraudulent rules and behaviors. The data become very important to detect fraudulent behavior that's why Internet of things because a major factor that collect huge data from different devices and share them via internet to prevent any fraudulent actions that's why data in banking institution become big data after each day that is why we tried to create a new design to involve big data technologies in Fraud systems as a new generation of fraud monitoring and fraud risk management in both type's Real time and Offline. Our analysis of Big Data technologies like Hadoop and MapReduce proves its huge potential, to reduce the detection and/or processing time of treatment to prevent Fraud before it happens. Another important advantage that we offer in our proposed system was the ability to handle all types of fraud and not limited to single type or scenarios, which mean a global solution based on IoT. All this, leads to the conclusion that the best solution for detecting fraud in the banking domain system is, at present, the optimized response during treatment of transaction in real time, and the optimized and research in terms of technologies and in terms of models of analysis, which make a huge difference between all international fraud systems providers.

References

- Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B.: APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support Syst. **75**, 38–48 (2015)
- Mahmoudi, N., Duman, E.: Detecting credit card fraud by modified fisher discriminant analysis. Expert Syst. Appl. 42(5), 2510–2516 (2015)
- Halvaiee, N.S., Akbari, M.K.: A novel model for credit card fraud detection using artificial immune systems. Appl. Soft Comput. 24, 40–49 (2014)
- West, J., Bhattacharya, M.: Payment card fraud detection using neural network committee and clustering. Comput. Secur. 57, 47–66 (2016)
- Zareapoor, M., Shamsolmoali, P.: Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Comput. Sci. 48, 679–685 (2015)
- Duman, E., Buyukkaya, A., Elikucuk, I.: A novel and successful credit card fraud detection system implemented in a Turkish bank. In: 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW). IEEE (2013)
- Bahnsen, A.C., et al.: Cost sensitive credit card fraud detection using Bayes minimum risk. In: 2013 12th International Conference on Machine Learning and Applications (ICMLA), Vol. 1. IEEE (2013)
- Wei, W., et al.: Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web 16(4), 449–475 (2013)
- Hormozi, E., et al.: Accuracy evaluation of a credit card fraud detection system on Hadoop MapReduce. In: 2013 5th Conference on Information and Knowledge Technology (IKT). IEEE (2013)
- http://www.jeuneafrique.com/404469/economie/gabon-bgfi-bank-secouee-fraude-massiveaux-cartes-visa-prepayees

Big Data Analytics Applied for Control Systems

Yousef Farhaoui^(⊠)

ASIA Team, Department of Computer Science, Faculty of Sciences and Technics, Moulay Ismail University, Errachidia, Morocco youseffarhaoui@gmail.com

Abstract. Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, search, sharing, storage, transfer, visualization, querying, and updating and information privacy. However, these huge data cannot easily handle since the most of CS systems are relational, and an adjustment is needed before any processing. With emergence of Big Data, new NoSQL systems come to deal with this relational data issue. So, we propose an approach to migrate historical CS data from relational to NoSQL system, and use a distributed environment containing many nodes. As experimentation, we migrate data generated by an oil and gas CS to an appropriate distributed NoSQL system, and we perform some data mining experiments on them in order to compare results and prove the obtained performance.

Keywords: Big data \cdot Data mining \cdot Control system \cdot NoSQL \cdot NewSQL \cdot Analytics

1 Introduction

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, and updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem."

Big Data could be defined as an emerging phenomenon which refers to the practice of treatment of large and complex data volumes, with technical systems associated such as algorithms used to visualize and analyze real-time or not (real-time or batch) these massive data, to create added value for the organization. In parallel, the Data Mining also experiencing a rapid development (neural networks, genetic algorithms...) and it is now possible to create self-learning IT structures. This is the Machine Learning: analysis and implementation of automated methods that allow a machine (at large) to evolve through a process of learning, and so perform tasks that are difficult or impossible to fill by more conventional algorithmic means.

ON-RELATIONAL and relational data models are different. The relational model takes data and separates it into many interrelated tables that contain rows and columns. Tables reference each other through foreign keys that are stored in columns as well. When querying data, the requested information will be collected from many tables, as if the user asks: what is the answer to my question?

Non-relational data models often start from the application-specific queries, as opposed to relational data models. Non-relational data model will be then driven by application-specific access patterns. An advanced understanding of data structures and algorithms is required [1], so that the main design would be to know: what questions fits to my data?

In this paper we will provide in the introduction, a preview of the different nonrelational data models. Next, we will focus on document databases in Sects. 2 and 3, by discussing evaluation criteria related to this data model. Finally, we will address the evaluation results in the conclusion.

Industrial companies and manufacturers are increasingly equipped with control systems (CS) that generate very large amounts of real time data. These data are used by specific applications to provide real time critical information, in-time graphs of evolution, real-time alarms, etc. Later, these data are stored for historical archives, and in many cases they are often deleted later. However, lots of users are increasingly interested in these historical data in order to use them in many business processes, and especially in data mining area, like extracting useful knowledge, providing early feedback means, improving future requests, etc. Nevertheless, this huge accumulated data needs progressively capacity and power to support processing and storage. Every machine or server arrives at its limits to support the storage and processing of huge data, whatever its physical capacity CPU, memory, or disk. Also, machines' upgrade or extension cannot be considered as a permanent solution. So, distributed platforms contained a lot of servers (nodes), such as clusters, grids can be efficiently used to deal with this issue.

This article has five sections. The first section is an introduction, and in the second, we present relational data issue on distributed environment and some related work. We give in the third section an approach to use distributed NoSQL system and data migration. In the fourth section, we prepare CS data for a company [2] on an experimental platform. In the fifth section, we carry out some interesting queries to test and compare performance. A conclusion closes this paper at the last section.

2 Characteristics of Big Data

It is therefore important to understand the Big Data - Volume, Speed and Variety.

2.1 Volume

The volume describes the amount of data generated by companies or individuals. Big Data is usually associated with this feature. Firms across all sectors will need to find ways to manage the ever-increasing volume of data that is created daily. Catalogs of more than 10 million products have become the rule rather than the exception. Some customers managing not only products but also their own customers can easily accumulate a volume exceeding the terabyte of data.

2.2 Speed

The speed describes the frequency at which data is generated, captured, and shared. Due to recent technological developments, consumers and businesses are generating more data in much shorter time frames. At this level of speed, companies can only capitalize on these data if they are collected and shared in real time. It is precisely at this stage that many analyses, CRM, personalization, point-of-sale and other systems fail. They can only process data in batches every few hours, at best.

2.3 Variety

The proliferation of data types from sources such as social media, Machine to Machine interactions and mobile devices creates a great diversity beyond traditional transactional data. The data is no longer part of a clear, easy-to-use structure.

New data types include content, geo-spatial data, hardware data points, geolocation data, connection data, machine-generated data, measurement data, mobile data, physical data points, processes, RFID data, Research data, trust data, flow data, social media data, text data, and web-based data.

Why is it important to understand all this?

Because Big Data helps us to get a better representation of customer interaction with the company. It allows a better understanding of what customers would like to achieve at each point of contact.

It minimizes the risk of losing these customers when switching from one point of contact to another and ensures the relevance of the information that is delivered to them. Thus, to improve both the quality of service, key aspect for customers, and the transformation rate of these customers, it is important for the company not to lose sight of the Big Data.

3 Characteristics of NoSQL Databases

The term "NoSQL" was invented in 2009 during an event on distributed databases. The term is vague, incorrect (some NoSQL engines use variants of the SQL language, for example Cassandra), but has the advantage of having a certain marketing and polemic effect. In this part, we will discuss the general characteristics of NoSQL engines, historically, conceptually and technically, in relation to relational databases, but also independently of this reference.

3.1 Principle of NoSQL Databases

The NoSQL databases, especially the document-oriented ones, neglect the strengths of relational databases, namely the notion of registration and the relations between elements, in order to focus on the notion of a document. NoSQL databases are much more flexible and more scalable. The organizational structure is no longer linked to a relational scheme that is difficult to modify, and the basis can therefore grow without constraint.

On the other hand, the "document" orientation facilitates the deployment of the database on multiple machines. Automatically, of course. The developer is not concerned with the location of documents, split or not. When the database becomes too large, it is enough to define new machines connected on the network, and the NoSQL database gets by.

This is the answer to new applications demanding speed of processing and quantity of data managed. Quantities of the order of several hundred terabytes. Note that there are also the NOSQL bases of type "columns" and bases of type "graph". Column bases are an excellent solution for massive analysis. The graphic bases, more delicate to be apprehended, are, as their denomination indicates, more adapted to the resolution of the questions of organization in network (structure in arcs and nodes).

NoSQL systems use replication to achieve multiple objectives.

- Availability. Replication ensures that the system is always available. In the event of a server, node or disk failure, the task performed by the defective component can be immediately supported by another component. This technique of failover is an essential asset to ensure the stability of a system that can include thousands of nodes, without having to swallow a monstrous budget in monitoring and maintenance.
- Scalability (reading). If data is available on multiple machines, it becomes possible to distribute (read) requests on these machines. This is the typical scenario for the scalability of Web applications.
- Scalability (writing). Finally, one can think of distributing also the requests in writing, but there one is faced with delicate potential problems of competing writings and reconciliation.

The technique is very classic and used by all the DBMS of the world. Instead of repeatedly writing to the disk without a pre-defined order (so-called "random" accesses), which each time require a displacement of the read head and therefore a latency of a few milliseconds, one writes sequentially in a file of Log (log) and the data is also placed in RAM memory.

3.2 The Emergence of Big Data and NoSQL Databases

Software evolutions follow naturally the material evolutions. The first DBMSs were built around mainframes and depended on the storage capacities of the time. The success of the relational model is due not only to the qualities of the model itself but also to the optimization of storage that allows the reduction of the redundancy of the data. With the widespread use of network interconnections, increasing Internet bandwidth and lowering the cost of moderately powerful machines, new possibilities have emerged in the area of distributed computing and virtualization, for example.

The shift to the twenty-first century has seen the volume of data manipulated by some organizations, especially those related to the Internet, increase dramatically. Scientific data, social networks, telephone operators, medical databases, national territorial defense agencies, economic and social indicators, etc., the increasing computerization of all types of processing implies an exponential increase in the volume of data Now in petabytes. This is what the Anglo-Saxons called the Big Data. Managing and processing these data volumes is seen as a new IT challenge, and traditional, highly transactional relational database engines appear to be completely outdated.

4 Distributed Data Issue and Related Work

In this section, we present the constraints of the relational model in distributed world, the NoSQL technology, and some related works given to deal with this issue. A distributed environment is a set of physical machines (nodes) which participate jointly to accomplish parallel processing and storage. All resources of the nodes (CPU, memory, disks) can be shared or not. The number of nodes can differ from a few to thousands of nodes. So, we can find simple clusters with few nodes which typically share disks, or more complex like grids that contain hundreds or thousands of nodes where resources are often not shared.

The main parts of the CS databases are based on the relational model which is built on the concept of table (relation between data) and operations of set-algebra.

This model is suitable for transactional needs due to the ACID properties (Atomicity, Consistency, Isolation, and Durability) [3], and it works very well in a single-node environment. Conversely, relational data cannot be well distributed and deployed in a distributed multi-nodes environment. The ACID properties become constraints for the model and then prevent data from being distributed effectively between nodes [4].

Note also that ACID constraints, although they ensure consistency, may become in some cases a blocking factor [5, 6]. For instance, an investigator in internet is often interested in having immediate response even if that response is not up-to-date. Consequently, new systems need to be created in order to dynamically distribute and manage data between nodes with more efficiency and usefulness. New systems for Big Data have been emerged like NoSQL and NewSQL.

5 Data Model Evaluation Criteria

The following criteria will be used to evaluate the document oriented model:

- 1. Nature of data (structured, semi-structured).
- 2. Data relationship (referential integrity, hierarchical relationships).
- 3. Data life-cycle (versioning, TTL).
- 4. Dataview (CRUD operations).
- 5. Data consistency (ACID properties).

- 6. Performance (indexing, partitioning).
- 7. Storage volume (BigData).
- 8. Data analysis (data aggregates).
- 9. Persistency and fault tolerance (data replication).
- 10. Data security (access rights, data encryption).

In general, data can be structured or semi-structured, depending on the related usecase.

In the relational data model, the tables contain a set of rows where every row is grouping a set of values. Existing relationships are based on joins between rows in different tables and they are not based on the semantic relationships between keywords (values).

6 Proposal

As proposal, we suggest migrating CS data to a suitable NoSQL system. We choose a NoSQL system according to specific criteria which are predominantly based on the fitting of NoSQL class to CS data. Also, since CS data must be consistent, the type of the chosen NoSQL system must be CP (Consistency/Partitioning). Migrated NoSQLCS



Fig. 1. Writing with logging



Fig. 2. Proposed approach

data will be distributed on a platform consisted of many nodes. The Sharding is a NoSQL property which distributes and allocates dynamically resources between nodes and adapt them according to data needs. "Figure 1," represents a schema for our proposed migration. This can greatly improve the performance on either the data storage or the processing time for ad hoc queries (Fig. 2).

7 Preparation of Target CS Data

To prepare CS data, we start implementing a distributed platform, and we perform migration data from Oracle 12c to MongoDB version 2.61. Some experiments and their results will be done in the next section. Firstly, we configure a distributed platform managed directly by MongoDB.

8 Results and Discussion

The new CS MongoDB database is now ready; our goal is to prove performance acquired by using elasticity given by distributed NoSQLdata. This elasticity signifies the ability to distribute data and queries processing on multiple nodes, contrary to the rigidity of relational data. So theoretically, good performance is expected. In the following, we use the same data mining queries for both the distributed MongoDBdata and the Oracle mono-node data. We will compare and analyze the results of the storage, run time, and shared query processing by varying each time the number of shares from 2 to 10. As experiments, we use collections which containing realtime data used frequently in looking for frequent itemsets. For data storage, we can see in Fig. 3 a graph showing



Fig. 3. Comparison the CS data storage repartition between MongoDB and Oracle

that MongoDB has shared automatically CS data on different shards, but for Oracle only one node supports data storage, the other nodes remain idle.

9 Conclusion

As experiment, after configuring a distributed environment with a number of nodes, we have installed MongoDB and migrated CS data. The NoSQLsharding property allows repartition of storage and processing between all nodes. Some experiments Data Ming queries have been done in order to compare performance of results between multi-nodes NoSQL data and mono-node relational data. Overall, the final results demonstrate an interesting improvement in run time, in addition of the data storage gained by joining all disk nodes. Finally, in perspective we look forward to enlarge this experimentation in widespread platforms environments with many nodes like Hadoop platform. Also, since NoSQL is new-fashioned and still in development, we are hearing of potential new NoSQL systems to test them and find the most appropriate for CS data. Document oriented clusters provide highly scalable architecture and better system availability. For this reason this model is one of the most used NoSQL models on the worldwide.

References

- Kaur, K., Rani, R.: Modeling and querying data in NoSQL databases. In: BigData, 2013 IEEE International Conference. INSPEC Accession Number 13999217 (2013)
- Hashem, H., Ranc, D.: An integrative Modeling of BigData Processing. Int. J. Comput. Sci. Appl. Print ISSN 0972-9038 (2014)
- Sharma, V., Dave, M.: SQL and NoSQL Databases. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2(8), 2–8 (2012). ISSN:2277 128X. Research paper available: www.ijarcsse.com
- Degroodt, N.: L'élasticité des bases de données sur le Cloud Computing. Master thesis in Sciences computer, FreeUniversity of Bruxelles, pp. 12–20 (2011)
- Li, Y., Manoharan, S.: A performance comparison of SQL and NoSQLdatabases. In: Communications, Computers and Signal Processing, 2013 IEEE Pacific Rim Conference (2013). ISSN 1555-5798
- 6. Farhaoui, Y.: Big data and NoSQL system for control system. IJEFT 14(2) (2017)

Detecting Network Intrusions Using Multi-class Logistic Regression and Correlation-Based Feature Selection

Taha Ait tchakoucht^(IZI) and Mostafa Ezziyyani

Mathematics and Applications Department, Faculty of Sciences and Techniques, UAE, Tangier, Morocco {t.aittchakoucht, m.ezziyyani}@fstt.ac.ma

Abstract. Because they're facilitating life, using computers and other intelligent devices associated with internet has become vital in those days. Banking transactions, education, trade marketing, texting ... are all daily and important operations that relies on such technology. Information systems that handle those operations must be kept secure from any intrusive activity. To help ensure that, we must take into consideration several subjects such as access control by managing confidentiality, integrity and availability, as well as deploying detection and prevention tools and mechanisms that help preparing for and dealing with attacks. In this perspective, we propose a network intrusion detection model based on multiclass logistic regression (MLR) and Correlation-based feature selection (CFS). Results will be discussed with respect to NSL-KDD Dataset, and compared to other techniques based on various classification methods.

Keywords: Intrusion detection system \cdot Logistic regression \cdot Attack \cdot NSL-KDD

1 Introduction

Intrusion detection system is a set of tools destined for traffic analysis in order to capture any abnormal or malicious activity that threatens the continuity of service, and alert security administrators for possible reactions.

Two families of intrusion detection are to mention; Misuse detection and Anomaly detection. The misuse detection approach consists of recognizing attacks that follow intrusion patterns already recognized and reported by experts [11]. Misuse detection systems are vulnerable to intruders who use new patterns of behavior or who mask their illegal behavior to deceive the detection system. Anomaly detection methods were developed to overcome this issue. Anomaly detection is based on the behavior of the user and/or application [10]. A third approach would be the hybrid intrusion detection which combines both techniques in the same time [12]. Often, these approaches are based on Data-mining techniques. Data-mining Techniques is a set of techniques for the extraction of motifs from large data sets, combining statistical and machine learning

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_37 methods with database management. Those techniques involve learning association rules, cluster analysis, classification and regression.

An IDS could also be whether a Network-based IDS (NIDS) or Host-based IDS (HIDS). The objective of HIDS is to focus on a single machine while he NIDSs look at the packets that pass through the whole network to determine if an attack occurs. In Anomaly-based approach, they consist of establishing a network profile that separates between normal and abnormal activity.

The remainder of this paper is organized as follows: Sect. 2 represents some related work. Section 3 explains our research methodology. In Sect. 4, experimental results and performance comparison are discussed. Finally, Sect. 5 describes conclusion and opens to some perspectives.

2 Related Work

In their application, IDSs are based on machine learning techniques, such as support vector machine, Artificial neural network), K-means algorithm, Naive bayes, decision trees...

In [1], Quinlin introduces a machine learning model C4.5, an extension version of ID3 algorithm, based on decision trees where he discusses several topics like tree construction, pruning as a remedy against high variance, conversion to rules, problems related to missing attribute values. The system is widely used, but still suffers from some limitations like the bias in favor of rectangular regions.

When using Naïve Bayes methods, the most usual problem faced is that one of handling continuous variables. One solution considers discretizing or assuming that the data are generated by a single Gaussian. According to [2], this assumption can be violated in some domains, and instead of it, they're using statistical methods for non-parametric density estimation. As a result, the system generalizes better than the former approach. This approach to Bayesian induction bears some similarities to other research in machine learning and statistics.

SVMs are also popular machine learning techniques. One example which is widely used is LIBSVM [3].

3 Proposed Model

3.1 NSL-KDD

NSL-KDD [4] is a dataset proposed to solve some of the inherent problems of the KDD'99 [5]. Although, this reformed version of KDD dataset is still suffering from certain problems mentioned by McHugh [6] and may not be a perfect representation of existing real networks, due to the lack of public datasets for network based IDSs, it still can be applied as an effective benchmark dataset to help compare different intrusion detection models.

The NSL-KDD dataset has the following advantages over the original KDD dataset:

- It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.
- There are no duplicate records in the proposed test sets; therefore, the performance of the learners are not biased by the methods which have better detection rates on the frequent records.
- The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD dataset. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.
- The number of records in the train and test sets is reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select small portion. Consequently, evaluation results of different research works will be consistent and comparable.

The dataset is built on 41 features including:

- 1. **Basic features:** this category encapsulates all the attributes that can be extracted from a TCP/IP connection.
- 2. **Traffic features:** this category includes features that are computed with respect to a window interval and is divided into two groups:
 - (a) *"same host" features:* examine only the last 100 connections having the same destination host as the current connection, and calculates statistics related to protocol behavior, service, etc.
 - (b) *"same service" features:* examine only the last 100 connections that have the same service as the current connection.
- 3. **Content features:** unlike most of the DOS and Probing attacks, the R2L and U2R attacks don't have any intrusion frequent sequential patterns. This is because the DOS and Probing attacks involve many connections to some host(s) in a very short period of time; however, the R2L and U2R attacks are embedded in the data portions of the packets, and normally involves only a single connection. To detect these kinds of attacks, we need some features to be able to look for suspicious behavior in the data portion, e.g., number of failed login attempts. These features are called content features.

3.2 Preprocessing

Converting symbolic data to numerical data. In addition to the improvements presented in NSL-KDD we perform some preprocessing to prepare an appropriate dataset for usage in the algorithm.

Indeed, there are some categorical features that need to be converted to numerical data such as, *protocol_type*, *Service* (service type) and *Flag* (connection status flag).

Class	Reference number	Label
Normal	1	Normal
Probe	2	Ipsweep, portsweep, nmap, satan, saint, mscan
DOS	3	smurf, teardrop, pod, back, land, apache2, udpstrom, mailbomb, processtable, neptune
R2L	4	dictionary, ftp_write, guess_password, imap, named, sendmail, spy, xlock, xsnoop, snmpgetattack, httptunnel, worm, snmpguess, multihop, phf, warezclient, warezmaster
U2R	5	perl, ps, xterm, rootkit, loadmodule, eject, buffer_overflow, sqlattack

Table 1. Classes and references.

protocol_type counts 3 values, *Service* regroups 70 values and *Flag* is precisely an 11 values feature. So, we replaced every categorical value with a reference integer.

We also have attributed a reference number to each class type including the normal class, as shown in Table 1.

Figures 1 and 2 respectively show 2 original NSL-KDD examples with symbolic data and their converted version with numerical data.

0,tcp,ftp_data,SF,491,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,1.00 ,0.00,0.00,150,25,0.17,0.03,0.17,0.00,0.00,0.00,0.05,0.00,normal

Fig. 1. Data from original dataset

Fig. 2. Data from Fig. 1 after conversion

Mean normalization and feature scaling. One important operation in a machine learning problem with a set of features is to proceed to mean-normalization and feature scaling. We usually encountering a situation where the features vary in different ranges of values. Here for example, in one hand we have the *protocol_type(integer)* varying from 1 to 3. In the other hand, the *duration(real)* varies from 0 to 58329. This will cause the algorithm to converge slowly. Therefore, we have to make sure that the features are on a similar scale, and make them have approximatively zero mean.

In order to do so, we replace x_i with

$$\frac{\mathbf{x}_{i} - \boldsymbol{\mu}_{i}}{\mathbf{s}_{i}} \tag{1}$$

Where x_i the feature value, μ_i the mean value of feature x_i over all training examples, and s_i the standard deviation.

Dimensionality Reduction. For this purpose, we are using the Correlation-based feature selection (CFS) [9] which belongs to the category of filters;

Feature-class and feature-feature correlations are computed. Then, we search for the best dataset in the subsets space.

The CFS's feature subset evaluation function:

$$Ms = \frac{kRfc}{\sqrt{k + k(k - 1)Rff}}.$$
(2)

Rfc: Mean feature-class correlation. Rff: Average feature-feature inter-correlation.

Algorithm. To compute the cost value, we are using the following cost function:

$$\operatorname{Cost}(\theta) = -\frac{1}{m} \sum_{i=0}^{m} y^{(i)} \log(h_{\theta}(\mathbf{x}^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(\mathbf{x}^{(i)})) + \frac{\alpha}{2m} \sum_{j=1}^{n} \theta_{j}^{2}$$

$$h_{\theta}(x) = sigmoid(z) \text{ and } z = \theta_{0} + \theta_{1}x_{1} + \dots + \theta_{n}x_{n}$$
(3)

 $x^{(i)}$ refers to the *i*th example in the dataset. $x_1, x_2, \ldots x_n$ are the selected features. h_{θ} is the hypothesis function (predictor) algorithm. $y^{(i)}$ is the *i*th example's label value which is whether 0 or 1. m the number of examples in the dataset. n the number of selected features, Θ is parameter vector which size is n + 1, and α is regularization parameter.

