

Lecture Notes in Electrical Engineering 276

James J. (Jong Hyuk) Park
Ivan Stojmenovic
Min Choi
Fatos Xhafa *Editors*

Future Information Technology

FutureTech 2013

 Springer

Lecture Notes in Electrical Engineering

Volume 276

For further volumes:

<http://www.springer.com/series/7818>

James J. (Jong Hyuk) Park · Ivan Stojmenovic
Min Choi · Fatos Xhafa
Editors

Future Information Technology

FutureTech 2013

Editors

James J. (Jong Hyuk) Park
Department of Computer Science
and Engineering
Seoul University of Science
and Technology
Seoul
Republic of Korea

Ivan Stojmenovic
University of Ottawa
Ottawa, Ontario
Canada

Min Choi
School of Information and Communication
Engineering
Chungbuk National University
Chungbuk
Republic of Korea

Fatos Xhafa
Department of Languages and Informatics
Systems
Polytechnic University of Catalonia
Barcelona
Spain

ISSN 1876-1100

ISBN 978-3-642-40860-1

DOI 10.1007/978-3-642-40861-8

Springer Heidelberg New York Dordrecht London

ISSN 1876-1119 (electronic)

ISBN 978-3-642-40861-8 (eBook)

Library of Congress Control Number: 2013947707

© Springer-Verlag Berlin Heidelberg 2014

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

Message from the FutureTech 2013 General Chairs

FutureTech 2013 is the FTRA 8th event of the series of international scientific conference. This conference takes place September 4–6, 2013, in Gwangju Korea. The aim of the FutureTech 2013 was to provide an international forum for scientific research in the technologies and application of information technologies. It was organized by the Korea Information Technology Convergence Society in cooperation with Korea Information Processing Society. FutureTech 2013 is the next edition of FutureTech 2012 (Vancouver, Canada), FutureTech 2011 (Loutraki, Greece), FutureTech 2010 (Busan, Korea, May 2010) which was the next event in a series of highly successful the International Symposium on Ubiquitous Applications & Security Services (UASS-09, USA, Jan. 2009), previously held as UASS-08 (Okinawa, Japan, Mar. 2008), UASS-07 (Kuala Lumpur, Malaysia, August, 2007), and UASS-06 (Glasgow, Scotland, UK, May, 2006).

The papers included in the proceedings cover the following topics: Hybrid Information Technology High Performance Computing, Cloud and Cluster Computing, Ubiquitous Networks and Wireless Communications Digital Convergence, Multimedia Convergence, Intelligent and Pervasive Applications, Security and Trust Computing, IT Management and Service Bioinformatics and Bio-Inspired Computing, Database and Data Mining, Knowledge System and Intelligent Agent, Game and Graphics Human-centric Computing and Social Networks, Advanced Mechanical Engineering, Computer Aided Machine Design, Control and Automations & Simulation. Accepted and presented papers highlight new trends and challenges of future and information technology. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to Steering Chairs: James J. (Jong Hyuk) Park (SeoulTech, Korea), Hamid R. Arabnia (The University of Georgia, USA). Our special thanks go to the Program Chairs: Min Choi (Chungbuk National University, Korea), Uyen Trang Nguyen (York University, Canada), Robert C. Hsu (Chung Hua University, Taiwan), Horacio Gonzalez-Velez (National College of Ireland, Ireland), Yejun He (Shenzhen University, China), all

Program Committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

September 2013

Doo-soon Park, SoonChunHyang University, Korea
Ivan Stojmenovic, University of Ottawa, Canada
Hsiao-Hwa Chen, National Cheng Kung University, Taiwan
Fatos Xhafa, Technical University of Catalonia, Spain

FutureTech 2013 General Chairs

Message from the FutureTech 2013 Program Chairs

Welcome to the FTRA 8th FTRA International Conference on Future Information Technology (FutureTech 2013), which will be held in Gwangju, Korea on September 4–6, 2013. FutureTech 2013 will be the most comprehensive conference focused on the various aspects of information technology. FutureTech 2013 will provide an opportunity for academic and industry professionals to discuss recent progress in the area of information technology. In addition, the conference will publish high quality papers which are closely related to the various theories and practical applications in information technology. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in these important subjects.

For FutureTech 2013, we received many paper submissions, after a rigorous peer review process, we accepted the articles with high quality for the FutureTech 2013 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least three reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation. We would like to sincerely thank the following invited speaker who kindly accepted our invitations, and, in this way, helped to meet the objectives of the conference: Prof. Victor C.M. Leung, University of British Columbia, Vancouver, Canada. Finally, we would like to thank all of you for your participation in our conference, and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

Min Choi, Chungbuk National University, Korea
Uyen Trang Nguyen, York University, Canada
Robert C. Hsu, Chung Hua University, Taiwan
Horacio Gonzalez-Velez, National College of Ireland, Ireland
Yejun He, Shenzhen University, China

FutureTech 2013 Program Chairs

Message from the DIIK 2013 Session Chairs

DIIK 2013 is the first Special Session on the Data-Intensive Intelligence and Knowledge co-organized by KISTI and ISTIC, which is held in conjunction with the 2nd International Conference on Ubiquitous Context-Awareness and Wireless Sensor Network (UCAWSN-13) at Jeju Island, Korea, in July 15–17, 2013. The aim of this session is to discuss key research issues and practices of intelligence and knowledge driven by big data. The research on data-driven intelligence and knowledge has been recently encouraged by the appearance of disruptive technologies/services such as Apple's Siri and IBM's Watson as well as the progress in distributed and parallel computing. DIIK 2013 will provide an opportunity for participants to share their research efforts and ideas between industry and academia.

For DIIK 2013, we had 20 submissions, among which 16 papers with high quality were accepted for publication. Every paper was reviewed by at least two reviewers in this session Program Committee. We take this opportunity to thank for their great contributions. We also wish to express our special thanks to the UCAWSN-13 chairs including Prof. James J. Park and Prof. Hwa-Young Jeong for allowing and helping this session to be successful. Finally, we heartily thank all the authors for their valuable contributions.

September 2013

Hanmin Jung and Lijun Zhu

DIIK 2013 Special Session Chairs

Organization

Steering Chairs

James J. Park
Hamid R. Arabnia

SeoulTech, Korea
The University of Georgia, USA

General Chairs

Doo-soon Park
Ivan Stojmenovic
Hsiao-Hwa Chen
Fatos