To minimize that cost function, we first compute its partial derivatives:

$$\frac{\partial Cost(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)} \qquad j = 0$$

$$\frac{\partial Cost(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)} + \frac{\alpha}{m} \theta_j \quad j \ge 1$$
(4)

The cost value and the gradient (partial derivatives) will be passed to the minimization function fmincg [7]. Overall, we employ the following algorithm for Multiclass Logistic Regression MLR. (Syntax is a mixture of Matlab and algorithmic's + vectorized representation).

```
program MLR (Output)
KDDTrain20 Dataset [4]
const lambda:= 1;
var
         X: Real % Training set
         v: Real % class labels vector
        m, theta, theta0, grad, cost, hypothesis: Real;
       % theta: the weights vector. cost: the cost
function. grad: partial derivatives of 'cost'.
hypothesis: hypothesis function h_{\theta}
  begin
    cost := 0;
    m = length(y);
    grad := zeros(size(theta));
    theta0 := [0; \text{theta}(2:\text{end})];
    hypothesis := sigmoid(X*theta);
    cost := (1/2*m)*sum(-y.*log(hypothesis)-(1-
y).*log(1- hypothesis))+...
   (lambda/2*m)*(theta0'*theta0);
   grad := (1/m) *X'*( hypothesis -y)+(lambda/m)*theta0;
   cost := fmincg(cost,grad) [7]
   Use the new \theta resulting of the minimal cost value
and recompute 'hypothesis'.
   Compare the new 'hypothesis' and y to evaluate
   accuracy on training set and predict classes for new
   examples test set 'KDDTest+ and KDDTest-21').
```

end

4 Experimental Results

To evaluate the model, we selected 11 features using feature selection. We applied the first 20% examples of the KDDTrain+ as the training set.

We then experimented our trained model on 2 datasets, the KDDTest+ and KDDTest-21. For comparison, we have selected the widely used J48 decision tree machine learning technique [1] and Naïve Bayes [2] from the Weka [8] collection with the default values as the input parameters for these methods.

In order to have an optimal system, we used a validation set of 12597 records which is 10% of the original dataset, and focused the processing on the parameter α ;

Normally if α is too small we fall into a high variance problem (overfitting). If it is too big the system will suffer from high bias (underfitting). Eight models are tested. Every model refers to a specific value of α . Figure 3 represent those models.



Fig. 3. Variation of Cost with α

Remark 1. Bigger values of α result in an increase in the Cost_{VALIDATION}(θ). We chose the model whose Cost_{VALIDATION}(θ) is minimal, which is ($\alpha = 1$ and Cost_{VALIDATION}(θ) = 1.78e5). We then trained our system according to that model and tested it on KDDTest+ and KDDTest-21.

To have a good insight on how the system is behaving on new sets of examples especially when treating the rare classes (skewed data), we used the Precision/Recall and FScore metrics;

$$Precision = \frac{\# True \ positives}{\# True \ positives + \# false \ positives}$$
(5)

$$Recall = \frac{\# True \ positives}{\# False \ Negatives + \# True \ positives}$$
(6)

$$FScore = \frac{2.Precision.Recall}{Precision + Recall}$$
(7)

We denote TP: True Positives, FP: False Positives, FN: False Negatives, P: Precision, R: Recall and FS: FScore.

Tables 2 and 3 represent confusion matrix respectively on KDDTest+ and KDDTest-21.

	Normal	Probe	DOS	R2L	U2R
Normal	9400	143	167	1	0
Probe	244	1701	476	0	0
DOS	1906	46	5506	0	0
R2L	2616	153	118	0	0
U2R	56	5	6	0	0
FP-rate	34%	16,8%	12%	-	-
FN-rate = 21,4%					

Table 2. Confusion matrix on KDDTest+

Table 3. Confusion matrix on KDDTest-21

	Normal	Probe	DOS	R2L	U2R
Normal	1935	25	192	0	0
Probe	747	349	1306	0	0
DOS	1894	34	2414	0	0
R2L	2856	14	16	1	0
U2R	61	1	4	1	0
FP-rate	74%	17,5%	38,6%	-	-
FN-rate = 46 , 9 %					

Table 4 represent the FScore indicator for all classes on both KDDTest+ and KDDTest-21.

Remark 2. FScore show good detection performance of Normal, Probe and DOS classes *ow performance compa* (respectively 0.79, 0.76 and 0.80) on KDDTest+.

Remark 3. Results on KDDTest-21 are medium, something due to the characteristics of this Dataset (It contains instances that not all of the 21 Machine learners could

	KDDTest+			KDDTest-21		
	Р	R	FScore	Р	R	FScore
Normal	0,66	0,97	0,79	0,26	0,90	0,40
Probe	0,83	0,70	0,76	0,83	0,15	0,25
DOS	0,88	0,74	0,80	0,61	0,56	0,58
R2L	0	0	0	0,5	~0	~0
U2R	-	-	-	-	-	-

Table 4. Fscore on both KDDTest+ and KDDTest-21

detect), leading to high values of False positive and False negative rates which affects negatively the Precision and the Recall metrics and consequently the FScore.

Remark 4. Large number of R2L and U2R instances is considered as normal behaviour; 90% of R2L instances and 83% of U2R attack instances with respect to KDDTest+. 98.9% of R2L instances and 91% of U2R instances with respect to KDDTest-21. This happens because of the similarities between R2L, U2R and Normal instances and the lack of R2L and U2R training records (only 209 R2L and 11 U2R in the 20%KDDTrain+, whilst there are 2887 R2L and 67 U2R instances on each test set); As a result, the system hardly detected some R2L attacks and couldn't detect U2R attacks.

Figures 4 and 5 show performance comparison with [1, 2].



Fig. 4. Performance comparison with respect to KDDTest+



Fig. 5. Performance comparison with respect to KDDTest-21

5 Conclusion and Outlook

In this article, we presented a NIDS based on Multiclass Logistic Regression. The model consists of several steps going from preprocessing to detection. Usually, the accuracy of classifiers on the original KDDTest of KDD99' is relatively high given that this test set contains skewed data. The accuracy results of classifiers on KDD99' cannot reflect the ability of the classifier. However, even if the accuracy rate is generally medium on KDDTest+ and KDDTest-21, those two test sets are good indicator of classifiers capability.

The method resulted in good accuracy rates compared with other widely used machine learning techniques. The model can be enhanced using additional/alternative machine learning techniques, with more targeted feature engineering and refined preprocessing, in order to:

- Reduce the false positives rate (FP).
- Reduce the false negatives rate (FN).
- Decrease the training costs.
- Increase performance detection on R2L and U2R attack types.

References

- 1. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
- John, G., Langley, P.: Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, pp. 338–345 (1995)

- 3. Chang, C., Lin, C.: LIBSVM: a Library for Support Vector Machines (2001). http://www.csie.ntu.edu.tw/~cjlin/libsvm
- 4. NSL-KDD (2009). http://www.unb.ca/research/iscx/dataset/iscx-NSL-KDD-dataset.html
- 5. KDD CUP (1999). http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
- McHugh, J.: Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln Laboratory. ACM Trans. Inf. Syst. Secur. 3(4), 262–294 (2000)
- Rasmussen, C.E.: fmincg minimization function. http://learning.eng.cam.ac.uk/carl/code/ minimize/
- 8. Waikato environment for Knowledge analysis (Weka) and Using Weka in Matlab. http:// www.mathworks.com/matlabcentral/fileexchange/50120-using-weka-in-matlab
- 9. Hall, M.A., Smith, L.A.: Feature subset selection: a correlation based filter approach. University of Waikato (1997)
- Denning, D.E., Edwards, D.L., Jagannathan, R., Lunt, T.F., Neumann, P.G.: A prototype IDES: a real-time intrusion detection expert system. Technical report, Computer Science Laboratory, SRI International, Menlo Park
- 11. Roesch, M.: Snort lightweight intrusion detection for networks. In: Proceedings of the 13th Systems Administration Conference, pp. 229–238. Usenix Association, Seattle (1999)
- Jagannathan, R., Lunt, T.F., Anderson, D., Dodd, C., Gilham, F., Jalali, C., Javitz, H.S., Neumann, P.G., Tamaru, A., Valdes, A.: System design document: next-generation intrusion-detection expert system (NIDES). Technical report, Computer Science Laboratory, SRI International, Menlo Park (1993)

The Optimization of Search Engines to Improve the Ranking to Detect User's Intent

Salma Gaou^{1(\boxtimes)} and Aissam Bekkari^{2(\boxtimes)}

¹ LIROSA Laboratory, Abdelmalek Essaadi University, Tétouan, Morocco Salma.gaou@gmail.com
² IRF-SIC Laboratory, Ibn Zohr University, Agadir, Morocco

a_bekkari@yahoo.fr

Abstract. The major evolution of the search engine is to understand the user's query and the intention of the user. The other change in size and all of you is the evolution of the mobile query. Indeed, research on search engines is trying today to bring results of research adapted to the intention of the user. Understand the search intentions of Internet users and paramount to the search engines. But it is also for SEO. By understanding the user's intention to search, we can define the type of content to produce in order to maximize your chances of positioning.

In this article, we focus on detecting and understanding the user's intent that motivates a user to search the web. The analysis of the history of research is the detailed examination of data on the Web different users for the purpose of understanding and optimizing web management. In this article we propose a new approach to detect user intent through search engine optimization to improve the ranking of a website in organic search results to increase visibility and quality.

Keywords: Search engine optimization (SEO) \cdot Intent user \cdot Information search \cdot Ranking of search results \cdot Search retrieval

1 Introduction

In recent years, a large number of information search engines have emerged on the Internet. To find information or website on a particular topic, a user uses the search engine search query writing with words, key words or sentences form. Most queries are short, ambiguous and multifaceted. Several queries can refer to the same thing many times the same queries can involve different user's information needs. In fact, people frequently make queries with more of an unspecified purpose [1-3]. In addition, user's information needs are numerous, exploration, and a single query can involve different user's information needs. To query "swine flu", doctors may be interested in the pathogenesis of treatment solutions, while patients can turn for transmission and preventive measures. Recently understand the intent behind user queries attracted much attention in the search for information retrieval (IR) [4].

So, when you type in your search term and press the search, and the search engine will try to match your words with the best and most relevant web pages, it can find a web search.
In seconds, you will see thousands of keywords containing the word appear in your list. There are several things you can do with this list, but we'll start talking, we make a list, sorted by popularity or search volume, competition, the pages have been directly optimized for each keyword and KEI - index keywords efficiency.

2 Preliminaries

2.1 Related Work

We can see some Approaches to classify user intents. Generally, they can be divided into several categories. Jansen et al. in [7-9] present a methodology developed to classify the user's intent in terms of content type specified by the query and other expressions of the user, a set of characteristics for each category in the taxonomy and Broder reported three levels of categories for users, respectively, the last two are obtained from manually classified queries. [5] A class is trying to raise queries with additional data, including search results returned for a query, the information from an existing corpus or an intermediate taxonomy. The second category uses data unmarked to help improve the accuracy of supervised learning. The third category is developing the training data by automatically marking some queries to certain click-through data by self-training. The anchor text and the results of search engines and the query text are used to represent an application. Wu et al. [10] has a relationship of dependency and characteristic words of the query text detection, bigramme duration and content features to represent a query. Some research current studies on identifying characteristics of each type of Queries. Ganti et al. [6] report used tag functionality to query based on the co-occurrence between the different types of tags and query terms.

Wen et al. cluster similar applications according to their content and user logs. They suggested a similarity function based on the application and the content of search results to compare two applications [12]. Chen and Dumais classified search results in predefined hierarchical categories such as Yahoo! directory or Web LookSmart [11]. Cobos et al. introduces a new description-centric algorithm for clustering web results, called WDC-KSB, which is based on the meta-search heuristic algorithm cuckoo, k-means algorithm, balanced Bayesian Information Criterion, split and merge methods on clusters, and common phrases to approach cluster labeling [13].

Then the documents were assigned to the relevant key phrases to form clusters candidates. Wang and Zhai learned one aspect of the application data users query logs with star clustering algorithm. They then categorized and search results organized according to the learned aspects [14]. Beeferman and Berger first built a bipartite graph with the click of data, including user queries and clicked URL. Then, they applied an agglomeration for the graph to the query and the URL [15] relative to the group. First found two interesting phenomena of the user's intention: A search for clarification of sub-theme and keyword. The first means that if a user clicks multiple URLs in a query, and then click the URL tend to represent the same facet. This means that users often add additional keywords to extend the queries to clarify their user's intent to search. Based on

these two phenomena, they grouped all clicked URLs and corresponding queries, where each group represents a plan [1]. Ranked first user's intents of the queries into two types according to their variation on the timeline: constant and sporadic [3].

Then they considered query logs as data flows continuously and divided into variable length partitions. Finally, they grouped each partition into groups of URLs that represent user's intents. [16] Summarized similar queries concepts by combining bipartite click-through queries and URLs recorded from query logs. [17] Random used approach on bipartite graphical URL queries to discover the attributes facet queries. [18] Found requests of user intentions with query logs. For a given query, they identified a first set of possibly linked queries, and then used the March likeness algorithm chance to find clusters of the user's intent. [19] Groups of user queries grouped in the underlying mine of the user's intent. They modeled the behavior of the users in the form of a Markov diagram combining occurrences of co-occurrence of documents, clicks and session co-occurrences, and then they made several random hikes on the graph to get clusters. Clustering/Classifier Terms related to the query. The existing work belonging to this management considers the intention of a request as a set of sub-user applications, namely the terms related to the request [20]. These candidate candidates can come from many sources such as search engine query suggestions, search queries related to log user queries, and so on. Currently, this area is one of the hot topics in the mining sector for application [21]. Proposed an algorithm called dual C-Means to group search results in double representation spaces with the query logs [22]. First found similar applications as candidates for a given query to query logs. Then they used a bipartite graph clicks to narrow these similar applications [23].

They provided taxonomy of key words of intent from the rigorous manual analysis queries. Recently, this problem has been highlighted by many researchers the task consists of two phases: the operation of the user's intent of the ranking. Participants of the job offer numerous methods [24]. Presented a method that gets the best performance in the Chinese data. Specifically, they ranked first user's intents into two types: the explicit role the subject and the subject of implicit role. For subjects of explicit role, they built a graphic modifier based on all the co-kernel-object chains [25]. Then the graph of change has been divided into clusters with high intra-cluster interaction and relatively low inter-cluster interaction.

Each modifier group intuitively reveals a possible user's intent. For subjects of implicit role that usually express single information should they used directly extracted sub-user's intents that user's intents.

In summary, although there is an increase in research investigating the user's intent queries recently, there are still some problems to solve. First, most of the current similarity metrics for keyword are typically constructed from a single viewpoint, namely either from queries or only document collections only. In addition, the combination of different similarity functions from several resources is usually defined heuristically. This can not accurately estimate the similarity between the keyword due to their short text characteristics. Second, existing approaches consider mining query intent and classification of static point of view. They ignore the question of user's intent derives some new user's intents could emerge and old user's intent could become unpopular. Furthermore, issues of diversity and redundancy are not carefully considered in the ranking of Keyword about the coverage of user's intent.

3 Proposed Work

Example. Suppose that user asks about:

When someone tape chocolate into the query box on a search engine page (such as Google), we have about 200.000.000 result and we can found easy what we want (Fig. 1).



Fig. 1. Example result box on a search engine page

In this case "Chocolate", user used only one word.

Into IR system the answer on this question is all information's existing in database about "chocolate".

But our approach is what you see when you go to a search engine - it is the end of what everyone thinks like a search engine.

So, when you type in your search terms and hit search, and the search engine will try to match your words with best pages most relevant web it can find a web search.

You can just enter a keyword unipare (as we have said, if you're researching chocolate).

In seconds, you'll see up to 20,000 keywords containing the word "chocolate" appear in your list. There are several things you can do with this list, but let's start by talking we make any list, sorted by popularity, or search volume, competition, the pages have been directly optimized for each keyword and KEI - effectiveness index keywords.

Search the Number of Queries: This is the number of times each keyword has been searched in the database.

Competition: The competition number gives us an idea of how many web pages already exist that has been optimized for each keyword.

Keyword competition is the measure of how difficult it will be to rank for a particular keyword. The competition for a keyword can vary depending on how popular the keyword is.

KEI (Keyword Effectiveness Index): is one of the quickest ways to find keywords with good potential - that is, those keywords which are likely to help your site attract more traffic and shoes a good result for your query.

Often, you'll find thousands of relevant keywords, you can't necessarily target all of them, we can see our process in Fig. 2.



Fig. 2. Process detect user intent by Search Engine Optimization

3.1 Calculate Our Keyword Efficiency Index (KEI)

The Keyword Efficiency Index is the Efficiency between the number of monthly queries (the value of the keyword) and the volume sites using the keyword in question (competition). Logically a good coefficient is a large number of applications for moderate competition. More commonly used the term called KEI for "Keyword Efficiency Index". Clearly, the higher the KEI of your keywords is high, the better! Now we will explain how to calculate the KEI factor for the keyword set of user.

To calculate the efficiency ratio of your keywords, prepare your spreadsheet with your keywords stored in a list in the first column.

Here's what your keywords are dependent:

• The number of Queries (Nr): This is the monthly volume of searches for this keyword on a given search engine.

As we see, we can find the number of monthly queries simply by visiting for example if we use Google Adwords, it is best to create a Google account (free) to use it without having to enter a captcha code for each research.

Go to the "Keyword Generator" section of the site, then simply enter your keywords in the "Word or phrase" and click search, the results are displayed below.

For example, we see here that "chocolate" was sought on average 2 245 564 times a month in France and 254 854 625 in the world. After obtaining these figures for all your keywords, you can simply click on "import" to obtain your entire table in spreadsheet format, then upload the results on your Excel chart in a "number of Queries" column.

• **Competition (C):** This number indicates how many web sites are more or less already set to your keyword, the lower the sites already positioned on a keyword, the higher the KEI of this word will be favourable in proposition to the number of monthly queries.

To find the number of competitors on a keyword, it is very simple and you probably do already know, indeed just search the keyword on google.fr and the number of results for keyword sought appears we can see in Fig. 3.

 $KEI = \frac{(Number of searches)^2}{Competition}$

Fig. 3. Calculate the Keyword Efficiency Index (KEI)

And we can see our new proposition for Keyword Efficiency to have list of keywords for user intent because logically if we want have a good coefficient is a large number of applications for moderate competition that we can propose our new coefficient:

$$Pi = \frac{(Number of searches * Number of searches)^{1/2}}{Competition} / 2$$

3.2 Our Proposition

It's easy to filter out keywords with high competition (high competition is bad) or low search volume (bad again, a low search volume means the keyword gets little search traffic).

We proposed new proposition for Keyword Efficiency to have list of keyword for user intent a k-means clustering, improved approach where the factor k is not insisted on the outside and produces different size groups. One of the disadvantages of k technical means is to specify k, before the combination. The final result of the aggregation depends on the k. K-means uses the square error method. To minimize the squared error, it splits large groups, even if they are well trained. It is not appropriate. We believe that clustering is a technique depends on field. The final clusters produced in a scenario combining the results of research on the web, depend on the nature of the application and documents. So, k technical means in its basic form is not appropriate in this context.

We determine the value of k using a heuristic search and calculate the effectiveness of keywords Index on the result set. For a better understanding, we will examine the pages and documents that nodes and links between them as edges. The heuristic we apply said; find a node with a high number of links. If a node has many links to other nodes, means that there are many edges between nodes. Clusters in a graphic can be identified by the large number of edges within and less number of edges between them.

After the application of heuristic we get K centers based hyperlinks. In our proposed algorithm to improve the understanding we use the concept of agents who are assigned to each search results page and then calculate the new effectiveness of keywords Index (KEI) to see a good keyword for user's intent. The function of each page of the agent is to maintain a list of related pages.

The documents represented by $p_1, p_2, ..., p_n$ where di \in D and D is the result set consisting of N documents. Let C_k represent the k centers of gravity. At the end of the heuristic phase provides C_k different centers of gravity $C_1, C_2, ..., C_k$ ie we can see in first step for our algorithm in Fig. 4,

$$\bigcup_{i=1}^{k} Cj = \{p1, p2, p3, \dots, p_n\} and Ci \cap C_i = \emptyset, i \neq l.$$

$$L(pi) = \bigcup pj, for all \ 1 \le j \le n \text{ and } j \neq i$$
(1)



Fig. 4. Search algorithm on the web grouping result on the heuristic research

In the list maintained by each page, we try to find a promising link page, many pages of all search results. To find promising p in a list page, we turn to the page that has the largest list and after calculating the keyword index efficiency is effective between the number of monthly queries (the value of the word- Word) and the volume of sites with the keyword in question (competition) as we can see in second step in Fig. 5.

$$pi(di) = pj : argmax \parallel \mathbf{L}(di) \parallel for all \ 1 \le j \le n \text{ and } j \ne i$$
$$Pi = \frac{(\mathrm{Nr} \ * \mathrm{Nr})^{1/2}}{\mathbf{C}}/2$$
(2)



Fig. 5. Calculate algorithm Centroid Construction Keyword Efficiency Index of pages

After obtaining promising page $p_i = p_j$ for di, di list now contains all pages that list already contains. This process is performed so that there is no change in the list of pages. The pages are merged into their respective center of gravity list contains the same elements. At this point, we obtain k different center of gravity, which eliminates the need to specify the factor *k* means method. To go further, all the documents and the centers of gravity are converted into vectors.

A document of the vectors be represented by $d_1, d_2, ..., dn$, where each $di \in D$, and D is the set of search results. The result of all the final consolidation process is k classes. Let $x_1, x_2, ..., x_k$ represent the poles of k such that:

$$\bigcup_{j=1}^{k} x_{j} = \{ d1, d2, d3, \dots, d_{n} \} and X_{j} \cap X_{l} = \emptyset, j \neq 1$$
(3)

The centroid of a cluster XJ can be calculated as

$$Cj = \sum di \in xj \, di / \parallel \sum d\mathbf{i} \in xj \, d\mathbf{i} \parallel$$
(4)

To map centroid vectors relevant documents, cosine similarity is used. We can see Cj is the cluster center of gravity x. The similarity between the vectors can be easily interpreted by the cosine method. The cosine similarity between the document vector and a centroid vector can be defined as (Fig. 6):

$$Sim = (di, cj) \frac{\mathbf{di}, \mathbf{cj}}{\| \mathbf{di} \| \| \mathbf{cj} \|}$$
(5)

The similarity is between 0 and 1. The value of 0 represents no similarity while the similarity value showed complete similarity. This measure is a policy measure, not size. A document vector di is compared to all the centroid vectors. di is assigned to this vector centroid di which has the greatest similarity.

$$xj = \{dj : y = : arg \max di c_l\} for all l = 1, 2, ..., k and 1 \le j \le k$$
 (6)

When non-vector document is assigned a UN center of gravity of the cluster, the center of gravity Must Be updated.

$$Cj = \frac{\sum di \in xj \, di}{\|\sum di \in xj \, di \|} \text{ and } 1 \le j \le k$$
(7)

1.Step3: allocation of document vectors to k centroids
2.For each pi ∈ P
3.Vectorize pi:
4.di ← pi
5.For each document vector di, $1 \le i \le n$
6.Compute di Cm for m= 1,2,, k
7.Assign di to Cm if
8.Cm \leftarrow di {y= argmax (di Cm)} and Y \ge a
9.Else create a centroid
10.Ck+1 = di {y= argmin (di Cm)}
11.Re-compute the centroids
$12.Cm = \frac{\sum di}{\sum di}$ for all di $\in Cm$
Σdi

Fig. 6. Search algorithm allocation of document vectors to k centroids

Documents that have little resemblance or centroid similarity below a given threshold, raise the problem of aberrant values. Since we group the result of the search, a document can not be dropped or rejected because of its similarity with fewer existing clusters. It can be important for the user. To deal with the document vector that has no similarity value greater than or equal to the threshold value, we create a new partition and hold the document as a new center of gravity. If there are more assignable vectors, we stand as a singleton cluster containing a single document.

$$C_{k+1} = \{dj : y = arg \min di c_l\} \text{ for all } l = 1, 2, \dots, k \text{ and } 1 \le j \le k$$
 (8)

The information behind the queries that the user's intent in a Web search engine is very useful for tasks such as in a Web search engine, can improve its performance.

Basically the intent of a user query can be classified sections for several categories. There are several approaches to represent the queries: some of them using the information obtained from query log and other add information from different sources. In this work, we analysis of the history of research is the examination detailed the Web data of different users for the purposes of understanding and optimizing web handling. Than we propose a novel approach to detect user's intent by Search Engine Optimization to improve website ranking in organic search results for increase visibility and quality. We use only the features extracted from the text contained in the queries We tested the query representation with algorithms.

We tested the query representation with Search algorithm on the web grouping result on the heuristic research base Keyword Efficiency Index and k-means, for after make application to have obtained favorable results for the classification of information and transactional queries, but low scores for those with navigation.

We examined the distribution of functions and make application for test our proposition. We can conclude that the use of the good keywords in queries is the important to classify all the user's intents.

4 Experimentation

We observe the same thing in Fig. 7 for the comparison of our algorithm KEI and the KEI with grouping. We found that the variation of the number of similarity and the grouping has a significant effect on the quality of the prediction. Indeed, the MAE decreases according to the number of neighbourhoods of items, the lower its value, the



Fig. 7. Comparison of the accuracy of our SEO algorithm with and without clustering

less the error is important. It is clear that the performance of the proposed algorithm improves with the increase of the number of similar items, which leads to a good precision in the classification of the key words to have the intention of the user, our comparison in Fig. 7.

5 Conclusion

Firstly, we propose a formal framework for searching user's intent in search retrieval by optimization keywords. Secondly, we propose an advanced SEO application. Our approach is tested using many knowledge bases in order to have interesting results. With our advanced SEO the user can ask system in different domain and with different language.

In future work, we plant to developed our approach advanced SEO based communication unit, also we will take increasingly to adopt different strategies according to the category of the keywords. Then we will need to specify the keyword categories more precisely, to have a good result list and improve the extraction of the keywords.

References

- Hu, Y., Qian, Y., Li, H., Jiang, D., Pei, J., Zheng, Q.: Mining query subtopics from search log data. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 305–314. ACM (2012)
- Sakai, T., Dou, Z., Yamamoto, T., Liu, Y., Zhang, M., Kato, M.P., Song, R., Iwata, M.: Summary of the ntcir-10 intent-2 task: subtopic mining and search result diversification. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 761–764. ACM (2013)
- Qian, Y., Sakai, T., Ye, J., Zheng, Q., Li, C.: Dynamic query intent mining from a search log stream. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, pp. 1205–1208. ACM (2013)
- Dou, Z., Hu, S., Chen, K., Song, R., Wen, J.R.: Multi-dimensional search result diversification. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM 2011, pp. 475–484. ACM (2011a)
- Cao, H., Hao Hu, D., Shen, D., Jiang, D., Sun, J.-T., Chen, E., Yang, Q.: Context-aware query classification. In: The 32nd Annual ACM SIGIR Conference, pp. 3–10 (2009)
- Ganti, V., König, A.C., Li, X.: Precomputing search features for fast an accurate query classification. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 61–70. ACM (2010)
- Jansen, B.J., Booth, D.: Classifying web queries by topic and user intent. In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4285–4290 (2010)
- Jansen, B.J., Booth, D.L., Spink, A.: Determining the informational, navigational, and transactional intent of Web queries. J. Inf. Process. Manag. Int. J. Arch. 44(3), 1251–1266 (2008)
- Jansen, B.J., Booth, D.L., Spink, A.: Determining the user intent of web search engine queries. In: Proceedings of the 16th International Conference on World Wide Web, pp. 1149–1150. ACM (2007)

- Wu, D., Zhang, Y., Zhao, S., Liu, T.: Identification of web query intent based on query text and web knowledge. In: First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA), pp. 128–131 (2010)
- Chen, H., Dumais, S.: Bringing order to the web: automatically categorizing search results. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, SIGCHI 2000, pp. 145–152. ACM (2000)
- Wen, J., Nie, J., Zhang, H.: Clustering user queries of a search engine. In: Proceedings of the 10th International Conference on World Wide Web, WWW 2001, pp. 162–168. ACM (2001)
- Cobos, C., Muñoz-Collazos, H., Urbano-Muñoz, R., Mendoza, M., León, E., Herrera-Viedma, E.: Clustering of web search results based on the cuckoo search algorithm and balanced Bayesian information criterion. Inf. Sci. 281, 248–264 (2014)
- Wang, X., Zhai, C.: Learn from web search logs to organize search results. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information retrieval, pp. 87–94. ACM (2007)
- Beeferman, D., Berger, A.: Agglomerative clustering of a search engine query log. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 407–416. ACM (2000)
- Cao, H., Jiang, D., Pei, J., He, Q., Liao, Z., Chen, E., Li, H.: Context-aware query suggestion by mining click-through and session data. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 875–883. ACM (2008)
- Fujita, S., Machinaga, K., Dupret, G.: Click-graph modeling for facet attribute estimation of web search queries. In: Adaptivity, Personalization and Fusion of Heterogeneous Information, RIAO 2010, pp. 190–197 (2010)
- Radlinski, F., Szummer, M., Craswell, N.: Inferring query intent from reformulations and clicks. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 1171–1172. ACM (2010)
- Sadikov, E., Madhavan, J., Wang, L., Halevy, A.: Clustering query refinements by user intent. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 841–850. ACM (2010)
- 20. Xue, Y., Chen, F., Zhu, T., Wang, C., Li, Z., Liu, Y., Zhang, M., Jin, Y., Ma, S.: Thuir at ntcir-9 intent task. In: NTCIR-9 Workshop Meeting, pp. 123–128 (2011)
- Aiello, L.M., Donato, D., Ozertem, U., Menczer, F.: Behavior-driven clustering of queries into topics. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 1373–1382. ACM (2011)
- 22. Moreno, J.G., Dias, G., Cleuziou, G.: Query log driven web search results clustering. In: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 777–786. ACM (2014)
- Dang, V., Xue, X., Croft, W.B.: Inferring query aspects from reformulations using clustering. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 2117–2120. ACM (2011)
- 24. Wang, C., Lai, J., Suen, C.Y., Zhu, J.: Multi-exemplar affinity propagation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(9), 2223–2237 (2013)
- 25. Wang, Q., Qian, Y., Song, R., Dou, Z., Zhang, F., Sakai, T., et al.: Mining subtopics from text fragments for a web query. Inf. Retrieval **16**(4), 484–503 (2013)

Hybrid HMM/MLP Models for Recognizing Unconstrained Cursive Arabic Handwritten Text

Mouhcine Rabi^{1(\Box)}, Mustapha Amrouch¹, and Zouhir Mahani²

 ¹ Laboratory IRF-SIC, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco mouhcineh@gmail.com, m.amrouch@uiz.ac.ma
 ² High School of Technology, Ibn Zohr University, Agadir, Morocco zouhir.mahani@uiz.ac.ma

Abstract. Recognizing unconstrained cursive Arabic handwritten text is a very challenging task the use of hybrid classification to take advantage of the strong modeling of Hidden Markov Models (HMM) and the large capacity of discrimination related to Multilayer Perceptron (MLP) is a very important component in recognition systems. The proposed work reports an effective method on improvement our previous work that takes into consideration the context of character by applying an embedded training based HMMs this HMM is enhanced by an Artificial neural network that are incorporated into the process of classification to estimate the emission probabilities. The experiments are done on the same benchmark IFN/ENIT database of our previous work to compare the results and show the effectiveness of hybrid classifier for enhancing the recognition rate the results are promising and encouraging.

Keywords: Arabic handwriting recognition \cdot Context \cdot Embedded training \cdot HMMs \cdot Multilayer Perceptron (MLP)

1 Introduction

Systems for handwriting recognition are referred to as off-line or on-line systems depending on whether ordinary handwriting on paper is scanned and digitized or a special stylus and a pressure-sensitive tablet are used. In both the ultimate objective is to convert handwritten sentences or phrases words or characters in analogue form (off-line or on-line sources) into digital form (ASCII).

More than 300 million people around the world speak Arabic and their derivative. Arabic is naturally written cursively in both handwritten and typewritten modes. In comparison to Latin Arabic seem to be more complex. For example many letters in this language have complementary diacritics such as dots madda and zigzag bars. In addition the letters have different shapes at different locations of the word.

Due to variability in handwriting styles and distortions caused by the digitizing process even the best handwritten word recognizer is unreliable when the number of word choices is large. This necessitates the use of advanced concepts to achieve a performance level comparable to that of humans. The researches focus on the use of

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_39

new methods and approaches to perform handwriting recognition. In this area the concept of combining multiple classifiers is proposed as a new direction for the development of highly reliable handwriting recognition systems and some preliminary results have indicated that the combination of several complementary classifiers will improve the performance of individual classifiers [1, 2].

In current work our system performs training and recognition of words and characters. In order to model the variations related to the character context in the corpus we have opted for a specific type of learning. Therefore character models are trained and obtained through embedded training; thereafter the decision has been done by the proposed hybrid classifier which mainly based on HMMs and neural network.

The remainder of this paper is organized as follow. Section 2 presents a literature review and related works of handwriting recognition system. Section 3 is focused on our contribution starting with our developed reference system then the incorporation of hybrid classifier. The performance of the recognition system has been experimented on the benchmark database IFN/ENIT and the obtained experimental results are shown and analysed in Sect. 4. The paper finally closed with a conclusion and perspectives.

2 Literature Review

Hidden Markov Models (HMMs) have proven to be one of the most successful and widely used classifiers in the area of text recognition. There are many reasons for success of HMMs in text recognition including avoidance of the need to explicitly segment the text into recognition units; characters or graphemes In addition HMMs have sound mathematical and theoretical foundation [3].

[4] The authors have investigated on contextual sub-characters HMMs for text recognition by using multi-stream HMMs where the features calculated from a sliding window frame form one stream and its derivative features are part of the second stream. The experiments were conducted with different train-test configurations on the IFN/ENIT database and the best recognition rate achieved was 85.12%. In [5] Azeem used an effective technique for the recognition of offline Arabic handwritten words using Hidden Markov Models. Besides the vertical sliding window two slanted sliding windows are used to extract the features. Three different HMMs are used: one for the vertical sliding window and two for slanted windows a fusion scheme is used to combine the three HMMs. [6] proposed a combined scheme for Arabic handwritten word recognition using a HMM classifier followed by re-ranking. Basically intensity features are used to train the HMM and topological features are used for re-ranking to improve accuracy. Experiments were carried out using IFN/ENIT database and the achieved recognition rate was 83.55%. An alternative approach is proposed in [7, 8] where multi-stream HMM models is used this paradigm provides an interesting framework for the integration of multiple source of information. Significant experiments have been carried out on two public available database the recognition rate was 79.8% in IFN/ENIT for Arabic and 89.8% in IRONFF for Latin script. In [8] Maggor proposed a system of Arabic handwriting recognition based on combining methods of decision fusion approach. The combination of the multiple HMMs classifiers was applied by using the different methods of decision fusion approach. The system is

evaluated using the IFN/ENIT database. Experimental results demonstrate that the Weighted Majority Voting (WMV) combination method have given better recognition rate 76.54% with Gaussian distribution. In [9] the authors present an analytical approach of an offline handwritten Arabic text recognition system. It is based on the Hidden Markov Models (HMM) Toolkit (HTK) without explicit segmentation. The feature extraction uses a sliding window on the line text image and processed by two groups of these features (the features of local densities and the statistical characteristics). The proposed system has been experimented in two different databases: "Arabic-Numbers" database where we achieved a rate of 80.26% for words and 37.93% for sentences and IFN/ENIT database where we achieved a rate of 78.95% for words.

Otherwise [10] proposed a new approach for the offline Arabic handwritten word recognition based on the Dynamic Hierarchical Bayesian Network (DHBN) using a free segmentation released by a smoothed vertical projection histogram with different width values. The model is consisting of three levels. The first level represents the layer of the hidden node which models the character class. The second layer models a frame set representing the sub-characters and the third layer models the observation nodes. The developed system has been experimented and the results are provided on a subset of the IFN/ENIT benchmark data base. These results show a significant improvement in the recognition rate because of the use of the DHBN. Most of the recognition errors of the proposed system can be attributed to the segmentation process error and to the poor quality of some data samples.

In many other works neural networks in its different applications have been extensively applied to classify characters as part of isolated or continuous handwritten word recognizers [11–15].

This paper focus on the impact of using a embedded training based on a hybrid classifier the motivation for the work on the hybrid HMMs and Artificial neural network models presented here originates from a critical analysis of the state of the art in offline handwritten text recognition [16–18] our previous work on offline handwriting recognition using HMMs [19] researches and experiences in using hybrid HMM/ANN models for automatic speech recognition [20–24] and for online handwriting recognition [25]. All these criteria making hybrid modeling an important factor in order to achieve an effective and efficient system.

3 Contribution

3.1 Reference System [19]

Our reference system (Fig. 1) was analytical without explicit segmentation based HMMs using embedded training to perform and enhance the character models.

Extraction features was preceded by baseline estimation; the approach used to estimate these baseline based on the horizontal projection curve that is computed with respect to the horizontal pixel density knowing that the skew and slant correction of words are made in pre-processing step to harmonize the direction of the sliding windows in the extraction features. These latter are statistical and geometric to integrate



Fig. 1. Synopsis of reference system



Fig. 2. Embedded training of character "chin" [19]

both the peculiarities of the text and the pixel distribution characteristics in the word image.

The sliding windows are shifted in the direction of writing (right to left). In each window we extract a set of 28 features represent the distribution features based on foreground pixels densities and concavity features. Each window is divided into a fixed number n of cells. Some of these features are extracted from specific areas of the image delimited by the word baselines.

These features are modelled using hidden Markov models and trained by the embedded training method (Fig. 2).

We used a model for each character right-left topology with four states and three transitions for each state. Word model is built by concatenating the appropriate character models.

The embedded training is to automatically identify relevant information letters without specifying them explicitly by exploiting the redundancy of information between words matched to changes in context and letters position.

The major problem of HMMs is the estimation of emission probabilities; this confirms that HMMs are powerful to model sequences but still limited compared to NN and SVM in classification [26] for this reason and the motivations cited above in Sect. 2 it is prominent and promising to use a hybrid classifier.

For more explanations about the baseline system refer to [19].

3.2 Hybrid Classifier

To improve the performance of the off-line handwritten recognition system either the accuracy of the classifier has to be increased. In this section we introduce the hybrid approach then we clarify the principle for our offline Arabic handwriting recognition system.

While HMMs are effective in modeling variation in handwriting they lack discrimination ability because of maximum-likelihood parameter estimation criteria. The strength of MLP is in the fact that they don't need to assume about statistical distribution of input as well as they can be trained to exhibit discriminant properties. As already mentioned recent works in various area of research tried to develop hybrid HMM/MLP systems in which MLPs are used to compute the emission probability associated with each state of HMM.

By HMM the goal of handwriting recognition is to retrieve the most likely grapheme character or word sequence W given a sequence of observation vectors x which is achieved by maximizing the a posteriori probability:

$$\hat{w} = \arg\max P(w|X) \tag{1}$$

Typically each grapheme is modelled by a right to left HMM and the number of states is chosen globally or individually for each character. Gaussian mixtures are used to model the output distributions in each state q given the feature vector x P(x|q). The Baum-Welch algorithm is used for training the HMMs whereas the Viterbi algorithm is used for recognition.

Hybrid models for handwriting recognition based HMMs were built with different neural networks. [27, 28] use an MLP. [29] built an hybrid CNN/HMM. In [30] CNNs is applied in the hybrid framework for handwritten word recognition using different segmentation methods.

In Hybrid HMM/MLP classifier neural networks can be considered statistical classifiers under certain conditions by supplying output of a posteriori probabilities. Thus it is interesting to combine the respective capacities of the HMM and the MLP for a new efficient recognition system inspired by the two formalisms.

The principal idea behind the MLPNN/HMM hybrid approach as illustrated in Fig. 3 is to estimate the output probability density function of each state of the used



Fig. 3. Global scheme design of the hybrid model HMM/MLP

HMM by the output nodes of the MLP classifier which received features as input. These input vectors are pre-processed to finally estimate the posteriori probability deciding whether the input vector belongs to the desired character class. The MLPNN's output weighted by the priori probability of each class forms the probability density function used for every state of the HMM.

In the hidden Markov modeling approach the emission probability density P(x|q) must be estimated for each state q of the Markov chains that is the probability of the observed feature vector x given the hypothesized state q of the model.

In the proposed hybrid HMM/ANN approach the emission probabilities are provided with a neural network since ANNs can be trained to estimate probabilities that are related to these emission probabilities. In particular an MLP can be trained to approximate the a posteriori probabilities of states P(q|x) if each MLP output unit is associated with a specific state of the model and if it is trained as a classifier. The a posteriori probability estimates from the MLP outputs P(q|x) can be converted to emission probabilities P(x|q) by applying Bayes rule:

$$P(x|q) = \frac{P(q|x)P(x)}{P(q)}$$
(2)

The class priors P(q) can be estimated from the relative frequencies of each state from the information produced by a forced Viterbi alignment of the training data. Thus the scaled likelihoods P(x|q)/P(q) can be used as emission probabilities in the proposed system since during recognition the scaling factor P(x) is a constant for all classes. This allows MLPs to be integrated into hybrid structural connectionist models via a statistical framework.

The advantages of this approach are the discriminate training criterion (all MLP parameters are updated in response to every input feature vector) and the fact that it is no longer necessary to assume an a priori distribution of the data. Furthermore if left and right contexts are used at the input of the MLP important contextual information can be incorporated into the probability estimation process. Another strength of this approach is that computing emission probabilities with hybrid HMM/ANN models is usually faster than conventional HMMs with Gaussian emissions since it only requires a forward pass of the MLP for all states of the Markov chains.

4 Experimental Results

To evaluate the performance of our recognition system experiments are conducted using IFN/ENIT [31] database of handwritten Arabic words. It was produced by the Institute for Communications Technology at the Technical University of Braunschweig (IFN) and the "Ecole Nationale d'ingénieur de Tunis (ENIT) "National school of engineering Tunis." This database is used by more than 110 research groups in about 35 countries [32].

4.1 HMM vs HMM/MLP Classification

Reference recognition HMM experiments were conducted using continuous density HMMs with diagonal covariance matrices of Gaussians in each state, a right-to-left topology was applied with four states for each character and three transitions for each state. The optical models were trained and tested using the HTK toolkit [33].

The developed system used hybrid classifier based HMMs enhanced by an artificial neural network that are incorporated into the process of classification to estimate the emission probabilities. Table 1 shows the experimental results of our developed system compared to reference system using the same benchmarking database IFN/ENIT to illustrate the reliability of our improvement models.

We compare the reference system [19] with the developed system using hybrid HMM/MLP classifier. Results in Table 1 show an improvement due to embedded training using the hydride classification: accuracy is increased by 1.1%.

System	Models	$RR^* \%$			
Reference	HMM	87.93			
Hybrid	HMM/MLP	89.03			
RR: Recognition Rate					

Table 1. Recognition results of improvement system compared to reference system

Table 2.	Recognition	results	of	various	systems
----------	-------------	---------	----	---------	---------

System	Models	$RR^* \%$
Irfan [4]	Contextual sub character	85.12
Alkhateeb [6]	HMM + re-ranking	83.55
Kessentini [7]	Multi-stream HMM	79.80
Maqqor [8]	Multiple classifier	76.54
ElMoubtahij [9]	HMM	78.95
Khaoula [10]	DBN	82.00
Our system	HMM/MLP	89.03

4.2 Comparison with Other Systems

A comparison of the recognition rates of our system with other state-of-the-art systems evaluated on the IFN/ENIT database is presented.

Table 2 shows the results of recognition rates for various offline systems recognition of cursive Arabic handwritten text using divers type of models and the same database with the same configuration (training-testing); sets a, b and c for training and d for test, to compare rates and infer the effectiveness of the proposed method.

As it can be noted most of the previous systems are based on HMM using various techniques exploiting the contextual approach, multiple HMM classifier, re-ranking or other techniques to improve HMM models and the recognition rate for the results of the systems mentioned does not exceed 86%. Others used hybrid classifier such as HMM and Dynamic Bayesian Network and the recognition rate achieved was 82.00%. Whereas the proposed system using HMM/MLP outperforms the results and achieve 89.03% due to enhancement of HMMs by incorporating an artificial neural network into the process of classification to estimate the emission probabilities.. This illustrates the effectiveness of embedded training to take account the context of characters to perform the models and using hybrid HMM/MLP classifier to improve the performance of recognition system.

5 Conclusion and Perspectives

In this paper we have enhanced the HMM based reference system by using a hybrid HMM/MLP classifier. Extracted features are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image. These features are modelled using hidden Markov models. These models as already

mentioned in [19] take into account the context of character by applying an embedded training to perform the models. In addition, the contribution in this paper is the improving of HMM modeling by incorporating MLPs to estimate emission probabilities that present the major HMM problem in order to take advantage of the strength of HMM modeling and neural networks classification. The modelling proposed has improved recognition and shown encouraging results to be perfect using annexes improvements.

Due to variability in handwriting styles even the best handwritten word recognizer is unreliable when the number of word choices is large. This forced the use of linguistic constraints to enhance HMMs modeling by a statistical language model that are incorporated as a post-processing into the process of recognition.

Statistical Language Modeling involves attempts to capture regularities of natural language in order to improve the performance of various natural language applications, e.g., Speech recognition, Machine translation, Handwriting recognition, Information retrieval and other applications.

The goal of Statistical Language Modeling is to build a statistical language model that can estimate the distribution of natural language as accurate as possible, which could improve significantly the results especially when we extend our system for line and paragraph recognition.

References

- Alkhateeb, J.H., Pauplin, O., Ren, J., Jiang, J.: Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition. In: Knowledge-Based Systems, vol. 24, pp. 680–688 (2011)
- 2. España-Boquera, S., Castro-Bleda, M.J., Gorbe-Moya, J., Zamora-Martinez, F.: Improving offline handwritten text recognition with hybrid HMM/ANN models (2011)
- Plots, T., Fink, G.A.: Markov models for offline handwriting recognition: a survey. Int. J. Anal. Recongit. 12(4), 269–298 (2009)
- 4. Ahmad, I., Fink, G.A., Mahmoud, S.A.: Improvement in sub-character HMM model based Arabic text recognition. In: Proceedings of the 14th International Conference on Frontiers in Handwrting Recognition (2014)
- Azeem, S., Ahmed, H.: Effective technique for the recognition of offline Arabic handwritten words using hidden Markov models. Int. J. Doc. Anal. Recognit. 16(4), 399–412 (2013)
- AlKhateeb, J.H., Ren, J., Jiang, J., Al-Muhtaseb, H.: Offline handwritten arabic cursive text recognition using hidden Markov models and re-ranking. Pattern Recogn. Lett. 32(8), 1081– 1088 (2011)
- Kessentini, Y., Paquet, T., Ben, A.: Hamadou off-line handwritten word recognition using multistream hidden Markov models. Pattern Recogn. Lett. 31, 60–70 (2010)
- Maqqor, A., Halli, A., Satori, K., Tairi, H.: Off-line recognition handwriting combination of multiple classifiers. In: Proceedings of the 3rd International IEEE Colloquium on Information Science and Technology IEEE, CIST 2014, October 2014
- 9. El Moubtahij, H., Halli, A., Khalid, S.: Using features of local densities statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition (2016)
- 10. Jayech, K., Mahjoub, M.A., Ben Amara, N.: Arabic handwritten word recognition based on dynamic Bayesian network (2016)
- 11. van der Zwaag, B.-J.: Handwritten Digit Recognition: A Neural Network Demo (2016)

- 12. Chen, X.: Convolution neural networks for chinese handwriting recognition (2016)
- 13. Tsai, C.: Recognizing handwritten Japanese characters using deep convolutional neural networks (2016)
- 14. Bluche, T.: Deep neural networks for large vocabulary handwritten text recognition (2015)
- 15. Obaid, A.M., El Bakry, H.M., Eldosuky, M.A., Shehab, A.I.: Handwritten text recognition system based on neural network (2016)
- AL-Shatnawi, A.M., AL-Salaimeh, S., AL-Zawaideh, F.H., Omar, K.: Offline Arabic text recognition an overview. World Comput. Sci. Inform. Technol. J. 1(5) 184–192, 2011
- Parvez, M.T., Mahmoud, S.A.: Offline Arabic handwritten text recognition: a survey. ACM Comput. Surv. 45(2), 23–35 (2013)
- Lawgali, A.: A survey on arabic character recognition. Int. J. Signal Process. Image Process. Pattern Recogn. 8(2), 401–426 (2015)
- Rabi, M., Amrouch, M., Mahani, Z., Mammass, D.: Recognition of cursive Arabic handwritten text using embedded training based on HMMs. In: Engineering & MIS (ICEMIS) International Conference. IEEE, September 2016. INSPEC Accession Number: 16467172. doi:10.1109/ICEMIS.2016.7745330
- Ettaouil, M., Lazaar, M., En-Naimani, Z.: A hybrid ANN/HMM models for Arabic speech recognition using optimal codebook. In: Proceedings of the 8th International Conference on Intelligent Systems: Theories and Applications (SITA). IEEE (2013)
- Surwade, S.S.: Speech recognition using HMM/ANN hybrid model. Int. J. Recent Innov. Trends Comput. Commun. 3(6), 4154–4157 (2015). ISSN: 2321-8169
- G-Moral, A.I., S-Urena, U., P-Moreno, C., D-Maria, F.: Data balancing for efficient training of hybrid ANN/HMM automatic speech recognition. IEEE Trans. Audio Speech Lang. Proc. 19(3), 468–481, 2011
- Mohamed, A., Ramachandran Nair, K.N.: HMM/ANN hybrid model for continuous Malayalam speech recognition. In: Selection and/or Peer-Review Under Responsibility of ICCTSD 2012 (International Conference on Communication Technology and System Design). Elsevier Ltd. (2012)
- Trentin, E., Gori, M.: Robust combination of neural networks and hidden Markov models for speech recognition. IEEE Trans. Neural Netw. 14(6), 1519–1531 (2003)
- Tagougui, N., Boubaker, H., Kherallah, M., Alimi, A.M.: A hybrid NN/HMM modeling technique for online Arabic handwriting recognition. Int. J. Comput. Linguist. Res. 4(3), 107–118 (2013)
- Rabi, M., Amrouch, M., Mahani, Z., Mammass, D.: Evaluation of features extraction and classification techniques for offline handwritten Tifinagh recognition. Glob. J. Comput. Sci. Technol. (USA) C Softw. Data Eng. 16(5) (2016). Version 1.0
- Dreuw, P., Doetsch, P., Plahl, C., Ney, H.: Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained Gaussian HMM: a comparison for offline handwriting recognition. In: Proceedings of the 18th IEEE International Conference on Image Processing (ICIP), pp. 3541–3544, September 2011
- Espana-Boquera, S., Castro-Bleda, M.J., Gorbe-Moya, J., Zamora-Martinez, F.: Improving offline handwritten text recognition with hybrid HMM/ANN models. IEEE Trans. Pattern Anal. Mach. Intell. 33(4), 767–779 (2011)
- Guo, Q., Wang, F., Lei, J., Tu, D., Li, G.: Convolutional feature learning and hybrid CNN-HMM for scene number recognition. J. Neurocomput. Arch. 184(C), 78–90 (2016). Elsevier Science Publishers B.V., Amsterdam
- Bluche, T., Ney, H., Kermorvant, C.: Tandem HMM with convolutional neural network for handwritten word recognition. In: Proceedings of the 17th International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 2390–2394. IEEE (2013)

448 M. Rabi et al.

- Pechwitz, M., Maddouri, S.S., Maergner, V., Ellouze, N., Amiri, H.: IFN/ENIT database of handwritten Arabic words. In: CIFED 2002, Hammamet, Tunisia, pp. 129–136 (2002)
- 32. Märgner, V., El Abed, H.: ICDAR 2011 Arabic handwriting recognition competition. In: International Conference on Document Analysis and Recognition, pp. 1444–1448 (2011)
- 33. Young, S., et al.: The HTK Book V3.4. Cambridge University Press, Cambridge (2006)

Reducing Crowding in Hospital Inpatient Unit Using Queuing Theory

Sara Jebbor^(∞), Abdellatif El Afia, and Raddouane Chiheb

Mohammed V University in Rabat, Rabat, Morocco sara.jebbor@gmail.com, {a.elafia,r.chiheb}@um5s.net.ma

Abstract. Nowadays, emergency department encounters several difficulties to provide quality service to patients, especially inpatient unit that faces a big number of patients random arrivals with different ages and acuities. Patients must be examined and treated in a restricted time, while the constraint of this unit which is limited capacity (human and materiel resources) given the big daily load creates high length of stay, long waiting times and then overcrowding. Those factors impact patient satisfaction and service quality. So, our goal is patients' length of stay and waiting time reduction by increasing inpatient unit service rate according to care load. In this paper, we present our approach which consists essentially in determining the adequate combinations of human and materiel resources to be attributed to each inpatient unit room, in order to insure and provide the optimal service rate. This approach is performed using queuing theory.

Keywords: Emergency department \cdot Inpatient unit \cdot Crowding \cdot Waiting time \cdot Length of stay \cdot Human and materiel resources \cdot Queuing theory

1 Introduction

Nowadays, supply chains improvement issue is becoming an interesting occupation for the majority of researchers in this field; there are works that have treated energetic supply chain, electronic supply chain like: [1–3] and hospital supply chain like: [4, 5] and also our previous works [6, 7] to improve hospital supply chain management and master their stochastic aspects, either for medical circuit or clinical circuit. In this paper, we propose the improvement of emergency unit clinical circuit.

In the recent years, emergency department especially inpatient unit management and their patients' taking-in-charge have become more difficult and complex [4]. This unit faces lots of problems like unqualified medical staff, few medical equipments and materials, ambulance diversion, inadequate planning and staffing, patients random arrivals... all those factors lead to high length of stay and long waiting times that create overcrowding [5] and therefore decrease service quality. In order to reduce overcrowding, hospital managers have provided several solutions such as: increasing the number of beds, adding a new care space, huge recruitments... [8]. Those solutions generate high cost and investigations without really bringing robust improvements. First of all, inpatient unit system must be well organized and mastered according to patients' needs variation and random arrivals. Indeed, increasing the number of human or material

© Springer International Publishing AG 2018 M. Ezziyyani et al. (eds.), *Advanced Information Technology, Services and Systems*, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_40 resources in inpatient unit is not the right solution allowing this unit improvement all over the coming years, because patients' number is in continuous growth. But, an adequate modeling to increase the inpatient unit service rate or flexible resources planning could be much more efficient. They are providing a robust and flexible capacity management according to care load [9, 10].

The remainder of this work is organized as follows: the first section is dedicated to literature about inpatient unit load/capacity and problems. In the second section, we present our approach argumentation according to literature. In the third section, we describe the approach of reducing inpatient unit crowding. Finally, we conclude with conclusion and perspectives.

2 Literature

2.1 Emergency Department Presentation

Emergency department is a unit of hospital center where several services are provided, we mention: reception, triage, evaluation, stabilization, investigation, hospitalization, treatment... [11]. The emergency department assesses and treats people with serious injuries and those who need emergency treatment by receiving and taking care of patients coming by themselves/with their families or patients who are brought by ambulances/ helicopter [5]. In addition, there are specialized emergency services such as maternity, psychiatry, cardiology... For each of those services there is a dedicated inpatient unit. Indeed, inpatient services are provided in different hospital wards according to several criteria like patient category, acuity, or pathology service such as maternity ward, cardiology, oncology, ophthalmology, nephrology, and psychiatry... In hospitalization cases, medical staff should extend personalized care to inpatients and make available all necessary resources [12]. Moreover, inpatient care units provide overnight stay for patients who are recovering from surgical procedures or have special medical conditions. A patient's length of stay until discharge or transfer to another facility is determined by patient's acuity level and needs. Care delivery services may be offered by physicians, nurses, physician assistants, nurse's assistants, dieticians, physical and occupational therapists, respiratory therapists... [13]. Concerning inpatients treatment, inpatient unit may need some of emergency department services like: radiology (scanners, vascular radio ...), imagery, operating room, doctors, nurses... So, those resources must be available at need in order to not extend the inpatients length of stay. According to [4], emergency department includes the following medical staff: administrative officers, reception nurses, nurses, physicians, general practitioners, doctors, surgeons, radiologists, gastroenterologists, pulmonologists, stretcher bearers, anesthesiologists and nurse anesthetists, pediatricians... In addition, we find several services, we mention mainly: operating rooms, maternity service, imagery, pediatrics, radiology, general nurses ward... [14].

2.2 Inpatient Unit Load and Capacity

Further [15] define a tension situation in inpatient unit and generally in emergency department as an imbalance between care load and care capacity, in which the threshold value is exceeded. First, we conclude inpatient unit care load and care capacity elements in the following figure according to [16, 17] descriptions (Fig. 1):

Care Capacity	Care Load
Rooms, beds, boards, medical equipments, diagnostic equipments, lingerie & meals services Operating rooms, other services if necessary (General wards, Ophthalmology) Laboratories, Radiology service (X-ray, scanners) Medical staff (nurses, physicians, doctors, radiologists, surgeons)	Admitted inpatients Inpatients waiting (in waiting rooms or fast track area) Patients to be admitted after programmed (planned) treatment (surgery, childbirth) Patients brought by ambulances/helicopter (random arrivals) Patients leaving without treatment

Fig. 1. Care load and care capacity elements in inpatient unit.

According to [18], increasing medical requirements, expensive resources, and high uncertainty and variability limit the capacity of many of emergency departments especially inpatient units and lead then to capacity shortage. In the same way [15] mention some factors making care load increasing and impacting patients' arrivals number such as seasonal epidemics and accidents. They mention also some factors influencing the speed of taking-in-charge patients and hence impacting the unit care capacity, for example: nursing staff competence (experience feedback), internal and external transfer capacity (availability of downstream care services).

Further [12] mentions that inpatient unit capacity planning taking in count this unit load is very important and primordial regarding the interaction with different services. Indeed, since capacity shortage impacts patients' admission and discharge processes and especially the scheduling with other connected services, inpatient unit capacity must be well identified and planned by using the adequate tool in order to be flexible with the load variations [19].

2.3 Inpatient Unit Problems

Inpatient unit management has become a very difficult and complex task for hospital managers. In one hand, the number of admitted patients whose hospitalization is included in the inpatient unit planning is in continuous increase. In the other hand this unit must take-in-charge random arrivals that are not planned and must be treated immediately due to their urgent case. This stochastic aspect leads to inpatient unit capacity shortage and therefore impacts the service quality [20]. So, inpatient unit is often overcrowded.

According to [21], overcrowding is the most critical problem in hospitals emergency departments over the world, affecting almost 10_74% of hospitals surveyed. This problem results in patients and emergency healthcare staff dissatisfaction. In addition, crowding often causes long waiting time, delayed treatment, overcharged medical staff, and patients leaving without being examined. Also, crowding can lead to other problems, such as medical errors/confusion and ambulance diversion [22].

Furthermore, [5] affirm that overcrowding in inpatient unit decreases the capability to provide immediate access and care to patients. As consequence, there is lack of resources & medical equipments such as beds, shortage of medical staff such as physicians, nurses, and inefficiencies in care processes like performing only some laboratory and radiology tests for inpatients and not all the required ones, inability to transfer admitted patients to inpatient beds; they stay in waiting rooms or wait in inpatient floors or they receive the necessary care (for urgent case) at fast track area. Moreover, [8] find that lack of beds is the most relevant cause of crowding in inpatient unit and can lead to high acuity for patients waiting for treatment.

Further, patients' high length of stay is one of the serious problems putting inpatient unit in overcrowding. Different factors may be involved in the variation of patients' length of stay; [23] mentions for example complementary examinations that take lot of time for hospitalized patients and especially for old patients who need several complementary examinations due to their complex pathology. Also, length of stay is getting higher with unqualified nurses, physicians or doctors [4]. In addition, when emergency department is overcrowded, inpatient unit practitioners (nurses, physicians ...) find difficulties to access the shared resources and services like general nurse ward, laboratories, radiology (scanners, X-ray...) to provide the tests recommended by doctors to evaluate the inpatients health state [24]. Also, some operations could be delayed due to the crowded operating rooms. Those factors lead to stretch the time the inpatient occupy a bed. Otherwise, they contribute to high length of stay, therefore a long waiting time (for patients waiting a treatment/bed) and then inpatient unit crowding.

In hospital sector, there are various metrics used as quality indicators or crowding markers for emergency department. Length of Stay (LOS) has been recently proposed as a quality indicator and a valuable marker of overall emergency department efficiency and overcrowding [21]. Also, waiting time could be used as performance indicator which formula is concluded from the adequate queuing theory model.

Front of all those problems, emergency department managers try always to find some simple solutions without generating additional costs. Also, in several research works, we find panoply of propositions and interventions in order to reduce length of stay and overcrowding. So, we report [8] who present some interventions for reducing crowding, that still inefficient in some cases, we mention mainly: developing alternative sources of care away from inpatient unit (hospitalization centers, collocation of primary care services within or adjacent to emergency care services), diverting ambulances to other hospitals, multi-skilled human resource staffing (helps to decrease bottlenecks), flexible scheduling for medical staff. We mention also discharge lounges, where patients to be discharged can receive final controls (this reduces beds charge) and early ward rounds of newly admitted patients help to match bed availability with demand. Finally, we mention patients' turnover optimization [25].

Certainly, the proposed interventions by researchers/hospital managers are very helpful and contribute to perform better inpatient services. However, they are not strong enough to improve effectively this unit system. In literature, we find several works developing some effectives solutions such as patients' arrivals prediction, care load smoothing over care capacity... by adopting either analytical models or simulation models like: queuing theory, linear programming, theory of constraints, multi-criteria decision analysis [26], what-if analysis, sensitivity analysis and optimization, discrete event simulation, agent-based simulation [27].

3 Our Approach Argumentation

According to the previous part "Inpatient unit problems", inpatient unit overcrowding reduction is expressed by length of stay and waiting time reduction. For this purpose, we propose improving inpatient unit service rate, without generating additional costs; by determining the adequate resources combination(s) providing an optimal service rate by using queuing theory. Since inpatient unit capacity is mainly defined by human and materiel resources, those are our model variables.

In the literature, we find several works that are interested in human resource or material resource or both determination and planning in order to reduce overcrowding in emergency services/inpatient unit using queuing theory. We report a non exhaustive review of those works: [28] have applied the M/G/c/c (or Erlang loss) queuing model to quantify the relation between the size of a hospital unit, the target occupancy rate, and probability of a refused admission. They have developed a decision support system which can be used to evaluate the current size of nursing units (beds) and prove the efficiency of merging departments. Those authors assume that an arriving patient who finds all beds occupied is refused and leaves the system. However, this assumption doesn't reflect the reality, actually a patient who finds all beds occupied must wait until a bed is available and here where we introduce the waiting queue in queuing theory. In addition, those authors work only on material resource determination which is bed, without integrating or taking into consideration the necessary and corresponding human resource that must be affected to the determined capacity of beds. Moreover, despite of several complexities such as time-varying demand, multiple types of patients, and resource sharing, [4] propose a heuristic iterative approach to determine the minimal hour by hour medical staffing levels in order to meet the government target (the 4-hour). The proposed heuristic algorithm uses queuing models to estimate the workload size.

We mention also that those authors have showed how queuing models equipped with simulation could be used to alleviate the congestion problem of emergency departments by modifying the staffing profiles. However, [4] have assumed having an infinite server networks, which is not realistic and is invalid in hospital sector. Furthermore, [29] have modeled a local emergency department as a single station queuing system and have used the Lagged Stationary Independent Period by Period approach to determine physicians staffing number. Those authors solution implementation has brought a considerable improvement for patients who left without being seen. In addition, [12] have performed queuing analysis to analyze a patient load in both inpatient and outpatient services. Having as purpose realistic resources planning simplification, The proposed queuing models (multi-channel queuing models (e.g. Erlang Loss Model)) constitute a basis model for medical staff size estimation (in outpatient services) and for beds number determination (in inpatient services), which are two vital resources in hospital. We mention also [14] that use queuing theory (M/M/n queuing model) to determine the minimal number of servers (providers) needed for a good service, in order to manage in an optimal way the patients flow and to provide accurate evaluations of the system's performance.

According to our deep reading on the inpatient unit organization, resources determination & planning and improvement, we classify research works, that are interested in emergency department/inpatient unit resources determination and planning, in three categories:

- Works that have considered only human resources; like determination of the minimum number of nurse that could perform an optimal service and their planning according to the service load. We mention for example [28, 30–32]...
- Works that have focused on material resources without taking into account the human resources; like beds number determination in inpatient unit. We mention [4, 14, 24, 29] ...
- Works that have performed the human resources and the material resources calculation and determination in inpatient unit but separately. We mention only [12] that we could find after a long search.

Each of the three works categories represents a number of shortcomings that we have extracted from literature [9, 10, 17, 33] and classified in three points as following:

- The determined human resource number (nurses for example) independently of material resource calculation, can either exceed or not meet the service load (for example beds or rooms' number to be taking in charge). So, in the first case we have a high generated cost and in the second case we have as result human resource shortage, and then a bad service quality.
- Sometimes, while fixing a standard service quality (a standard service rate), hospital managers focus especially on material resource determination (for example beds number). They think that increasing material resource number will reduce the problem of inpatient unit overcrowding and long waiting time. However, the material resource number increase generates sometimes high cost without bringing an important improvement. This is because of the lack of adequate human resource to be attributed to each material resource (beds, scanners...).

• When determining separately the human resource and material resource numbers to meet an identified service quality level, we cause inpatient unit disorganization, especially when some services are overcrowded and we desire attribute more human resource (nurses for example) from other services, but we can't determine exactly the convenient human resource number to be attributed which will serve the overcrowded service without putting their original service in shortage.

So, it is essential to ensure the adequacy and availability of both human resource and material resources through the inpatient unit activity programming and the upstream estimation of care load [34]. Indeed, [35] insist on the convergence of human, material and information resources at the same place and at the same time for inpatient unit performance. In addition, on the same level as the human resources optimization, the optimization and the planning of material resources also contribute to optimize the access and the use of technical-medical rooms. As a result, these two resources combination and planning could better organize examinations according to demand [34].

Hence, in our paper, we fill this gap in the literature; about inpatient unit resources combination determination for an optimal service or a given service quality, by generating combinations of minimal number of human resource and materiel resource given us the optimal service rate.

4 Our Approach to Reduce Inpatient Unit Crowding

In our study, we focus on Moroccan hospitals where we find several inpatient units, for example maternity inpatient unit, pediatric inpatient unit, surgical inpatient unit... all included in inpatient department. Our approach is applied generally on one inpatient unit and can be generated for the others. Inpatient unit includes several rooms providing different (or the same service for some ones) service and each one equipped with a number of beds and medical equipments. We consider a room as a server denoted by S, so in our system (inpatient unit), we have multiple servers $(S_1, S_2 \dots S_m)$. For hospital case, we consider an infinite population source situation; patients arrivals are unrestricted, and could exceed the system's capacity at any time [14]. Patients' arrivals follow a Poisson distribution and service durations follow an exponential distribution [28]. So, we adopt the following queuing model: M/M/m. In addition, according to queuing theory principal, the queue discipline refers to the order in which patients are processed. In emergency department and especially in inpatient unit, patients are served according to their acuity level; the most seriously ill patients, are treated first [12]. Thus, patients are categorized according to their assigned acuity level. In each category, treatment is done on a First-In, First-Out basis, because patients of the same category have the same importance. Also, if a patient with high acuity level comes, he is examined immediately in the place of the one (with lower acuity) who is being treated [24]. So, we have M/M/m queuing model with priority rule and preemption, illustrated by the Fig. 2:



Fig. 2. Illustration of inpatient unit servers and patients' arrivals.

Where we denote by:

- λ : Inpatient unit arrivals rate;
- μ: Inpatient unit service rate;
- m: Number of server (rooms);

 $\lambda_1, \lambda_2 \dots \lambda_m$: arrivals rates to the servers $S_1, S_2 \dots S_m$ respectively;

 $\mu_1, \mu_2 \dots \mu_m$: Service rates of the servers $S_1, S_2 \dots S_m$ respectively;

First of all, in this paper we consider only two main resources: beds (material resource) and nurses (human resource). Our objective is to write μ_i for i = 1...m, in form of combination of those two resources.

Let $x_1, x_2, ..., x_m$ be the number of beds in $S_1, S_2, ..., S_m$ respectively.

Let $y_1, y_2, ..., y_m$ be the number of nurses in $S_1, S_2, ..., S_m$ respectively.

We consider S_i , for $i = \{1, 2, ..., m\}$ as a main service which includes x_i beds and y_i nurses; beds and nurses are considered as servers. S_i beds' provide the same service $(x_i \text{ identical servers})$ and constitute the stage 1, also S_i nurses' provide the same service $(y_i \text{ identical servers})$ and constitute the stage 2. So, we represent the service S_i in the Fig. 3 as two stages with multiple servers system:

Let μ'_i be beds service rate and μ''_i nurses' service rate. μ_i is the service S_i service rate. We have two stages with multiple servers system. λ_i is the arrival rate to the service S_i , or to the stage 1 and also to the stage 2. So, the service mean time (T) in S_i is equal to the sum of the service mean time (T1) in the stage 1 and the service mean time (T2) in the stage 2. We have:

$$T = T1 + T2.$$
(1)



Fig. 3. Two stages with multiple servers service S_i

According to queuing theory, we have:

$$T = \frac{1}{\mu_i}.$$
 (2)

$$T1 = \frac{1}{\mu'_i}.$$
(3)

$$T2 = \frac{1}{\mu_i''}$$
 (4)

So, according to (1), (2), (3) and (4) we have:

$$\frac{1}{\mu_{i}} = \frac{1}{\mu_{i}'} + \frac{1}{\mu_{i}''}.$$
(5)

According to (5), we can write:

$$\mu_{i} = \frac{\mu_{i}'\mu_{i}''}{\mu_{i}' + \mu_{i}''}.$$
(6)

According to queuing theory, we have:

$$\rho_i' = \frac{\lambda_i}{x_i \mu_i'}$$
(7)

$$\rho_i'' = \frac{\lambda_i}{y_i \mu_i''}$$
(8)

 ρ_i' and ρ_i'' are beds servers and nurses servers use rates respectively.

So, according to (7) and (8) we can write:

$$\mu_i' = \frac{\lambda_i}{x_i \rho_i'}.$$
(9)

$$\mu_i'' = \frac{\lambda_i}{y_i \rho_i''}.$$
(10)

As result, according to (6), (9) and (10) we have for $i = \{1, 2, ..., m\}$:

$$\mu_{i} = \frac{\lambda_{i}}{x_{i}\rho_{i}' + y_{i}\rho_{i}''}.$$
(11)

Finally, we obtain the service rate μ_i , $i = \{1, 2..., m\}$ expression in form of combination of beds and nurses numbers for a server S_i .

The following parameters: λ_i , ρ'_i , ρ''_i for $i = \{1, 2..., m\}$ and m will be extracted from the studied hospital inpatient unit database. Then, we will use "MATLAB" software to generate several combinations of beds and nurses numbers from (11) by changing x_i and y_i ($i = \{1, 2..., m\}$) values in order to obtain the minimal number of x_i and y_i providing the optimal service rate μ_i .

5 Conclusion and Perspectives

In this paper, we have studied hospital inpatient unit problems in order to identify the critical key allowing this unit improvement by reducing waiting times, length of stay and therefore overcrowding. So, according to literature and our deep reading of research works, we have proved that providing an adequate combination of human and materiel resources could offer an optimal service rate. For this purpose, we used queuing theory by adopting the queuing model M/M/m with priority rule and preemption. We have focused on two main resources which are beds and nurses. So, we developed several equations considering two main assumptions: in one hand, we assumed that beds are the service S_i servers and we obtained the necessary equations. In the other hand, we assumed that nurses are the service S_i servers to obtain the function F expression. As result, we have written the service rate μ_i (i = {1, 2..., m}) as combination of beds and nurses numbers for the server S_i. In our future work, we are going to simulate the obtained formula (11) in a Moroccan hospital inpatient unit framework, after providing the necessary data such as: λ_i , ρ'_i , ρ''_i and m. Indeed, we will use "MATLAB" software to generate several combinations from the Eq. (11) by changing x_i and y_i (i = {1, 2..., m}) values in order to obtain the minimal number of x_i and y_i providing the optimal service rate μ_i . Then, we will improve the obtained formula (11) by integrating other resources like: medical equipments (e.g. radios), doctors... and considering other parameters such as: time intervals, shared resources, nurses and radios categories.

References

- Mezouar, H., El Afia, A., Chiheb, R.: A new concept of intelligence in the electric power management. In: 2nd IEEE International Conference on Electrical and Information Technologies, Tangier (2016)
- Mezouar, H., El Afia, A., Chiheb, R., Ouzayd, F.: Proposal of a modeling approach and a set of KPI to the drug supply chain within the hospital. In: 3rd IEEE International Conference on Logistics Operations Management, Fez (2016)
- Mezouar, H., EL Afia, A., Chiheb, R., Ouzayd, F.: Toward a process model of Moroccan electric supply chain. In: IEEE International Conference on Electrical and Information Technologies, Marrakech (2015)
- Izady, N., Worthington, D.: Setting staffing requirements for time dependent queuing networks: The case of accident and emergency departments. Eur. J. Oper. Res. 219, 531–540 (2012)
- 5. Paul, J.A., Lin, L.: Models for improving patient throughput and waiting at hospital emergency departments. Am. J. Emerg. Med. **43**, 1119–1126 (2012)
- Jebbor, S., El Afia, A., Chiheb, R., Ouzayd, F.: Comparative analysis of drug supply and inventory management methods literature review. In: 4th IEEE International Colloquium on Information Science and Technology, Tangier (2016)
- Jebbor, S., El Afia, A., Chiheb, R., Ouzayd, F.: Management and control of stochastic drug supply chain by KANBAN and Petri Net. In: 3rd IEEE International Conference on Logistics Operations Management, Fez (2016)
- Boyle, A., Beniuk, K., Higginson, I., Atkinson, P.: Emergency department crowding: time for interventions and policy evaluations. Emerg. Med. Int. 2012, 1–8 (2012)
- Chen, T.L.: Decision support system based on distributed simulation optimization for medical resource allocation in emergency department. Lecture Notes in Computer Science, vol. 8527, pp. 15–24 (2014)
- Ben Bachouch, R., Guinet, A., Hajri-Gabouj, S.: An integer linear model for hospital bed planning. Int. J. Prod. Econ. 140, 833–843 (2012)
- Guide de gestion de l'unité d'urgence, la Direction des communications du ministère de la Santé et des Services sociaux, bibliothèque nationale du Québec. http://www.banq.qc.ca/ dotAsset/ae7e54d4-6379-4fea-88d8-bdac78173e06.pdf
- Mital, K.M.: Queuing analysis for outpatient and inpatient services: a case study. Manag. Decis. 48, 419–439 (2010)
- Department of Veterans Affairs, Office of Construction & Facilities Management, Medical/ surgical inpatient unit & intensive care nursing units. https://www.cfm.va.gov/til/dGuide/ dgInpatientNU.pdf
- 14. Vass, H., Szabo, Z.K.: Application of queuing model to patient flow in emergency department case study. Procedia Econ. Finan. **32**, 479–487 (2015)
- Berquedich, M., Kamach, O., Masmoudi, M., Deshayes, L.: Méthodologie organisationnelle des processus en environnement incertain et perturbé: application au domaine hospitalier. In: The 10th International Conference: Conception et Production Intégrées, CPI 2015, Tangier, pp. 1–7 (2015)
- Vissers, J.M.H.: Patient flow-based allocation of inpatient resources: A case study. Eur. J. Oper. Res. 105, 356–370 (1998)
- Hulshof, P.J.H., Kortbeek, N., Boucherie, R.J., Hans, E.W., Bakker, P.J.M.: Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. Health Syst. 1, 129–175 (2012)

- Bai, J., Fügener, A., Schoenfelder, J., Brunner, J. O.: Operations research in intensive care unit management: a literature review. Health Care Manage. Sci. 1–24 (2016)
- 19. Shi, P., Dai, J.G., Ding, D., Ang, J., Chou, M.C., Jin, X., Sim, J.: Patient Flow from Emergency Department to Inpatient Wards: Empirical Observations from a Singaporean Hospital (2013)
- Tancrez, J.S., Roland, B., Cordier, J.P., Riane, F.: Étude de la perturbation par les urgences du planning opératoire. Logistique Manage. 19, 4–52 (2011)
- 21. Casalino, E., Choquet, Ch., Bernard, J., Debit, A., Doumenc, B., Berthoumieu, A., Wargon, M.: Predictive variables of an emergency department quality and performance indicator: a 1-year prospective, observational, cohort study evaluating hospital and emergency census variables and emergency department time interval measurements. Emerg. Med. J. **30**, 45–638 (2013)
- 22. Zeng, Z., Ma, X., Hu, Y., Li, J., Bryant, D.: A simulation study to improve quality of care in the emergency department of a community hospital. J. Emerg. Nursing **38**, 8–322 (2012)
- 23. Jlassi, J., El Mhamedi, A., Chabchoub, H.: Networks of queues with multiple customer types: application in emergency departments. Int. J. Behav. Healthc. Res. 1, 400–419 (2009)
- Zeltyn, S., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Marmor, Y.N., Mandelbaum, A., Shtub, A., Lauterman, T., Schwartz, D., Moskovitch, K., Tzafrir, S., Basis, F.: Simulation-based models of emergency departments: operational, tactical and strategic staffing. ACM Trans. Model. Comput. Simul. 21, 1–25 (2011)
- Askarian, M., Hesami, S.A., Kharazmi, E., Hatam, N., Haghighinejad, H.A., Danaei, M.: Evaluation of the patients' queue status at emergency department of nemazee hospital and how to decrease it, 2014. Glob. J. Health Sci. 9, 1916–9744 (2017)
- Amaral, T.M., Costa, A.P.C.: Improving decision-making and management of hospital resources: An application of the PROMETHEE II method in an emergency department. Oper. Res. Health Care 3, 1–6 (2014)
- 27. Gül, M., Guneri, A.F.: A comprehensive review of emergency department simulation applications for normal and disaster conditions. Comput. Ind. Eng. **83**, 327–344 (2015)
- de Bruin, A.M., Bekker, R., van Zanten, L., Koole, G.M.: Dimensioning hospital wards using the Erlang loss model. Ann. Oper. Res. 178, 23–43 (2010)
- Green, L.V., Soares, J., Giglio, J.F., Robert, R.A.: Using queuing theory to increase the effectiveness of emergency department provider staffing. Acad. Emerg. Med. 13, 61–68 (2006)
- Beeknoo, N., Jones, R.P.: Achieving economy of scale in critical care, planning information necessary to support the choice of bed numbers. British J. Med. Med. Res. 17, 1–15 (2016)
- Green, L.V., Liu, N.: A study of New York city obstetrics units demonstrates the potential for reducing hospital inpatient capacity. Med. Care Res. Rev. 72, 86–168 (2015)
- 32. Boulton, J., Akhtar, N., Shuaib, A., Bourke, P.: Waiting for a stroke bed: Planning stroke unit capacity using queuing theory. Int. J. Healthc. Manage. 9, 4–10 (2016)
- Lane, D.C., Monefeldt, C., Rosenhead, J.V.: Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. J. Oper. Res. Soc. 51, 518–531 (2000)
- 34. Defachelle, C.: L'organisation des soins en hospitalisation de jour: quelles contraintes pour quels enjeux?. Mémoire de l'Ecole Nationale de la Santé Publique, France (1999)
- 35. Chaabane, S.: Gestion prédictive des blocs opératoires. INSA de Lyon, France (2004)

Hybrid Penguins Search Optimization Algorithm and Genetic Algorithm Solving Traveling Salesman Problem

Ilyass Mzili¹⁽¹⁾, Mohammed Essaid Riffi¹, and Fatiha Benzekri²

¹ Department of Computer Science, Faculty of Sciences, Chouaïb Doukkali University, El Jadida, Morocco dr.mzili.ilyass@gmail.com, saidriffi2@gmail.com ² Department of Mathematics, Faculty of Sciences, Chouaïb Doukkali University, El Jadida, Morocco F.Benzekri@hotmail.fr

Abstract. This paper is to present a hybrid technique of two metaheuristic algorithm Penguins Search optimization Algorithm (PeSOA) and the genetic algorithm (GA) called HPeSOA, which was proposed to solve the combinatorial optimization problem NP-hard Traveling salesman problem. In this algorithm, we improve the population of the solutions by the integration of the genetic operators, namely the crossover and the mutation in the algorithm PeSOA. The experimental results of the application of HPeSOA algorithm on the instances TSPLIB are reported and compared, with the results of Penguins Search optimization Algorithm and the genetic algorithm.

Keywords: Metaheuristic \cdot Combinatorial optimization \cdot Genetic algorithm \cdot Penguins search optimization algorithm \cdot HPeSOA \cdot Traveling salesman problem

1 Introduction

Travelling salesman problem (TSP) [1] is one of typical NP-hard problems in combinatorial optimization problem, which wishes to visit a certain number of cities, beginning and finishing its course in the same city by visiting each other's city one and only once. It wishes to select the best tour, which minimizes the traversed total distance.

The resolution of this problem is very important because of its application in various fields such as combinative data analysis, data-processing wiring, machine sequencing, the routing of the vehicles and scheduling, planning and logistics.

In the 19th century, several researchers used heuristic and metaheuristics methods for this problem: Local Search [2], Simulated Annealing [3], Tabu Search [4], Genetic Algorithm [5, 6], Ant Colony System [7], Particle Swarm Optimization [8], Bee Colony Optimization [9], and Penguins Search Optimization Algorithm [10]. Hybrid methods: A novel hybrid penguins search optimization algorithm [11], Hybrid genetic

[©] Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems,

Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_41

algorithm [12], Particle Swarm Optimization with Simulated Annealing [13] and Simulated Annealing Algorithm with Greedy Search [14].

In this paper, we have added GA operators such as selection, crossover and mutation to the PeSOA algorithm in order to find a new hybrid approach HPeSOA. This approach aims to solve the traveling salesman problem. The computational results show that the proposed HPeSOA algorithm has faster convergence, higher computational precision and is more efficient to solve the TSP.

The organization of this research is as follows: In Sect. 2, presentation of the traveling salesman problem; In Sect. 3, description of metheuristics.; In Sect. 4, the adaptation of Hybrid genetic algorithm and Penguins Search Optimization Algorithm to solve traveling salesman problem; In Sect. 5, the results of tests using TSPLIB instances and finally conclusions are summarized in Sect. 6.

2 The Travelling Salesman Problem

Travelling salesman problem (TSP) is one of the most ancient and widely studied problems in combinatorial optimization. Having N cities to visit wishes to establish a tour which permit it to pass exactly once by each city and to return to its starting point for lower costs, traversing the smallest possible distance.

The distances between the cities are known. It is necessary to find the way, which minimizes the distance from displacement. This problem is modeled in linear programming in integer numbers, associating to a binary variable x_{ij} the value 1 if the city i is immediately visited before the city j otherwise, this variable takes the value 0. The cost of this visitor is noted c_{ij} . We then obtain the following model:

$$\operatorname{Min} \quad \sum_{i,j} c_{ij} x_{ij} \tag{1}$$

The constraints are:

$$\sum_{j} x_{ij=1} \quad \forall i \neq j \tag{2}$$

$$\sum_{i} x_{ij=1} \quad \forall j \neq i \tag{3}$$

$$0 \le x_{ij} \le 1 \quad \forall i, \, \forall j \tag{4}$$

The objective function (1) is expressed as a scalar product between a real vector c and a vector of binary variables x. This model is also composed of constraints (2) and (3), allow to visit all cities exactly one and only once. The third constraint (4) prohibits solutions, which will form sub-rounds; it is generally called removal sub-rounds constraint.

3 Description of Metaheuristics

3.1 The Genetic Algorithm

Genetic algorithm [15], introduced by John Holland, is a search algorithm based on the mechanism of natural selection and genetics.

Genetic algorithm first generates an initial population randomly which consist of a set of individual solutions to the problem. The best individuals of the population are those, which have a better chance to reproduce and to transmit part of their genetic heritage to the next generation. A new population of generation is then created by combining the genes of the parents. We expects that certain individuals of the new generation have the best characteristics from both parents, so they will be better and be a better solution to the problem. The new group (the new generation) is then subjected to the same selection criteria, and roughly generates its own kids. This process is repeated several times, until all the individuals have the same genetic heritage. The members of this last generation, who are usually very different from their ancestors, have genetic information g which corresponds to the best solution to the problem. The basic genetic algorithm comprises two simple operations, which are not more complicated than algebraic operations: Crossover–Mutation.

The Pseudo code of this Genetic algorithm is as follows:

```
Algorithm 1. Pseudo code of Genetic algorithm (GA) .
Initialize population
Initial probabilities of crossover (pc) and mutation
    (pm)
do
    Generate new solution by crossover and mutation
        if PC >rand then
            Crossover;
        end if
        if PM >rand then
            Mutate;
        end if
    Accept the new solution if its fitness increases.
    Select the current best for the next generation.
While (the stop condition is not satisfied)
```

3.2 Penguins Search Optimization Algorithm

Penguins Search Algorithm optimization [16] is a new class of metaheuristics proposed in 2013 by the Gheraibia and Abdelouahab Moussaoui. This algorithm is inspired by the strategy of the hunting of Penguins, which is based on the concept of collaboration between the penguins to profit from their diving. This algorithm is based on the
construction of a population, it is divided into groups, each group starts looking for food in a random position, the oxygen reserve exhausted, the groups return to the earth, and choose the best group that consumed the largest amount of fish, the best individuals in the selected group immigrate to other groups to guide them to find rich positions. This technique is repeated several times in order to find one of the best positions. Each position of penguins in the search space is a solution to the problem correlated. Penguins cooperate to determine the best position (solution) in a specific hole and level according (solution space). The movement of penguins between the groups is carried by this equation:

$$D_{new} = D_{last} + Rand \otimes |X_{best} - X_{id}|$$
(5)

Where rand is a random number of the distribution; the current Solution (D_{last}) , the best local solution (X_{best}) , the last solution (X_{id}) and the new solution (D_{new}) .

The Pseudo code of this Penguins Search Optimization algorithm (PeSOA) is as follows:

```
Algorithm 2. Pseudocode Penguins Search Optimization
    algorithm.
Initialize parameters
Generate random population P
Find the best in population
While (the stop condition is not satisfied) do
     For each i individual of P do
         While (RO2> 0) to
              Adjusts a new position using the Eq5
              Choose the best Position.
            End while
 End for.
   Select the best group Gbest
 If (f (Gbest) <f (Xbest)) then
      Xbest = Gbest
        End If
End while
Return Xbest.
```

3.3 HPeSOA Algorithm

The HPeSOA hybrid method is a combination of two Metaheuristics PeSOA and GA. The objective of this method is to improve PeSOA research techniques, by integrating techniques of the genetic method to find solutions of good quality.

The genetic algorithm operator's crossover and mutation implemented to create a population of childhoods solutions from the current solutions and the best solution in

the previous population. The PeSOA mechanism, presents to improve this population by local search techniques in order to find the best solution.

The pseudo code for this HPeSOA algorithm is as follows:

```
Algorithm 3. Pseudo code of HPeSOA
Initialize parameters
Generate random population P
Sort the P in ascending order
While (stopping critical is
                              satisfied) do
   While (stopping critical GA is satisfied) do
        Generate new solution by crossover and mutation
             If (PC >rand) then
                   Crossover;
                End If
             If (PM >rand) then
                     Mutate;
                 End if
   End wihle
     Xbest=min(P) :
     For each i individual of P do
           While (RO2> 0) to
             Adjusts a new position using the Eq4
             Choose the best Position.
           End while
     End for.
     Gbesti = Min (Gi)
   If (f (Gbesti) <f (Xbest)) then
      Xbest = Gbesti
   End If
End while
Return Xbest.
```

4 Adaptation of HPeSOA to Solve TSP

This section presents the HPeSOA method adaptation to solve a TSP.

The adaptation of HPeSOA consists in redefining algebraic operators of genetic algorithm (crossover - mutation), penguins search optimization algorithm and the HPeSOA algorithm structures and steps.

4.1 The GA Operators (Crossover - Mutation)

Suppose that the two positions randomly selected are 4 and 7, so we get the children q1 and q2, but q1 and q2 are not legitimate. We consider the duplicated cities in q1 as superfluous cities that are 2 and 10 in Fig. 1, and duplicate cities in q2 as being devoid of cities that are 3 and 9. Then exchange for the new law q1' and q2'.



Fig. 1. An example of the GA operators (crossover - mutation).

The mutation operator creates random changes in the order of cities. By randomly selecting a solution and two positions, and then changing the first position city with the second one. In Fig. 1 an example of a mutation of a solution q1' with positions 4 and 8 gives the solution q1".

4.2 The PeSOA Operators (Substraction – Multiplication-Addition)

The new solution X_{new} is calculated by using Eq. (5), which is based on the operator's substraction, multiplication and addition: The substraction operation (–) is an operation between two solutions X_{best} and X_{id} , the result of this operation is a set of permutations Q, which can pass from the first solution to the second.

The multiplication operation (*) is an operation effected between a real $k \in [0; 1]$ and a set of permutations Q. The result Q' is a part of the set Q according to the value of k.

The addition operation (+) is an operation effect between the solution X_{id} and the set of permutations Q', the result is a new solution X_{new} . This operation consists of applying permutations of Q' on X_{id} to obtain a new solution X_{new} (Fig. 2).



Fig. 2. A simple example of the PeSOA operators.

4.3 The Flowchart of the HPeSOA Algorithm

The Flowchart of the HPeSOA algorithm is shown in Fig. 3. The structures of two PeSOA and GA algorithms are present in the HPeSOA method to solve Travelling salesman problem.

• 1st step: Initialization.

The initialization of the parameters.

Initialize a population of penguins with random positions on the dimensions D, a position represents a solution to our problem.

• 2nd step: Fitness.

Calculate the objective function of each solution in the population (the physical condition of each penguin in the population) $\{f(x_1), f(x_2), f(x_3), ..., f(x_M)\}\}$

• 3rd step: Selection.

Select the best solution (X_{best}) and the best group (g_{best}) of the population.

• 4th step: Crossover Operator.

Select solutions with PC probability of the population, and the crossover with the best solution g_{best} , then the best child of each crossover, are selected as a child solution to the second population.



Fig. 3. Flowchart of the HPeSOA Algorithm

• 5th step: Shift Operator (mutation).

Select solutions with PM probability of the population, and then generate solutions by exchanging two randomly selected positions within the solution.

• 6th step: update population/gbest

Update the population by child solutions, which are generated by genetic operators and update the best group g_{best} of this population.

• 7th step: update population using PeSOA

Update of the population, using Eq. (5) of PeSOA algorithm.

• The last step: stopping critical

Check the stop condition if the case leaves the loop and shows the best solution found, otherwise, return to step 2.

5 Experimental Results

We applied the HPeSOA method to the different instances of TSPLIB. This experiment implemented by language C in a PC with an Intel (R) Core (TM) 2 Duo processor 2.00 GHz M370@2.40GHZ 2.40GHZ and 4.00 GB of RAM. Each instance runs 10 times. The parameters of HPeSOA are Presented in the Table 1.

Parameters	Value
Gene	1000
М	80
RO2	10
PC	0.95
PM	0.1

Table 1. Value of HPeSOA parameters

Table 2. Numerical results obtained by HPeSOA applied to some TSP instances of TSPLIB

Instance	Ν	Optimum	Best	Worst	Err(%)	Tps(s)
Att48	48	33522	33522	33522	0	1.56
Berlin52	52	7542	7542	7542	0	0.39
Eil51	51	426	426	427	0.11	4.32
Eil76	76	538	538	538	0	28.28
Eil101	101	629	629	629	0	76.92
KroA100	100	21282	21282	21282	0	94.13
KroB100	100	22141	22141	22199	0.13	106.67
KroC100	100	20749	20749	20749	0	121.76
St70	70	675	675	675	0	18.53
Rd100	100	7910	7910	7910	0	84.59
Pr76	76	108159	108159	108159	0	21.33
Pr144	144	58537	58537	58537	0	159.63
Pr107	107	44303	44303	44303	0	174,82
Pr124	124	59030	59030	59030	0	104.65
Lin105	105	14379	14379	14379	0	28.96

The numerical results of this adaptation appearing in Table 2 presented as follows: The first column contains the names of the instances (Instance). The second column represents the number of cities (N). The third is the best result (Optimum) shown in the TSPLIB. The fourth and fifth columns represent the best (Best) and worst results (Worst) obtained by the HPeSOA method. The sixth is the percentage error (Err) and the last column contain the average time of execution (Tps).

$$Err = \frac{\frac{Best + Worst}{2} - Optimum}{Optimum} \times 100\%$$
(6)

Table 3 shows the comparison between the HPeSOA method, Penguins search optimization algorithm (PeSOA) and genetic algorithm (GA). This comparison is based on three criteria: Best solution value (Best), percentage of error (Err) of optimal solution value and average solution value of 10 runs, and the average time of execution (Tps) in second (Fig. 4).

This figure show that the error percentage of HPeSOA method is negligible with respect to GA and PeSOA (Fig. 5).

Instance		PeSOA [10]			GA [6]			HPeSOA		
Name	Optimum	Best	Err (%)	Tps(s)	Best	Err (%)	Tps(s)	Best	Err (%)	Tps(s)
Berlin52	7542	7542	0	0.42	7542	0.24	5.10	7542	0	0.39
Eil51	426	426	0.11	2.5	426	0.63	4.59	426	0,11	4.32
Eil76	538	538	0	24.42	538	0.87	128.94	538	0	28.28
Eil101	629	629	0.07	209.54	634	1.12	226.42	629	0	76.92
KroA100	21282	21282	0	114.04	22142	4.37	48.75	21282	0	94. 13
KroC100	20749	20749	0	75.9	21122	2.77	123.76	20749	0	121.76
Pr76	108159	108159	0	20.63	108278	1.43	131.10	108159	0	21.33
Pr107	44303	44303	0.09	232.96	I	I	I	44303	0	174,82
Lin105	14379	14379	0	74.38	14741	2.67	185.90	14379	0	28.96

A
HPeSC
with
GA,
PeSOA,
performance
the
Comparing
Э.
Table



Fig. 4. Percentage of PeSOA, GA and HPeSOA methods for different instance.



Fig. 5. The overall execution time of the different instances except Pr107 by methods PeSOA, GA and HPeSOA

This figure show that the HPeSOA method solve a set of instances at a small average time of execution, compared with PeSOA and GA methods.

6 Conclusion

In this paper, we have developed a hybrid algorithm HPeSOA that combines the advantage of the genetic algorithm and Penguins search optimization algorithm. This approach is an improvement of the two algorithms: the genetic algorithm so as not to be trapped in the local optimal and PeSOA algorithm by minimizing the time of search for the best solutions.

The result of HPeSOA shows its effectiveness to resolve some instance of TSP according PeSOA and GA, In the future; we plan to add more parameters to the proposed algorithm to make it more robust and flexible to solve the most difficult combinatorial optimization problems.

References

- 1. Lin, S.: Computer solutions of the traveling salesman problem. Bell Syst. Tech. J. 44(10), 2245–2269 (1965)
- Johnson, D.S.: Local optimization and the traveling salesman problem. In: International Colloquium on Automata, Languages, and Programming, pp. 446–461. Springer, Berlin (1990)
- 3. Bayram, H., Şahin, R.: A new simulated annealing approach for travelling salesman problem. Math. Comput. Appl. 18(3), 313–322 (2013)
- Fiechter, C.N.: A parallel tabu search algorithm for large traveling salesman problems. Discrete Appl. Math. 51(3), 243–267 (1994)
- Chatterjee, S., Carrera, C., Lynch, L.A.: Genetic algorithms and traveling salesman problems. Eur. J. Oper. Res. 93(3), 490–510 (1996)
- 6. Ahmed, Z.H.: Genetic algorithm for the traveling salesman problem using sequential constructive crossover operator. Int. J. Biometrics Bioinf. (IJBB) **3**(6), 96 (2010)
- Dorigo, M., Gambardella, L.M.: Ant colonies for the travelling salesman problem. Biosystems 43(2), 73–81 (1997)
- Wang, K.P., Huang, L., Zhou, C.G., Pang, W.: Particle swarm optimization for traveling salesman problem. In: The 2003 International Conference on Machine Learning and Cybernetics, pp. 1583–1585. IEEE Press, Xi'an(2003)
- Wong, L.P., Low, M.Y.H., Chong, C.S.: A bee colony optimization algorithm for traveling salesman problem. In: 2008 Second Asia International Conference on Modelling & Simulation (AICMS), pp. 818–823. IEEE Press, Kuala Lumpur (2008)
- Mzili, I., Riffi, M.E.: Discrete penguins search optimization algorithm to solve the traveling salesman problem. J. Theor. Appl. Inf. Technol. 72(3), 331–336 (2015)
- Mzili, I., Bouzidi, M., Riffi, M.E.: A novel hybrid penguins search optimization algorithm to solve travelling salesman problem. In: 2015 Third World Conference on Complex Systems (WCCS), pp. 1–5. IEEE Press, Marrakech (2015)
- Yugay, O., Kim, I., Kim, B., Ko, F.I.: Hybrid genetic algorithm for solving traveling salesman problem with sorted population. In: Third International Conference on Convergence and Hybrid Information Technology(ICCIT 2008), pp. 1024–1028. IEEE Press, Busan (2008)
- Fang, L., Chen, P., Liu, S.: Particle swarm optimization with simulated annealing for TSP. In: Proceedings of the 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED 2007), Corfu Island, Greece, pp. 206–210 (2007)
- Geng, X., Chen, Z., Yang, W., Shi, D., Zhao, K.: Solving the traveling salesman problem based on an adaptive simulated annealing algorithm with greedy search. Appl. Soft Comput. 11(4), 3680–3689 (2011)

- 15. Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. MIT Press, Cambridge (1992)
- Gheraibia, Y., Moussaoui, A.: Penguins search optimization algorithm (PeSOA). In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 222–231. Springer, Heidelberg (2013)

The Particularities of the Counter Propagation Neural Network Application in Pattern Recognition Tasks

Khatir El Haimoudi¹⁽⁾, Ikram Issati², and Ali Daanoun²

¹ Laboratory of Science and Advanced Technologies, Faculty Poly-Disciplinary of Larache, University Abdelmalek Essaadi, Tétouan, Morocco helkhatir@gmail.com

² Applied Physics Research Team, Faculty of Sciences and Technologies, University Abdelmalek Essaadi, Tangier, Morocco

Abstract. Currently, pattern recognition is an attractive and popular field of research, and requires the emergence of new tools. Artificial intelligence techniques like neural networks are efficient candidates. Various neural algorithms have been employed to solve recognition tasks. In this paper, we focus on the use of Counter Propagation network "CPN" that, combines two modes of learning. A special attention is given to the linear regularities in learning data, and its impact on the success of the Counter-propagation. This work investigates the abilities of CPN for recognition, and detects the ambiguities encountered in the learning process. Experimental tests approve the influence of this problem on the exactitude of the recognition during its application, and allow suggesting an optimization procedure called the Principal components analysis in order to eliminate the detected problems, and to increase the accuracy of the algorithm.

Keywords: Neural networks · Counter Propagation · Pattern recognition · Linear dependence · Principal components analysis

1 Introduction

Pattern recognition is a challenging topic of scientific research, most targeted by the methods of machine learning. In this work, the recognition is viewed as a classification problem; it aims to search the description of an object from a defined database. Various recognition techniques have been used for many years, such as, Data Clustering, Fuzzy Sets, and Support Vector Machine "SVM", a valuable reference is reported in [3].

In this paper, Artificial Neural networks (ANNs) algorithms are suggested as a useful tool for pattern recognition tasks. They concern a category of systems inspired from the functioning of biological nervous systems. ANNs are basically statistical methods with intelligent features such as learning, self-adapting, and the ability to generalize results. Due to these intelligent characteristics, they arrive to classify an object from a defined data set. Also, neural networks can be used to solve other tasks such as modeling, prediction, and control [4, 7, 9].

Neural network (NN) system consists of a set of elementary units called neurons, connected to each other; according to different connections, several neural architectures

could be derived. This paper focuses on the Counter Propagation Neural Network "CPNN" architecture, which is a popular model widely used in different domains, particularly to solve problems of discrimination (classification, clustering, regrouping). CPNN is characterized by special abilities such as learning, adaptation and the possibility of visualization of Multi parameter objects with a reduced space [11]. For pattern recognition tasks, several neural paradigms can beings used, such as Back-propagation network (BPN), self-organization map (SOM), etc. The choice of CPNN is justified by its intelligentes capabilities, in particular, the ability to combine two popular modes of learning (supervised and unsupervised). The last skill gives the advantage to classify data faster and with a better level of accuracy. From the other side, some ambiguities and relations between input data reduce the performance of the classification tasks, which may limit the use of CPNN in different areas of artificial neural network applications.

In order to develop the efficiency of CPNN, several tests have been suggested to identify the nature of ambiguities and obstacles encountered, due to the complex relations between data. Identification of obstacles allows searching for possible solutions to eliminate them. For that reason, a preliminary theoretical study of the learning process and an investigation has been performed respecting the pattern recognition tasks. On the basis of achieving results, the nature and details of the problem are detected which triggers to suggest an optimization procedure called the principal components analysis (PCA).

The PCA is a data analysis method; aims to reduce the data space by eliminating the linear relations between the components of inputs and keep the most informative elements. Using this technique allows exceeding the encountered problem in CPNN algorithm and enables to improve the learning process of the algorithm.

The rest of the paper is organized as follows, in Sect. 2, the principles of CPNN architecture and learning strategy are reviewed. In Sect. 3, some numerical tests and applications that show the abilities of CPNN for pattern recognition tasks, and the investigation of the nature of data ambiguities and their influence on the learning process will be treated. The description of the optimization procedure, and its advantages are described in Sect. 4. Section 5 shows the capabilities of the improved CPNN for classification tasks. Finally, a conclusion and perspective subjects of research are provided.

2 Architecture and Learning Strategy of Counter Propagation Neural Network

2.1 The Structure of the Counter Propagation Neural Network

The Counter Propagation network is a combination of a portion of the Kohonen network and Grossberg outstar structure [6]. The Kohonen maps (Self Organization Maps (SOM)), is one of the basic types of artificial neural networks, composed of three layers: an input layer receives *m* input signals $X = (x_1, x_2,...,x_m)$, a middle Kohonen layer with *N* processing units, and a final output layer. Its architecture represents a two-dimensional grid of connected neurons, which are multidimensional vectors [6, 13]. SOM falls into the category of unsupervised learning methodology in which the relevant multi-variable algorithms seek clusters in the data, in other words, it allows the investigator to regroup objects together on the basis of their perceived closeness in n dimensional space [9]. The basic architecture of Kohonen network is described in (Fig. 1), details and mathematical expressions can be found in several papers, (see [6]). In general, Kohonen maps are designed for visualization, either to show views of the data which indicate groups. The second architecture is the Network of Grossberg, which is a general competitive model with a supervised learning process.



Fig. 1. The typical model of the Kohonen map.

By combining the Kohonen and Grossberg architectures and learning processes, the Counter Propagation Network is obtained, see (Fig. 2), in this case the Kohonen layers regroup data into clusters, then, the Grossberg layer adjusts the output signal relative *X* to the desired output $Y_d = (y_1, y_2, ..., y_n)$ [7, 8].



Fig. 2. The typical model of the Counter Propagation Network.

2.2 The Learning Strategy of the Counter Propagation Network

During the supervised learning process, the Counter Propagation Network is exposed to a couple of training data, input signals and desired outputs $\{x_b, y_i\}_{i=1,...,m_i}$. The learning strategy or algorithm consists of the application of successive rapprochement method, beginning with the random choice of the disposition of cluster centers. Then, the algorithm gradually improves them to perform the learning data clustering [6]. The learning procedure begins with the normalization of input data and synaptic weights to reduce the learning time [10]. This operation is based on the following algebraic formula:

$$x_i = x_i / \sqrt{\sum_{j=0}^{m-1} x_j^2}$$
(1)

 x_i : ith element of the input signal *X*.

In the next step the algorithm computes the winner neuron of the Kohonen layer on the basis of the above formula:

$$k: \|\mathbf{w}_k - \mathbf{x}\| \le \|\mathbf{w}_o - \mathbf{x}\| \quad \forall o$$
⁽²⁾

When the *X* vector is submitted to the input layer, each of the *N* processing elements of the Kohonen layer receives all of its components. A competition is then held among the units of the network, and the process whose weight is close to *x* wins. Subsequently, once the winner is determined the algorithm performs a correction of synaptic weights w_{ij} between input vector *X* in the Kohonen layers is performed at the base of the following formula [6, 11, 13]:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)h(t) \cdot [x_i - w_{ij}(t)]$$
(3)

Where:

 x_i :The value of the winner neuron *i*. $w_{ij}(t)$ and $w_{ij}(t+1)$:The synaptic weights at t and t + 1 iterations. $\alpha(t)$:Learning rate, this coefficient takes a value between 0 and 1.

$$\alpha(t) = \alpha_0 \frac{1}{\exp(MNI)} \tag{4}$$

And MNI: is the max number of iterations. h(t): Neighborhood function.

$$h(t) = \exp(-\frac{d}{2\delta(t)}) \tag{5}$$

$$\delta(t) = \delta_0 e(\frac{t}{\mu}), \ \mu = \frac{1}{Log_{10}(\delta_0)}$$
 (6)

d: the distance between the winner neuron and *X* vector. δ_0, α_0 : Given parameters, in general between 0 and 1.

After the stabilization of the Kohonen layer, the connection weight adjustment for Grossberg layer starts to learn the desired outputs. It is performed by the following two expressions:

$$y_k = \sum_{i}^{N} w_{ki} z_i \tag{7}$$

$$v_{ij}(t+1) = v_{ij}(t) + \beta (y_d - v_{ij}(t)) z_i$$
(8)

 y_d : and y_k : are respectively, the desired output and the estimated output of the network.

- z_i : ith output signal of the Grossberg layer.
- β : Learning rate.

The stabilization criterion of the Counter Propagation Network is the average error between the calculated outputs and the desired outputs.

3 Investigation of the CPNN Abilities for Classification Tasks and the Detection of Data Ambiguities

In this section, some numerical simulations addressed to show the role of Counter Propagation network in pattern recognition tasks, with special attention to detect the ambiguities encountered during the learning process. This investigation aims to show the ability of CPNN in pattern recognition tasks, and study some of its limitations, which gives a motivation for a further improvement treated in the next section. Consequently, three basic examples were suggested.

3.1 Examples of the Recognition Process by the Counter Propagation Network

Approximation of the XOR function, the counter propagation network can not only deal with the classification problems; also it is able to approximate functions and learn nonlinear input-output relationships. Indeed, given a sufficiently large network, an arbitrarily good approximation can be achieved [6, 9]. However, a basic example of nonlinear functions is the XOR input-output relation. The performed results obtained for this function are described in the first example of (Fig. 6) [3, 10].

Results show that the CPNN was able to compute the exact desired outputs in two iterations and with a small level of error 0.00001.

The recognition of Latin alphabets, in the Latin alphabet recognition test, the data set consists of five patterns that represent the letters A, B, X, H, and I. for representing each character we used a pixel grid with a dimension of 5×7 , (Fig. 3).

Fig. 3. Representation symbol A in a grid value 5×7 .

Later, the grid matrix is represented by a vector structured row by row, which consists of 35 components. The description of symbol vectors, A, X, H, B, and I, in the form of input signals, is shown in (Fig. 4).

A 001000101010001100011111111000110001 X 100010100010000100001000101010001 H 10001100011000111111100011000110001 B 11111100011000111111110001100011011111 E 00100001000010000100001000010000100

Fig. 4. Vectors for the letters.

For each input vector, there is a desired output described by it 8-bit binary, ASCII code, see (Fig. 5).

A	:0	1	0	0	0	0	0	1
Х	:0	1	0	1	1	0	0	0
Η	:0	1	0	0	1	0	0	0
В	:0	1	0	0	0	0	1	0
Ι	:0	1	0	0	1	0	0	1

Fig. 5. Desired output vectors for the letters.

The obtained results are represented in the second example; see (Fig. 6), which shows that, the CPNN also was able to compute the same desired output, with 3 numbers of iterations and a smaller level of error less than 0.00001. In fact, the both previous examples were designed to show the ability of CPNN in recognition tasks. The case of the approximation of XOR function, can be viewed as a special case of classification problem, here the SOM layer gives 4 independent clusters (Winner neurons, see (Fig. 6)) with a sum equal to 100%. This indicates that the SOM layer was able to perform a high level of classification. The same results are obtained in the case of the recognition of Latin alphabets (Fig. 6). In this two examples, input data were completely independent and of a nonlinear nature 0 or 1.

Linear or dependent input data, the Last example, aims to study the behavior of CPNN when exposed to linear input data, contrariwise the cases of the first and the second previous examples. Consequently, we selected a vector of 3 elements, such that there is a high level of correlation between the components of the vector. The dependence of data is performed when the elements are close to each other, or they are linearly dependents (see the third Example, (Fig. 6)).

The results Analysis of the application of the Counter Propagation Network in Example 3 shows that, the information about the initial length of the different vectors is lost, and the ratio between the absolute values of the corresponding components and the computed components less than 0.11. In this case, the CPNN cannot perform accurate results; also it is time-consuming, takes 500 of iterations. On the other, hand the SOM

	First example	Second example	Third example
Inputs data	00 01 10 11	Fig 4	0.304 0.403 0.608
Desired outputs	0 1 1 0	Fig 5	0.3 0.4 0.6
Calculated output	0 1 1 0	Fig 5	0.46 0.40 0.44
Iteration rate	2	3	500
Learning rate	1	1	1
Learning rate	3	3	4
Weightinitialvalues	1	0.4	0.4
The average error	<0.00001	<0.00001	0.1100
Winner neurones	0.6, 60, 66	0,5,60,66,9	0, 2, 0

Fig. 6. Learning rate and iteration number for each example.

layer was unable to classify data into 3 independent clusters. (Figure 6) shows that the Kohonen layer gives the same cluster 0 to two different components. This implies a low level of the classification. Consequently, the linear dependence between inputs prevents the layer of Kohonen well classify input vectors, and even the network cannot find the corresponding output vectors. Therefore, our perspective is to improve and create a new model able to eliminate the ambiguities fixed. We suggest developing intelligent systems based on neural network models with competitive learning. These systems will be characterized by news capabilities to identify and classify correctly the linear or nearly linear objects.

3.2 The Proposed Approach for Improving the Learning Process of the Counter Propagation Network

The previous section aimed to detect the reason of the ambiguities encountered during the learning process of the Counter Propagation Network. From the three basic examples, it was clear that the linear dependence between input data influences the classification task performed by the Kohonen layer. Also, it was noticed that the CPNN gives a high level of accuracy in recognition tasks when it is not exposed to these ambiguities. This investigation gives a good motivation to suggest data analysis methods that enable us to eliminate the linear dependence, and improve the accuracy of CPNN.

Let consider a class of data analysis methods, also known as multidimensional scaling techniques, which produce low-dimensional data. These techniques can extract feature parameters and reduce the similar data points and spatial dimensions, then, enable us to optimize the set of learning data. For that various techniques have been developed, such as Principal Component Analysis, Gram Schmidt, *K* means methods, and more [4]. In this work, the Principal component analysis (PCA) is suggested as an optimization technique able to reduce the dimension of the input data by eliminating the linear dependence. Later allows to accelerate the convergence of the CPNN and make it more accurate in the classification tasks [1, 2, 10]. Details about the application of this method are treated in the next section. The design of the optimization block can be suggested as follow (see Fig. 7):



Fig. 7. Optimization block for the Counter Propagation Network.

4 The Improved Model of the Counter Propagation Network

The aim of this section is to eliminate the ambiguities of the CPNN mentioned in the last section. The proposed algorithm in this section combines two well-known methods: the PCA that will calculate the correlation matrix, and Power iteration algorithm (IP) which will search the Eigenvectors that constitute the rows of the resulting matrix. Consequently, a variant of the classical Principal component analysis method is suggested based on a paper developed by the same author [5]. This technique adds new skills that enable to eliminate the linear regularities between the object's components of the learning data, and reserves only the most informative inputs [12]. By combining two learning algorithms and by adding the PCA as an optimization procedure. The performance of CPNN and learning accuracy in pattern recognition will be increased.

The initial data is stated in the form of a Matrix, whose rows are the individuals and columns are the characters of individuals. Each row designs an input vector for the network.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ & \ddots & \ddots & \\ & \ddots & \ddots & \\ x_{n1} & \dots & x_{np} \end{bmatrix}$$

Where:

- x_{ij} : Value of the *ith* observation for the *jth* individual.
- *n*: number of individuals.
- *p*: number of observations

In the first step, the algorithm calculates the vector of mean point g. This point is the center of the points cloud in a space F.

$$g^{t} = (\overline{x}^{1}, \dots, \overline{x}^{p}) \tag{9}$$

And

$$\bar{x}^{j} = \frac{1}{n} \sum_{i=1}^{n} x_{i}^{j}$$
(10)

Where g^t : is the transposed of g.

Then, the data centered matrix Y is calculated basing on g and X, expressed by:

$$Y = X - 1g^t \tag{11}$$

Where: 1 is the identity matrix.

The term centered signifies that the means of the vectors y^{i} are zero.

The centered data matrix *Y* is used in this step for calculating the variance-covariance matrix *V*, which is written according to the following way of *Y*:

$$V = \frac{1}{n}Y^{t}Y \tag{12}$$

Where Y^{t} : is the transposed of *Y*. Such that:

$$V = \begin{pmatrix} S_1 & . & . & S_{1p} \\ S_{21} & . & . \\ . & . & . \\ S_{p1} & . & . & S_p \end{pmatrix}$$

Where: s_{kl} is the covariance of the variables k and l, and s_k is the variance of the variable k.

The last step aims to develop the correlation matrix *R*, for that the two matrices D_{1/S^2} , and D_{1/S^2} , that are computed based on *V*, such that:

$$D_{1/S} = \frac{1}{Diag(V)} \tag{13}$$

$$D_{1/S^2} = \frac{1}{Diag(V^2)}$$
(14)

Where: $D_{1/s}$, is a diagonal matrix containing $\left(\frac{1}{s_{1,1}}, \dots, \frac{1}{s_{p,p}}\right)$, the diagonal coefficients of V.

The correlation matrix R comprises the coefficients of linear correlation between p variables taken two by two. It values show the level of correlation or dependence between p variables. R is symmetric and of a diagonal equal to 1. The last characteristics facilitate the computation, and the data become easy to manipulate. R, is written according to the following way by V.

$$R = D_{1/S} \cdot V \cdot D_{1/S} \tag{15}$$

Such that:

$$R = \begin{pmatrix} 1 & \dots & r_{1,l} \\ & \ddots & \ddots \\ & \ddots & \ddots \\ & & \ddots & \ddots \\ & & r_{1,l} & \dots & 1 \end{pmatrix}$$

Using the IP algorithm, we calculate the Eigenvectors of the matrix R, to constitute the rows of the resulting matrix C.

$$C = \begin{pmatrix} c_{1,1} & \dots & c_{1,l} \\ & \ddots & \ddots \\ & \ddots & \ddots & \ddots \\ c_{n,1} & \dots & c_{n,l} \end{pmatrix}$$

The figure below presents the proposed algorithm flowchart (See (Fig. 8)). The first 6 steps allow computing the correlation matrix, then the IP algorithm enable to calculate the resulting matrix C [5, 12].

The last one is reduced, and the elements of each row are linearly independent, contrariwise the matrix X. Consequently the Matrix X will be used to train the CPNN. Some numerical results show the positive impact of the proposed methods (O. B see (Fig. 7)) are performed in the next section.



Fig. 8. The proposed algorithm.

5 Experimental Results of the Pattern Recognition Tasks by the Improved CPNN

This section aims to show the positive impact of the improved methods or the suggested optimization block by choosing the basic examples of the investigation, a comparison of the classic method and the proposed method are treated.

Approximation of XOR function, in this example, the CPNN with and without the optimization procedure gives good results based on the same initial parameters. The average of the error between the desired output and the computed outputs is less than 0,00001. Also, the kohonen layer classifies inputs into 4 different clusters for the 4 inputs, and the sum of the probabilities of winner neurons is equal to 100%, which proves the accuracy of the classification task, (see example 1, (Fig. 9)).

	Examp	ole 1	Examp	le 2
	S	М	S	М
Input vectors	0 0 0 1 1 0 1 1	0 0 0 1 1 0 1 1	0.3 0.4 0.4 0.3 0.6 0.8	0.3 0.4 0.4 0.3 0.6 0.8
Desired outputs	0 1 1 0	0 1 1 0	0.3 0.4 0.6	0.3 0.4 0.6
Computed outputs	0 1 1 0	0 1 1 0	0.46 0.40 0.44	0.3 0.4 0.6
Iteration rate	2	2	500	3
Learning rate	1	1	1	1
Weight initial values	1	1	0.4	0.4
The average error	< 0.00001	< 0.00001	0.1100	< 0.00001
Winner neurones	0,6,60,66	0,59,63,9	0,2,0	0,50,6

Fig. 9. Data and obtained results in the two variants of the CPNN (S-Standard and M-improved)

Linear input data, in this example, three inputs (objects) are used, of which two elements are linearly depend: the first and the third (see example 2 (Fig. 9)). For the standard CPNN, the mean error is less than 0.11, and for the enhanced CPNN, the error is less than 0.00001. In addition, the kohonen layer gives 3 independent clusters for the three inputs.

	Tes	st 1
	S	М
Input vectors	Voir Fig. 4	Voir Fig. 4
Iteration rate	80	80
Learning rate	1	1
Learning time	40	40
Weight initial values	-1	-1
The average error	0.00001	0.00001

Fig. 10. Test obtained results.



Fig. 11. Graph representing the capacity of the restoration with respect to the level of deformation (M-improved variant of the CPNN, S-Standard variant of the CPNN).

Recognition of Latin Alphabets, in this test, each alphabet is represented in the form of a two-dimensional grid of 5×7 pixels, where the solid symbols are coded by unique values, whereas those which are not filled are coded by zero. The grid components are transformed into input component vectors, and for each input vector there exists the desired output vector which represents the binary ASCII code corresponding to the character, (see (Figs. 3, 4 and 5)). The purpose of the tests is to compare the improved model of the network with the standard one, in order to reveal the advantages and disadvantages of each method.

In the first (Sect. 3), for the CPNN learning test the input data were used in their normal state without deformation, but subsequently in a second test, we tried to make changes in the values of the components of the vectors, and their positions to evaluate and compare the capacity of the restored data distorted by the two variants of the CPNN. The level of deformation provided to each of the vector ranges from 1 to 35 changes.

The obtained results show that the learning capacity of the two variants of the CPNN, and with the same initial parameters in the object recognition task, made 100% of the

recognized objects (see Fig. 10). However, test results in operation mode (Fig. 11) with a deformation level from 0 to 35 indicate that the use of the improved CPNN variant yielded good results compared to the standard variant, at the level of accuracy of recognition and restoration of objects deformed, especially when the level of deformation does not exceed 12 values. But with higher deformation levels, both CPNN variants lose recognition and restoration capability.

6 Conclusion

This paper focuses on the topic of pattern recognition by the tools of artificial neural networks. In particular, the Counter Propagation neural algorithm, thanks to its intelligent characteristics; the ability to learn, to adapt and to integrate two learning modes (supervised, unsupervised), makes it an elegant candidate for pattern recognition tasks. In another side, this algorithm suffers from several limitations due to the complex relations that can be found between input data, especially for data having certain regularities between their components. This problem is due to the inability of the Kohonen layer to correctly identify the linear patterns belonging group, which prevents the Grossberg layer from producing the right output for this type of pattern. Consequently, we investigated the exact nature of this problem and later, we suggested the Principal component analysis as an optimization method that enables to improve the learning process of the CPNN. Final results show that, the improved method was able to exceed this problem by allowing better classification with a high level of accuracy.

The interest, of this paper, appears in two essential points. Firstly we suggested a more useful architecture for pattern recognition tasks. Secondly, we improved the learning algorithm of Counter Propagation Networks, which adds many advantages in the area of neural networks and their applications. Allows choosing a complex recognition problem as a perspective of this work.

References

- 1. Bolley, C.: Analyse numérique. Ecole d'ingénieur. Nantes, France (2012)
- 2. Christian, R., Yvan, S.A.: Mathmatiques et Technologie, 1st edn. Springer, New York (2009)
- Dutt, V., Chaudhry, V., Khan, I.: Pattern recognition: an overview. Am. J. Intell. Syst. 2, 23– 27 (2012)
- 4. Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S.: Apprentissage Statistique, 2nd edn. Groupe Eyrolles, Paris (2008)
- El Khatir, H., Fakhouri, H., Cherrat, L., Ezziyyani, M.: Towards a new approach to improve the classification accuracy of the kohonen's self-organizing map during learning process. Int. J. Adv. Comput. Sci. Appl. 7(3), 230–236 (2016)
- Hecht-Nielsen, R.: Applications of counter propagation networks. Neural Netw. 1, 131–139 (1988)
- Kohonen, T.: Self-organized formation of topologically correct features maps. Biol. Cybern. 43, 59–69 (1982)
- 8. Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer, Berlin (2001)

- 9. Lek, S., Gugan, J.F.: Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. **120**, 65–73 (1999)
- 10. Ripley, B.D.: Pattern Recognition and Neural Networks. Cambridge University Press, NewYork, Cambridge (1996)
- Ravisankar, P., Ravi, V.: Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP. Knowl. Based Syst. 23, 823–831 (2010)
- 12. Saporta, G.: Probabilites, Analyse des données et Statistique Reliée, 3rd edn. Editions Technip, France (2011)
- 13. Vracko, M.: Kohonen artificial neural network and counter propagation neural network in molecular structure-toxicity studies. Curr. Comput. Aided-Drug Des. 1, 73–78 (2005)

Converting Temporal Relational Database into Temporal Object Relational Database

Soumiya Ain El Hayat^(IZI) and Mohamed Bahaj

LITEN Laboratory, Faculty of Science and Technology, University Hassan 1, Settat, Morocco

 $\verb|soumya.ainelhayat@gmail.com, mohamedbahaj@gmail.com||$

Abstract. This paper presents an approach for migrating existing temporal relational database (TRDB), into temporal object relational database (TORDB). This is done by enhancing a representation of a varying time database's structure, in order to make hidden semantic explicit. In contrast to other studies, our main goal here is to offer a first and better solution to mentioned limits to existing works, in order to provide the efficient and the correct method for the translation from TRDB to TORDB. We are going to take an existing RDB using valid time features as input, enrich its metadata representation, and generate a new valid time data Model (NVTM), which captures the most important characteristics of temporal databases for conversion. From the NVTM, we will develop our TORDB design scheme in order to simplify the implementation of a temporal object. Through this UML profile, we precede to the last step, the creation of temporal object relational tables integrating valid time aspects.

Keywords: Valid time · TORDB · TRDB · SQL: 2011 · Temporal database

1 Introduction

The Time-varying data management system represents a model of the information in the real world. The temporal database has appeared since the 1980s, there were a number of authors produced articles and books about varying-time, but the commercial adoption have been slow [1]. During a long time, developers and researchers started to provide new applications support a data type features dependent on the time. The storage time is one of the most important properties are characterized an attribute in systems. Therefore, the need to retain trace and audit the change made to a data and the ability to plan based on the past or future assumptions are important uses cases for temporal data [2]. This is why the applications must support management of temporal data.

The relational database model has come a long way since the 1970s and it has become the dominant model of database [3], most traditional database applications are based on traditional management systems. Today, in majority of management database system, time of event presents one of the most important rows in its application. Therefore, a need to shift time manipulation from application to RDB is implemented in SQL: 2011 standard, which adds period definition as metadata to tables on the relational database. A period definition is a named table component, identifying a pair of columns that capture start and the period end time [1]. Although the relational

© Springer International Publishing AG 2018

M. Ezziyyani et al. (eds.), Advanced Information Technology, Services and Systems, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4_43

database according to SQL: 2011 have been accepted as a new solution for time varying data management, many problems have been emerged. The weakness of such Temporal RDBMSs in supporting complex data structures, Object types and data persistence required by temporal object relational database. ORDBs with time-varying features has addressed of these problems, which have a relational base and append object concept, to enhance ORDB performance and give the correct description of data, we associate time at rows.

The purpose of this paper is to present a method for migrating RDB including valid time data into Temporal ORDB integrating time-varying features. The method comprises three basic steps. In the first, the method takes the entire temporal relational database and stores it in a temporal structured table that contains several parameters, attributes, class, relationships, cardinalities, integrity constraint and valid time period, in order to enrich and realize the schema translation (NVTM). The NVTM so obtained is converted into a temporal ORDB design, which handles complexes object and data semantic that can be expressed in its metadata in the second step. The third step deals with the transformation of the ORDB design, which holds the necessary for the correct description of Object associated with time, to TORDB tables. The prototype has been developed to demonstrate the migration process.

2 Related Works

As we already stated, there are several works dealing with the temporal data storage and query language. Atay presented a comparison between interval-based attribute and tuple time stamped with bitemporal data models, and evaluated their usability using the same data and same queries for both concepts [4]. According to this comparison, Petkoviç's work examined the performance implication for tuple timestamping and time varying attribute, his test stored data using two different forms, and examined the 36 query on both [5].

This work in [6] introduced a temporal object relational SQL language handling valid time dimension at the attribute level in a temporal environment. Comparison of three different data storage models (OODB, ORDB, and XML) for the parametric temporal data model in order to estimate storage costs are discussed in [7].

The ISO (international organization for standard) and IEC (International Electrotechnical Commission) committee, initiated a project to provide a language extension to support temporal database, is given in [8]. The most important features in SQL: 2011 to create and manipulate relational database including temporal data, which implemented by IBMDB2 is discussed in [1]. Slavimir Vesić presented a temporal concept and focused on temporal features defined in SQL: 2011 in DB2 [9]. Sandro Radovanović evaluated the performance of traditional DBMS (RDBMS) and temporal databases using Oracle 12c DBMS [10].

In summary, these related works covered only the technical part of storage and retrieval of data. We feel that some semantic concepts of temporal data are not considered. Furthermore, these studies don't give solutions that will allow providing a mapping method between different data models, and gain from the offered advantages in temporal ORDB. Our study is based on semantic enrichment techniques that give the possibility to understand the structure and meaning of temporal data, and facilitate the construction of schema translation which is enhanced by additional data semantic. We create our approach by combine several results from the previous works, and apply our enhancement using some semantic concepts.

Our aim goal is to define the rules that facilitate the migration from RDB according to SQL: 2011 standard and including valid time dimension into temporal database based on Temporal ORDB model. Furthermore, this solution proposes a meta-model to define set of stereotype for specifying the new characteristic of the valid time associated with UML class diagrams concepts, in order to simplify the generation of TORDB query. In the creation of the conceptual schema it will be considered into Temporal object relational tables.

3 Semantic Enrichment of Temporal Relational Database

3.1 Definition of the New Valid Time Data Model

The NVTM is a representation of Relational Database contains valid time dimension, which is enriched with semantic data in order to provide a new kind of tables representing the classes extracted from temporal RDB, with the data necessary for the creation of temporal Object Relational Database. This phase provides a data reference model that is designed to allow the exchange the temporal schema and the sharing of information to reuse.

The NVTM is defined in our approach as a set of element: NVTM: = $\{C|C: = \{Cn, ANVTM, RelNVTM, Clas, PK-NVTM, FK-NVTM\}\}$

Where:

- Each class has a name Cn
- ANVTM = means a set of Attributes of class C

ANVTM = $\{a|a: = (NA,TA,L,NL,D)\}$, where NA: attribute name, T: attribute type, L: data length, NL: if the attribute accepts the parameter null (N/NoN), D: Default value.

• RelNVTM: Relations NVTM

each class C has a set of relationships with other classes, where rel is defined in C with another class C' RelNVTM = {rel| rel: = RelType,DirC,Crd}, where each class C has a relationship RelType with other classes, DirC is the name of C' that interacts with C,Crd means cardinalities describing the relationship.

RType offers five types of relationships:

- "Ass": Association
- "Agg": Aggregation
- "comp":Composition
- "inher" and "inherBy": Inheritance
- Clas: classification

Classification divides classes into two different kinds of categories:

- Temporal class (TCls): class contains a varying time period.
- Simple class (SCls): class without temporal data.

Ľ	Clas	ANVTM					ReINVTM			PKNVTM		FKNYTM	
		NA	TA	Г	NL	D	RelType	DirC	Crd	PKn	NB	FKn	BB
Employee	TCLS	EmpNo	Number	25	NoN		Ass	Department	1N	EmpNo		dept	_
		Name	Varchar		NoN		Ass	Works-on	1N	Vt_start	7	D-Start	7
		Birthday	Date		NoN		IhnerBy	Salaried-emp	11	Vt_end	3	D-End	3
		Vt_start	Date		NoN		Agg	Kids	1N				
		Vt_end	Date		NoN								
		dept	Number		NoN								
		D-Start	Date		NoN								
		D-End	date		NoN								
Department	TCLS	Deptno	Number	25	NoN		Ass	employee	11	Deptno	1		
		Deptname	Varchar		NoN					Vl-Start	5		
		Vt-Start	Date		NoN					Vt-End	33		
		Vt-End	date		NoN								
Kids	SCLS	NoK	Number	25	NoN		Agg	employee	11	NoK	1	Numemp	1
		Kname	Varchar	10	NoN								
		Sexe	Varchar		z								
		Numemp	Number		NoN								
Project	TCLS	Numproj	Number	30	NoN		Ass	Works-on	1N	Numproj	1		
		Nameproj	Varchar	255	NoN					Vl-start	2		
		Details	Varchar		z					V1-End	33		
		Vt-start	Date		z								
		Vt-End	Date		z								
Works on	SCLS	Empno	Number		NoN		Ass	Project	11	Empno	1	Empno	1
		Numproj	Number		NoN		Ass	Employee	11	Numproj	2	Numproj	2
Salaried_emp	TCLS	Empno	Number	25	NoN		ihner	employee	11	Empno	1	Empno	1
		Grade	Varchar		NoN								
		Salary	Number		NoN								
		Vt-Start	Date		NoN								
		Vt-end	Date		NoN								

Table 1. Result of the generation of NVTM

491

EmpNo	Name	Birthday	VT-Start	VT-End	dep	t D-Start	D-End		
1	Hajar	10/04/1986	15/02/2007	31/12/9999	1	15/07/1998	831/12/9999	Empno	NumProj
								1	2
2	Amine	24/08/1976	03/05/20043	31/11/2011	2	03/05/2004	431/12/9999	5	1
3	Ahmed	24/08/1980	03/05/20083	31/12/2013	4	21/12/2010	031/12/9999	3	17
4	Ilyas	30/01/1979	20/12/20053	31/12/2010	2	03/05/2004	431/12/9999		

3 Ahmed24/08/198010/12/201329/11/2016 3 20/12/200529/11/2016

5 Imane 31/05/197503/05/200631/12/2007 3 20/12/200529/11/2016

DeptNo	deptname	VT-Start	VT-End				
1	Computer	15/07/1998	31/12/9999	NoK	Kname	sexe	Numemp
	-			10	sarah	W	17
2	Accounting	03/05/2004	31/12/9999	15	Mehdi	М	1
3	After-sales	20/12/2005	29/11/2016				
4	Marketing	21/12/2010	31/12/9999				

	Emp	ono	Grade		salary	Vt-Start	Vt-	End
	1		enginee	r	7000	15/02/2007	31/12	/9999
	1		Manage	r	8000	15/02/2015	31/12	/9999
	3		commerc	ial	5000	03/05/2008	31/12	/2013
Nur	nProj	Na	imeproj		Det	ails	VT_Start	VT_END
	1	Payn ag	nent Man- gement	Crea age	ation of p ment app	ayment man- lication web	15/05/2006	601/01/2007

2 HR Management Integration of a module in an erp source 30/10/201301/04/2014

Fig. 1. The tables representing temporal RDB

- PK-NVTM: each class C has a primary key PK. PK = {P|P: = PKn, NB}, where PKn is the primary key name, and NB is a number of Pk in case of a composite key.
- FK-NVTM: denotes foreign key of Class C, FK = {F|F: = FKn, NB} where: FKn is the Foreign key name, and NB is a number of Fk.

3.2 Generation of the NVTM from TRDB

The NVTM presents the first step of the migration process from TRDB into TORDB. Consider the TRDB Example shown in Fig. 1. That example presents an RDB includes

valid time features. In Table 1, we will generate the NVTM of the TRDB described in the example (Fig. 1).

4 Translating NVTM into TORDB Design Schema

In this work, after obtaining the NVTM, we focus on the use of UML notation for creating the TORDB design scheme, which can facilitates the transition towards the object by a set of rules for transposition, and promotes the description of the complex type and possible navigation paths. The model of navigation introduces the logical links between object of the type ref or nested table in order to Decrease the redundancy of the temporal data.

The main goal behind developing a TORDB design is to simplify the comprehension of essentials information stored in temporal data. The TORDB modeling plays a pivot role between the conceptual schemes and implementation object.

5 Translation of the TORDB Design Schema to an ORDB Query

This section describes the schema definition of the previous prototype (Fig. 2), using the commercial database oracle 12C. Through the studies presented in the previous sections, it can be able to produce temporal ORDB queries for relationships. The temporal and no temporal queries formed as shown in Fig. 3:



Fig. 2. Temporal ORDB design schema

```
Create type t Department as object(
  deptno NUMBER ,
  Namedept VARCHAR(25),
  Vt Start Date,
 Vt End Date)/
 Create table Department of t_Department (CONSTRAINT
dept-PK PRIMARY KEY(Numdept, Vt Start Date, Vt End
Date));
  Create type t kids as object(
 Nok NUMBER.
 Namek varchar(25), sexe varchar(10),
 emp Ref t_employee) /
 Create table kids of t kids (CONSTRAINT kids-PK
PRIMARY KEY(Numdept));
  Create type NT_Emp as object(
  emp REF t_employee,
  Vt Start Date,
  Vt End Date )/
  Create type works on is table of NT Emp;
  Create type t_project as object(
  Numproj NUMBER ,
  Nameproj varchar(30),
 Details VARCHAR(255),
 Vt Start Date,
  Vt End Date, employee works on)/
 Create table project of t project (CONSTRAINT proj-
PK PRIMARY KEY(Numproj, Vt_Start Date, Vt_End
                                              Date),
NESTED TABLE employee STORE AS works on tab;
 Create type NT dept as object(
 Dept REF t_Department,
  Vt Start Date
  Vt End
         Date)/
  Create type dept_emp is table of NT_dept;
  Create type t_employee as object(
 NOEMD NUMBER
  Name VARCHAR(25),
  Birthday DATE,
  Vt Start Date,
  Vt End Date,
  Department dept emp) Not Final /
 Create table employee of t employee (CONSTRAINT emp-
PK PRIMARY KEY(NoEmp, Vt_Start ,Vt_End )), NESTED
TABLE Department STORE AS dept tab;
  Create type NT Salary as object (
 Salary Number,
 Vt_Start Date,
 Vt End Date)/
  Create type Salary is table of NT_Salary ;
  Create type T Salaried emp UNDER t employee(grade
varchar(25), salary Salary_emp ) Final;
  Create table Salaried emp of T Salaried emp UNDER
employee NESTED TABLE salary STORE AS salary_tab ;
```

Fig. 3. TORDB queries

6 Conclusion

In this article, we have proposed the basics phases to convert an RDB based on SQL: 2011 standard into ORDB, which contains valid time features, with a simple and practical method to capture the different relationships between classes, association, aggregation, composition, as well as inheritance. Currently, no approach has provided such a solution to extract data model from RDB including temporal data and implemented by SQL: 2011.

This method is done by creating a NVTM from a TRDB, and we use it as an input enriched with semantic data, in order to provide a TORDB design to capture the characteristics of temporal and non-temporal SQL query.

In the future work, we will present an algorithm for mapping method from TRDB with valid time features into ORDB including temporal data that not requires any human interference.

References

- 1. Kulkarni, K.: Michels, J.E: Temporal features in SQL: 2011. ACM SIGMOD Rec. 41(3), 34–43 (2012)
- Kaufmann, M., Fischer, P.M., May, N., Kossmann, D.: Benchmarking bitemporal database systems: ready for the future or stuck in the past? In: EDBT, pp. 738–749 (2014)
- Brdjanin, D., Maric, S, Pavkovic, Z.S.: On suitability of standard UML notation for relational database schema representation. In: Enterprise Business-Process and Information Systems Modeling, Vol. 248, pp. 399–413. Springer (2016)
- Atay, C.E.: A comparison of attribute and tuple time stamped bitemporal relational data models. In: Proceedings of the International Conference on Applied Computer Science, pp. 479–489 (2010)
- Petković, D.: Performance issues concerning storage of time-variant data. Egypt. Comput. Sci. J. 38(2), 1–11 (2014)
- Chau, V.T.N., Chittayasothorn, S.: A temporal object relational SQL language with attribute timestamping in a temporal transparency environment. Data Knowl. Eng. 67, 331–361 (2008). Elsevier
- Noh, S.Y., Gadia, S.K., Jang, H.: Comparisons of three data storage models in parametric temporal databases. J. Central South Univ. 20, 1919–1927 (2013)
- 8. ISO/IEC 9075-2:2011: Information technology Database languages SQL Part 2: Foundation (SQL/Foundation) (2011)
- Vesić, S., Babarogić, S., Aničić, N.: Use of the temporal concepts in transaction database. In: Proceedings of the SYMORG, pp. 850–857 (2014)
- Radovanović, S., Milovanović, E., Aničić, N.: Performance evaluation of temporal features defined in Oracle 12C database. In: Proceedings of the SYMORG, pp. 858–866 (2014)

Implementing of a Binary Data Generator on a FPGA Card

M. Benzaima^(SC), Mensouri Mohammed, Aaroud Abdessadek, and Ali El Hore

Department of Computer Science, Faculty of Sciences, El Jadida, Morocco mo.benzaima@gmail.com, mensourimohl@hotmail.com, a.aaroud@yahoo.fr, aelhore@gmail.com

Abstract. The technique of binary data sequence generator based on LFSR is used a variety of cryptographic applications and for designing an encoder/decoder in different communication channel. It is more important to test and verify by implementing on any hardware device to get a better effective result. As FPGA is used to implement any logical function for faster prototype development, it is necessary to implement the existing LFSR design on FPGA to test and verify the result of simulation and synthesis between different lengths. The total number of random states generated on LFSR depends on the feedback polynomial. The binary data generator is implemented using an LFSR shift register. This is a 23-bit shift register. Random Number Generator allows you to generate a random number with the selected length. The maximum length is $2^{23}-1$.

Keywords: Binary data generator \cdot LFSR \cdot Implementation VHDL \cdot Quartus \cdot FPGA

1 Introduction

To generate a sequence of data, random numbers are very useful in various applications such as communication channel [1, 2]. It is used to design an encoder and decoder for sending and receiving data in a noisy communication channel. They have also been used aesthetically, for example in literature and music, and are of course always popular for games [1]. When we speak of single numbers, a random number is one that is derived from a set of possible values, each of which is also probable, that is, a uniform distribution. The Random Number Generator is a computing device designed to generate a sequence of numbers. There are various methods for knowing pseudo-random numbers. Most of them are based on linear equations and require a number of long arithmetic operations. On the other hand, the use of feedback shift registers allows a very rapid generation of binary sequences. The maximum length register sequences (m-sequences) are well suited to simulate truly random binary sequences [3–5].

The FPGA configuration is usually defined using a hardware description language (HDL) similar to that used for an application-specific integrated circuit (ASIC) (circuit diagrams were previously used to specify the configuration, as they were for ASICs, but this is increasingly rare). FPGAs can be used to implement any logic function that an ASIC can perform [6].

The design flow of an FPGA allows configuring a chip from a starting description in a predefined order of steps that we can mainly break down into two: design and simulation. Design tasks allow you to go from one description to the other to arrive at the configuration. In fact, a logical synthesis makes it possible to pass from an RTL description of the architecture to a description at the logical gate level. The description of logical elements is optimized according to the speed, surface or consumption constraints imposed by the designer. The synthesis tool replaces the generic logical elements with those specific to the targeted FPGA. Placement and routing convert the hardware description into a configuration file. The synthesis tool generates this file which is used for the configuration of the interconnection matrices of the FPGA circuit (Fig. 1).



Fig. 1. LFSR basic block diagram

We used the environment QUARTUS II version 7.0 of the company Alteras both for the design and for the implantation on chip.

Quartus is software developed by the company Altera, allowing the complete management of an FPGA design flow. This software makes it possible to make a graphical input or a description HDL (VHDL or verilog) of digital architecture, to realize a simulation, a synthesis and a reprogrammable target implementation [10].

2 Line Offset Register

2.1 LFSR Shift Register

LFSR is a shift register whose input bit is a linear function of its previous state. The most commonly used linear function of simple bits is XOR. Thus, an LFSR is most often a shift register whose input bit is controlled by the exclusive or (XOR) of certain bits of the overall value of the shift register [1, 8]. The initial value of the LFSR is called the seed. Since the register has a finite number of possible states, it must eventually enter a repeating cycle. However, an LFSR with a feedback function can produce a sequence of bits that seems random and has a very long cycle. Applications of LFSRs include pseudo-random number generation, pseudo-noise sequences, fast digital counters, and bleach sequences. The hardware and software implementations of LFSR are Common [7].

2.2 Implementation of LFSR-Based PRNG

The pseudo-random number sequence generator is generated in VHDL according to the following circuit as a function of the shift register concept. The bits in the LFSR state

that affect the input are called taps. A maximum-length LFSR produces a sequence (is it traverses all possible 2^n-1 state in the shift register except the state where all bits are zero), unless it contains all the zeros, in which case it will never change. The sequence of numbers generated by this method is random. The period of the sequence is (2n-1), where n is the number of shift registers used in the design. For generating a 23-bit LFSR shift register, the period is 8388607. This is large enough for most practical applications. The arrangement of the valves for feedback in an LFSR can be expressed in arithmetic of the finite fields as a polynomial mod. This means that the coefficients of the polynomial must be 1 or 0. This is called the feedback polynomial or the characteristic polynomial. For example, if the valves are at 23, 18, the feedback polynomial is

$$X23 + X18 = 1$$
,

3 Method for Generating a Binary Data Sequence

The Data Generator is implemented using an LFSR, a shift register whose input bit is a linear function of its previous state. The most commonly used linear function of simple bits is XOR. Thus, an LFSR is most often a shift register whose input bit is driven by the exclusive or (XOR) of certain bits of the overall value of the shift register.

3.1 Design of a 6-Bit Binary Generator on Matlab

For the chosen polynomial generator:

$$P(z) = z6 + z + 1,$$

The model of Fig. 2 generates a periodic binary data sequence of period 63.



Fig. 2. LFSR circuit diagramme 6-bit

The simulation result of a 6-bit binary data generator on matlab is shown in Fig. 3



Fig. 3. Simulation generator 6 bit

3.2 Design of an 8 Bit Data Generator

The 8-bit generator, with return polynomial of maximum length X8 + X6 + X5 + X4 + 1, generates 28 - 1 = 255 random outputs, which is checked from the simulation waveform. The circuit diagram for 8-bit LFSR with maximum length polynomial is shown in Fig. 4 [9].



Fig. 4. Circuit diagram of 8-bit LFSR

3.3 Design of a 16 Bit Data Generator

LFSR 16 bits, with return polynomial of maximum length X16 + X14 + X13 + X11 + 1, generates 216 - 1 = 65535 random outputs, which is checked from the simulation waveform. The circuit diagram of the 16-bit LFSR with maximum length polynomial is shown in Fig. 5.


Fig. 5. Circuit diagram of 16-bit LFSR

4 Architecture of a Binary Data Generator

The binary data generator is realized using a LFSR (Linear Feedback Shift Register). Figure 6 shows the structure of the LFSR that we have implemented. This is a 23-bit shift register. The recursively of the register is defined using the following generator polynomial:



Fig. 6. Binary data generator

Indeed, the signals (x1, x2, and x23) of the register give a value in binary representation. The value of the register at a periodicity which depends on the generator polynomial. So the LFSR does not generate a random data sequence. If this period is sufficiently large, we consider that the generated data are random. The structure of Fig. 6 makes it possible to generate a data sequence with a period of $2^{23}-1$.

5 Implementation of a Binary Data Generator

The binary data generator has input and output ports shown in Fig. 7.

- Clk: Clock signal active in the high state
- **Reset:** Signal to reset the generator
- Active: Signal to activate the generator

Random: Output signal of the bits generated by the generator



Fig. 7. Diagram of inputs and outputs of the binary data generator

6 Simulation of a Binary Data Generator

To achieve simulation results, we test the binary data generator in the Quartus II environment, you can see the input signals and output signals in Fig. 8.



Fig. 8. Simulation of a generator

7 Conclusion

In a digital communication chain, the data is encoded and transmitted on a communication channel; the information source is the first link in the transmission chain. The binary data generator supplies the message carrying the digital information. The binary data generator produces a periodic sequence that appears to be random. The sequence is not statistically random but will pass many random tests. This sequence is called pseudo-random numbers or pseudo-noises.

In the future improvement, we expect to increase the number of tapping, the degree of chance as well as the stroke length can be increased and again one can compare the sequence of maximum length between LFSR of different length with an implementation on FPGA.

In this work and for an improvement purpose, a controller can be added to the binary data generator to divide the frame entering the encoder to add the redundancy bits and transmitted via the transmission channel.

References

- 1. Luby, M.: Pseudo randomness and Cryptographic Applications. Princeton University Press, Princeton (1996)
- 2. Hao, J., Li, Z.: On the production of pseudo random Numbers in Cryptogrzphy. J. Chzngehou Teach. Coll. Technol. **7** (2001)
- 3. L'Ecuyer, P.: Random numbers for simulation. Commun. ACM 33, 10 (1990)
- Katti, R.S. Srinivasan, S.K.: Efficient hardware implementation of a new pseudo-random bit sequence generator. In: IEEE International Symposium on Circuits and Systems, ISCAS 2009 (2009)
- 5. Goresky, M., Klapper, A.M.: Fibonacci and Galois representations of feedback-with-carry shift registers. IEEE Trans. Inf. Theory **48**, 2826–2836 (2002)
- 6. Brown, S., Vranesic, Z.: Fundamental of Digital Logic Design with VHDL, 2nd edn. McGraw Hill
- Panda Amit, K., Rajput, P., Shukla, B.: Design of multi bit LFSR PNRG and performance comparison on FPGA usingVHDL||. Int. J. Adv. Eng. Technol. (IJAET) 3(1), 566–571 (2012)
- 8. Bhasker, J.: A VHDL Primer. P T R Prentice Hall, Englewood Cliffs (2013)
- Hao, J., Li, Z.: FPGA design flow based on a variety of EDA tools. Micro-comput. Inf. 11-2(23), 201–203 (2007)
- 10. http://www.polytech.univmontp2.fr/pravo/cours/Logique/Altera/Notice%20Quartus.pdf

Towards a Hybrid Method of Construction of a Normalized Domain Ontology Used by Machine Teaching PERO2

Mostafa Chahbar, Ali Elhore, and Younes Askane^(IX)

Department of Computer Science, Chouaïb Doukkali University, EL Jadida, Morocco mostafa.chahbarl@gmail.com, elhore.a@ucd.ac.ma, ay.askane@gmail.com

Abstract. The challenges facing learning of reasoning require a general refounding of knowledge modeling expressed by the development and integration of intelligence and reasoning techniques in the processing of this knowledge. In this paper, we present a part of the research work of PERO2 project. This intelligent system is a machine teaching dedicated to the learning of reasoning and problem-solving of exercises in physical science. Our research is focused on representing the knowledge base of PERO2 by integrating a semantic layer based upon a domain ontology, capable of adding some intelligence and enriching the reasoning in our system. In order to build this ontology, we propose a hybrid method based on two main phases: (*) design phase of our domain ontology called "OntoPhyScEx". (**) Semantic validation phase of this ontology.

Keywords: Domain ontology \cdot Normalization \cdot Intelligent tutoring system \cdot Knowledge base \cdot Learning of reasoning \cdot Problem-solving

1 Introduction

Machine teaching is a discipline of Artificial Intelligence which concerns the interaction in appropriate manner of humans and machines during the learning process in order to offer an approach to improve the education and training of personnel. In principle, we can use machine teaching to design the optimal lesson for each and individual student [1]. This coupling between machine and user concerns the design and development of systems endowed with algorithms and techniques of scientific thought allowing reasoning on data similar to that of human beings, and thus to teach the Machine to perform certain difficult tasks to be performed by more conventional algorithmic capabilities. In this context, our research work aims to add and integrate a semantic layer within our machine teaching architecture "PERO" developed by [2], by providing a knowledge base based on ontologies that supply a formal representation of domain knowledge in order to be used by a machine and to be eventually understood by a community of people [3]. Most often, it is a structured set of concepts, actions, and predicates providing a semantic description to knowledge. PERO is a learning environment based on a reference frame consisting of three conceptual learning models and resolution process to solve a problem (exercises):

- Learner model focuses on adapting the learning content to the learner's characteristics in order to allow the learner or student to learn at his or her level. PERO1considers that the knowledge of the learner are organized onto (learner's knowledge) and (know-how).
- Process model concerns the methods and the pedagogical approaches which should focus on knowledge to be offered to the learners, e.g.: for a lesson on "RLC electrical circuits" the concepts of this lesson are related to each other by relationships as (RLC is sequentially composed of others concepts, to teach RLC concept it is necessary to have prerequisites/knowledge as example on: resistance R, coil L and capacitor C).
- Resource model allows the description of the pedagogical nature of a resource. Resource can be: a concept, a law, a lemma, etc. our contribution in this paper concerns this model to be implemented via domain ontology model allowing a semantic description of this one by proposing a hybrid construction method.

The resolution process of a given problem is based upon an explanatory model [4]. The model proposed at this stage is action oriented; it takes into account the context/ intention of actions that are performed by the system to reach a solution. The context/ intention of actions is set of knowledge represented by the goal of the action, the means used to execute the action and the reasons justifying the action. This formalism (intention of an action (Goal, Mean, Reason)) allows the system to manage and to predict the actions of the learner, in the other hand it conducts the learner to construct the solution of a given problem step by step by interpreting these actions in the form of States.

The problem of PERO lies in its data storage tools, it uses a traditional models (relational database MySQL) which has disadvantage of its low availability, particularly if the volume of information circulating is important, as well as the data stored in a database are textual data which remains a major obstacle for the search and extraction of relevant information. The integration of a semantic model of data allows an intrinsically agile semantic processing and an in-depth data analysis capable of endowing PERO2 with capacities comparable to human reasoning. In this focused-research work we will discuss an ontology construction method to represent the knowledge of physical sciences domain especially what concerns pedagogical knowledge defined by PERO's Resource Model. Our choice will be a hybrid method based on the methodology of [5] Which requires a manual modeling during the ontology construction stages. The process of construction requires two phases: (1) a phase of the construction of the initial ontology and (2) a phase that concerns the correction and the semantic validation of the initial ontology.

In the next sections of this document we discuss related works in the utilization of Artificial Intelligence technics especially "ontologies" to design intelligent machines teaching. Next we introduce our hybrid Methodology to create normalized domain ontology for our machine teaching PERO2 but firstly we must present briefly PERO2 architecture, then we present in detail the first phase of initial taxonomy construction and we give an overview of second phase of semantic validation of our initial taxonomy

which is not detailed in this scope. We then address a discussion on perspectives of this research work.

2 Ontologies in the Design of Machine Teaching

Ontologies become a main structural framework in the explicit modeling and organization of information in many areas of artificial intelligence such as Cognitive Science, Semantic Web, Systems Engineering, Business Intelligence, Big data, etc. Being a result of an attempt to formulate exhaustively and rigorously the conceptualization of a domain, an ontology provides through this formulation a formal representation of the domain knowledge in the form of a semantic network called taxonomies of relevant concepts, Relationships between these concepts, rules and axioms that constrain them [6]. In order to produce intelligent learning environment that approximates to human education, an immense effort has been made in recent years described by the use of a discipline of artificial intelligence "ontologies", this axis research has given rise to many approaches based on ontologies for modeling teaching subjects: MONTO approach [7] which offers Machine-Readable Ontology for teaching problem solving in mathematics at the high school level, [8] Have created an ontological representation of an educational scheme to improve the discovery and resource recovery potential of an intelligent elearning system. [9] Presented an integrated intelligent tutoring system in order to support distance learning. [10] Proposed an optimal teaching framework aimed at learners who employ Bayesian models, This framework is expressed as an optimization problem over teaching examples that balance the future loss of the learner and the effort of the teacher. [11] Proposed an architecture supported by several ontologies as a way of addressing the problems. Ontologies have been widely used for student modeling mainly for two reasons: (a) ontologies support the formal representation of abstract concepts and properties in a way that they can be reused by many tasks or extended if needed and (b) they enable the extraction of new knowledge by applying inference mechanisms (e.g. inference engine) on the information presented in the ontology [12].

In this context, our contribution is to increase intelligence in PERO and to explore its potential to an improved version PERO2 that offers a comprehensive approach to learning taking into account aspects of planning (planning a solution to exercises), explanation (generate an explanation related to each step and which leads to the resolution along the exercises resolution process) and finally indexing (index traces indicating the various exercises have been done during solving process). However, we favored adding and integrating a semantic layer that consists of a formal representation of declarative knowledge of type ontology coupled with an inference mechanism to analyze and reason on it. The usefulness of this ontology is to enable an interaction in direct mode when querying the system rather in batch mode (the case of database) while following the inference rules defined therein.

3 Brief Description of PERO2 Architecture

PERO2 machine teaching architecture preserves the same objectives (Resolution, explanation and Indexation) of the previous version as it is shown on the left side of Fig. 1, the change is focused on means of storage and treatment of declarative knowledge manipulated by the system (Fig. 1; right side) that retain their integrity whatever the use of them, which is not the case when knowledge is stored in a relational model (case of PERO). This knowledge base is based on a domain meta-ontology that provides a general structure in terms of concepts and relationships between concepts [13, 14] whose role is to produce ontologies related to the different subjects of teaching physical science in the form of instances (e.g., teaching subject "Electricity" is an instance of the domain meta-ontology), thus expanding the scope of the system to affect several disciplinary of physical science (mechanics, electronics, etc.).



Fig. 1. Simplified system architecture of PERO version 2.

4 Methodology to Create Normalized Ontology for PERO2

The design of ontology raises a number of difficulties. One of these difficulties is the gap between syntax and semantics. This gap leads to divergences between the designer's intention and his modeling: unexpected inferences, absence of expected inferences, even inconsistencies [15]. Therefore, our choice is a hybrid method of ontology construction to represent the domain knowledge of physical science that will be adapted and integrated into our system PERO2. The implementation of such ontology requires a manual modeling based on two phases (Fig. 2):

The phase of the construction of the initial taxonomy taking into account the following steps: (1) requirements specification based on a textual corpus, (2) Conceptualization, (3) Internal Structure of concepts, (4) Define extensional relationships of

concepts, (5) instantiation, while providing an ontological refinements all Throughout these steps.

- The semantic validation phase taking into account (6) normalization of the semantic meaning lent to concepts using the ONTOCLEAN meta-properties [16], and the implementation of this normalization to ensure modularity of ontology (7) and finally, the formalization of the ontology in a formal language in order to ensure relevant criteria during the construction process.



Fig. 2. Phases of construction of our domain ontology "OntoPhyScEx".

4.1 The Extraction of the Initial Taxonomy of Our Ontology of Concepts

This phase is widely used by different methods of ontology designs, for example in the domain of networking computer and communications [17] Proposed a conceptual model and a software framework using ontologies to build semantic context databases. [18] Offers medical ontology for information retrieval. [19] Proposed an ontological approach for creating a metadata profile for Remote Learning.

The modeling of this phase takes into account five steps: (1) specification of the domain that aims to specify the domain of knowledge that the ontology must represent the operational objective and the users of the ontology. (2) Conceptualization of data devoted to the analysis and extraction of relevant concepts. (3) Internal structuring of concepts consists in defining the differential semantics of each concept. (4) Define extensional relations between concepts and finally, (5) instantiation.

(1) Specification of the knowledge domain:

In the first place, it is necessary, for the creation of this ontology, to define its domain of knowledge that it must represent its operational objective, and the users of this ontology. The domain we are trying to model is not physical science in itself, but the exercises of physical science. In turn, it will be necessary to define questions of competence, which are concrete examples of questions to which the ontology must answer, in order to formulate its objective. And finally users of ontologies that are students, teachers and domain experts.

(2) Conceptualization:

The second step consists of defining a specification in advance via interviews with actors in the domain and the analysis of the data resources relating to Corpus¹. of the application domain (the books, series of directed works and websites of physical science). Next we have to enumerate the important terms in the ontology. Indeed, it is useful to note in a list form, all the terms to be treated and to specify their conceptual nature (concepts, relations, properties, theorems, laws, rules, lemmas, constraints, etc.). in relation to the physical science domain we have defined in this list the most salient concepts of the top level domain {Mechanics, Electricity and Magnetism, Light, Vision, Sound, Hearing, Relativity, Astrophysics, Quantum, Nuclear Physics, Condensed Matter, Heat, Thermodynamics}, then we defined hierarchy of concepts: we have a choice to use a "top-down" process, (begins with a definition of the general concepts of the domain or the most important and continues with the specialization of concepts), or "bottom-up" (it begins with the definition of the specific classes and continues with the regrouping of these classes in more General concepts), or we can combine the two processes [20].

The following (Fig. 3) shows on the right an example of the concepts of physical science domain and on the left an informal initial hierarchy of these concepts.



Fig. 3. Hierarchical extraction of concepts of the sub-domain «electricity and magnetism» .

(3) Internal structuring of Concepts:

After defining the concepts, we must describe their internal structure. A concept can have an extension property (includes the objects manipulated through the concept) called Data Properties allow connecting individuals (instances) of concepts to data values (e.g. String, number, Boolean, enumerated) and an intended property (contains the semantics of concept) known as Object Properties allow

¹ It may contain domain-specific knowledge corresponding to the physical and mathematical knowledge that helps solving the exercises of physical science and others participate in the expression of domain knowledge.

liaising instances of concepts to other individuals (e.g. composition, aggregation, association, etc.) [21].

Figure 4 shows an example of internal structure of Resistor's object which describes these characteristics; in most cases, Resistor's object is presented with color rings (bands) around it. Each color corresponds to a digit. The mapping between numbers and colors of the bands is named (Resistor Color Code): This code is used to determine the value and type '4-band, 5-band, 6-band' of "Resistor". Other properties can be specified as 'Tolerance', 'Temperature-Coefficient' and physical size 'Shape'.

(4) Define the extensional relations of concepts: Concept has a referential semantic imposed by its extension [21] provides a connection by reference to other domain's concepts using set theory operations (reunion,

tion by reference to other domain's concepts using set theory operations (reunion, intersection, complementary...), laws concerning relationships (symmetry, reflex-ivity, transitivity, ...) and the laws of logical axioms.



Fig. 4. An example of internal structuring of concept (resistor).

The notion of subsumption (relationship of generalization/Specialization "is-a") is a particular binary relationship which implies the semantic commitment according to [22]: a concept c1 subsumes a concept c2 if the concept c2 is more specific than the concept c1, and the instances relating to the concept c2 will be instances of c1, on the other hand only a part of the instances of c1 will be instances of c2.

Figure 5 shows an exemplary case of electrical circuit, the concept parent "Electric Circuit" has two concepts child: "Direct Current Electric Circuit" and "Alternating Current Electric Circuit", and each of two concepts have some sub-concepts. The set of concepts is structured hierarchically within a network of concepts, and are linked by

conceptual properties of type "is-a" and semantic relations. To simplify the example, the sign ' $^{\prime}$ indicates that the concept is subsumed by other concepts.



Fig. 5. Example of extensional Relations (a specific relation "is-a" and semantic relations).

(5) Instantiation:

Instances also known as (Individuals) correspond to a concrete objects models that constitute a formal part of ontology and describe the interest entities.



Fig. 6. Example of conception of an RLC electrical circuit in alternating regime fitted in series.

Considering the following scenario as concrete example based upon an AC RLC Electrical Circuit fed by a Frequency Generator f = 50 Hz and Amplitude E = 311 V. Phase at the origin of the Voltage e (t) delivered by the Generator is taken equal to zero. with a Resistance (R) equal to 40 Ω , Inductance (L) equal to 0.2H and finally a Capacitance (C) equal to 5 μ F. All electronic components are fitted in series.

Figure 6 shows the conception of our use case of our RLC Electrical Circuit in alternating current in our domain ontology:

As result following these standard steps of producing an initial taxonomy of our domain ontology as a conceptual graph in which the edges correspond to every entities or concepts and the arcs describe the several relationships types such as semantic relationship, generalization/Specialization "is-a" relationship and so on.

Figure 7 below illustrates part of our ontology that represents the sub-domain "Electricity and Magnetism". We suppose that the relations between the concepts are a specific relation "is-a" and we seek to correct these relations through the validation phase.



Fig. 7. Brief example of our ontology "OntoPhyScEx" representing the sub-domain "Electricity and Magnetism".

4.2 Overview on Normalization of Our Domain Ontology

The normalization of ontologies is inspired by the normalization of information models for relational databases. Indeed, semantics in a conceptual model is expressed by: classes, properties, relationships, instances and constraints of a problem domain. [23] Has studied the differences and similarities between domain ontologies and conceptual data models, they have much in common.

The ontology as well as a conceptual model consists of concepts, properties, individuals and constraints (restrictions and axioms). [24–26] Have discussed the normalization that states the adding of constraints to the construction of ontology so that the source ontology meets the five criteria proposed by [5]; (1) Accuracy of domain, (2) reuses (3) modularity, (4) maintainability, (5) scalability. The semantic validation requires passage through two phases:

(1) Normalization of semantic meaning lent to concepts:

Seeking to find an abstract interpretation of the variations of meanings of concepts related to other domains and represented in the ontology: A concept is a meaning, its place in a system of meanings helps to understand it, to distinguish it and to differentiate it from other concepts [27]. The same concept can have different meanings, interpreted by different domain experts. Many concepts have been borrowed by computer science from mathematics and logic, but this borrowing has often resulted in a skewed meaning. In particular, the terms property and class are used in computer science with often drastically different meanings from the original [5].

The aim in the current stage is about to explicit and adjust the meaning given to the concept by associating an independent interpretation of its context of use. In our case, the choice is brought on this principle by using the definition of meta-properties of concepts (identity, rigidity, unity, dependency)². In order to structure and to test the coherence of the hierarchy by imposing some constraints on the use of these meta-properties.

Assigning meta-properties to concepts that requires a combination of these metaproperties (which are not independent). This combination produces eight types of properties that help to structure the taxonomy hierarchy and that are classified into: Sortals: "Type", "Quasi Types", "Mixins", "Material Roles", "Sortals Phased". Non-Sortals: "Attributions", "Formal Roles", "Categories". The projection of the meta-properties on the concepts is described by the assignment beside each concept the meta-properties notations preceded by the sign "+", "-" or "~" corresponding to: carrying the meta-property, not carrying the meta- property, and "anti-metaproperty". These notations are assigned on a simple and natural intentional reasoning.

As example, the concept "Electric Circuit" has combination of (+I+R-D-U) which corresponds to the classification of "Sortals" category of property types "TYPE". Indeed, this concept carries out an Identity Condition (IC); "electric circuits may have same power source type of DC or AC regime", in the other hand

² Evaluating Ontological Decisions with OntoClean (see reference [15]).

the concept doesn't carry unity condition (UC); "electric circuits are not necessarily wholes, so our assignment is ~U", finally the concept carries no dependency at all.

To satisfy the normalization of the meaning attributed to the concepts, we must go through a check of constraints in order to correct the taxonomy. "OntoClean" methodology imposes some constraints [16] on the hierarchy of the taxonomy throughout the assigned meta-properties, to help the designer to infer modeling inconsistencies in the hierarchy.

(2) Implementation of normalization: The previous step of normalization provides an independent explicit analysis of the ontology of any implementation tool. [25] Has proposed a solution for the implementation of this normalization in a formal language. The purpose of this one is to achieve modularity which aims to decompose the ontology taxonomy to a set of hierarchies ("modules") homogeneous disjoint. This hierarchical decomposition must meet specific criteria of normalization [28] related to identify primitive concepts that constitute a backbone taxonomy domain in which we distinguish between (a) "Self-Standing" concepts correspond to all types of concepts to represent the physical and conceptual world (e.g. ideas, processes, human beings live, organizations, etc.). (b) The "Refining" concepts are concepts that represent types of values or quantitative or qualitative values (e.g. small, medium, large, mild, moderate, severe, etc.). The consequences of such decomposition is to support the evolution and the update of the ontology following the requirement changes (e.g. the context of use is changed, or the domain knowledge is expanded) in the ontology. Such changes must lead to updates in a small number of modules.

5 Conclusion

In this work, we have proposed a hybrid method of construction of normalized domain ontology to represent and to exploit the knowledge base of our system for learning of reasoning and problem-solving PERO 2. The main aim is to replace the existing relational database with a base of ontological concepts. The strong point of this solution is that it is based on meta-properties inspired by the philosophical foundations and allows defining concepts and their relationships in a formal, explicit and independent manner of any context. The prospect of this approach is to implement the normalization in a formal language which requires a decomposition of the hierarchical structure of the ontology into a set of disjoint homogeneous "modules" [26]. This modularity consists in supporting the evolution and updating of ontology and then integrating this solution in order to interface the system functioning through Machine learning algorithm with the knowledge base expressed by normalized ontological model.

References

- 1. Zhu, X.: Machine teaching: an inverse problem to machine learning and an approach toward optimal education. In: AAAI 2015 Proceedings of the Twenty Ninth Conference on Artificial Intelligence, Austin, Texas, pp. 4083–4087, 25–30 January 2015
- El-Hore, A., Tazi, S.: PERO a planning system for the explanation of problem-solving in physical sciences. In: Workshop on Mixed Language Explanations in Learning Environments, AIED, Amsterdam, Hollander, 18–22 Juillet 2005. http://andes3.lrdc.pitt.edu/ xlang/. Classée A par CORE
- Gruber, T.R., Olsen, G.R.: An ontology for engineering mathematics. In: KR 1994, pp. 258– 269 (1994)
- El Hore, A., Tazi, S.: Explaining and indexing solutions for physics learning. In: IEEE Conference International CELDA 2005, pp. 532–538, 13–17 December 2005
- Guarino, N., Welty, C.: An overview of OntoClean. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 151–171. Springer (2004)
- Chandrasekaran, B., Josephson, J.R., Benjamins, V.R.: What are ontologies, and why do we need them? IEEE Intell. Syst. Appl. 14(1), 20–26 (1999)
- Lalingkar, A., Ramnathan, C., Ramani, S.: Ontology based smart learning environment for teaching word problems in mathematics. J. Comput. Educ. 1(4), 313–334 (2014)
- Nikolopoulos, G., Kalou, A., Pierrakeas, C., Kameas, A.: Creating a LO metadata profile for distance learning: an ontological approach. In: Research Conference on Metadata and Semantic Research, pp. 37–48. Springer, Heidelberg (2012)
- 9. Panagiotis, S., Panagiotopoulos, I., Goumopoulos, C., Kameas, A.: Combining agents and ontologies for building an intelligent tutoring system. In: CSEDU (1), pp. 15–24 (2015)
- 10. Zhu, X.: Machine teaching for Bayesian learners in the exponential family. In: NIPS (2013)
- Panagiotopoulos, I., Kalou, A., Pierrakeas, C., Kameas, A.: An ontology-based model for student representation in intelligent tutoring systems for distance learning. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 296– 305 (2012)
- da Silva Jacinto, A., de Oliveira, J.M.P.: An ontology-based architecture for intelligent tutoring system. Interdiscip. Stud. Comput. Sci. 19(1), 25–35 (2008)
- Müller, J.P.: Mimosa: using ontologies for modeling and simulation. In: Proceedings of Informatik 2007, Bremeen, Allemagne, Germany, s.l:s.n (5), 25 September 2007
- Zniber, N.: Conception de parcours pédagogique: une approche orientée service. In: La 3ième conférence en Environnement Informatique pour l'Apprentissage Humai – EIAH 2007, 27– 29 Juin 2007
- Ferré, S.: A proposal for extending formal concept analysis to knowledge graphs. In: International Conference on Formal Concept Analysis, pp. 271–286. Springer International Publishing (2015)
- Guarino, N., Welty, C.: Evaluating ontological decisions with OntoClean. Commun. ACM 45(2), 61–65 (2002). http://doi.org/10.1145/503124.503150
- Loukill, M., Ghariani, T., Jouaber, B., A semantic database framework for context management in heterogeneous wireless networks. In: Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 35–39, 11–13 October 2010
- Charlet, J., Declerck, G., Dhombres, F., Gayet, P., Miroux, P., Vandenbussche, P.Y.: Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et de modélisation. In: 23th journées francophones d'Ingénierie des connaissances, IC 2012, Paris, France, pp. 33–48, June 2012

- Kalou, A., Solomou, G., Pierrakeas, C., Kameas, A.: An ontology model for building, classifying and using learning outcomes. In: 12th IEEE International Conference on Advanced Learning Technologies, pp. 61–65 (2012)
- Uschold, M., Grüninger, M.: Ontologies: principles, methods and applications. Knowl. Eng. Rev. 11(2), 93–155 (1996)
- Bachiment, B.: Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances. In: Charlet, J., Zacklad, M., Kassel, G., Bourigault, D. (eds.): Ingénierie des connaissances : évolutions récentes et nouveaux défis, pp. 305–323. Eyrolles, Paris (2000)
- 22. Guarino, N., Welty, C.: Supporting ontological analysis of taxonomic relationships. Data Knowl. Eng. **39**, 51–74 (2001)
- El-Ghalavini, H., Odeh, M., McClatchey, R., Solomonides, T.: Reverse engineering ontology to conceptual data models. In: Hamza, M.H. (ed.): IASTED International Conference on Databases and Applications, Part of the 23rd Multi-Conference on Applied Informatics, Innsbruck, Austria, 14–16 February 2005
- Guarino, N., Welty, C.: Identity, unity, and individuality: towards a formal toolkit for ontological analysis. In: Proceedings of ECAI 2000: The European Conference on Artificial Intelligence. IOS Press, Berlin (2000)
- Rector, A.L.: Normalization of ontology implementations: towards modularity, re-use, and maintainability. In: Proceedings Workshop on Ontologies for Multi-agent Systems (OMAS) in Conjunction with European Knowledge Acquisition Workshop, pp. 1–16 (2002)
- Rector, A.L.: Modularization of domain ontologies implemented in description logics and related formalisms including owl. In: Proceedings of the 2nd International Conference on Knowledge Capture. ACM (2003). https://doi.org/10.1145/945645.945664. ISBN 1-58113-583-1
- Bachimont, A.I., Troncy, R.: "Semantic commitment for designing ontologies": a proposal. In: Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Lecture Notes in Computer Science, vol. 2473, pp. 114–121. Springer, Heidelberg (2002)
- 28. Vyšniauskas, E., Nemuraitė, L., Paradauskas, B.: Preserving semantics of OWL 2 ontologies in relational databases using hybrid approach. Inf. Technol. Control **41**(2), 103–115 (2012)

Author Index

A

Abdessadek, Aaroud, 21, 496 Achatbi, Iman, 129 Adnani, Fatiha El, 349 Ain El Hayat, Soumiya, 488 Ait tchakoucht, Taha, 416 Aksasse, Brahim, 78, 86, 95, 179, 362 Ali, El Hore, 21 Alsulami, Majid H., 3 Amechnoue, Khalid, 50, 129 Amrouch, Mustapha, 138, 438 Aoulad Allouch, Saloua, 129 Askane, Younes, 503 Atkins, Anthony S., 3 Aziz, Abdel Karim, 208 Azrour, Mourade, 239

B

Bahaj, Mohamed, 69, 119, 196, 374, 396, 488 Bekkari, Aissam, 427 Ben Chigra, Younes, 280 Ben Laadar, Hajar, 196 Benzaima, M., 496 Benzekri, Fatiha, 461 Bouhorma, Mohamed, 280 Boukil, Samir, 349 Boumlik, Abdeljalil, 396 Bousalem, Zakaria, 383 Bousseta, Mohamed, 324

С

Campion, Russell J., 3 Chahbar, Mostafa, 503 Chawki, Youness, 95 Cherrat, Loubna, 162, 349 Cherti, Ilias, 196, 383 Chiheb, Raddouane, 449

D

Daanoun, Ali, 474

Е

Ech-Charrat, Mohammed Rida, 50 El Afia, Abdellatif, 103, 449 El Asnaoui, Khalid, 95, 179 El Azhari, Maryam, 227 El Fazazy, Khalid, 150 El Haimoudi, Khatir, 474 El Hankouri, M., 59 El Hore, Ali, 496 El Moussaid, Nadya, 227 Elhore, Ali, 503 Ezzahir, Redouane, 150 Ezziyyani, Mostafa, 162, 349, 416

F

Farhaoui, Yousef, 239, 362, 408 Fehis, Saad, 302

G

Gaou, Salma, 427 Ghadi, Abderrahim, 280

H

Hajar, Ben Laadar, 216 Hamdaoui, Youssef, 36, 269

I

Ilias, Cherti, 12, 216 Issati, Ikram, 474

J

Jebbor, Sara, 449 Jebrane, Asma, 324, 333

K

Kechadi, Mohand-Tahar, 302 Khallouki, Hajar, 69 Kharbach, M., 59 Khoukhi, Faddoul, 208

© Springer International Publishing AG 2018 M. Ezziyyani et al. (eds.), *Advanced Information Technology, Services and Systems*, Lecture Notes in Networks and Systems 25, https://doi.org/10.1007/978-3-319-69137-4 517

L

Laabira, Hicham, 150 Laadidi, Yassine, 119

М

Maach, Abdelilah, 36, 269 Mahani, Zouhir, 438 Maizate, Abderrahim, 252 Makhlouf, Dardour, 291 Meddah, Naima, 324, 333 Medouri, Abdellatif, 315 Mezouar, Houda, 103 Mohammed, Mensouri, 21, 496 Mostafa, Ezziyyani, 291 Moutaouakkil, Abd Elmajid El, 349 Mzili, Ilyass, 461

Ν

Nouali, Omar, 302

0

Okba, Kazar, 291 Ouanan, Hamid, 78, 86 Ouanan, Mohammed, 78, 86, 95, 179, 239 Ouardouz, M., 59 Ouhda, Mohamed, 179 Ouzzif, Mohamed, 252

R

Rabi, Mouhcine, 138, 438 Riffi, Mohammed Essaid, 461

S

Salah, Mohammed Saïd, 252 Salaheddine, Akazzou, 12 Soltane, Merzoug, 291 Sossi Alaoui, Safae, 362 Soussi, Nassima, 374

Т

Toumanari, Ahmed, 227, 324, 333 Toumi, Mohamed, 252

Z

Zhao, Gansen, 383 Ziani, Ahmed, 315 Zouadi, Tarik, 50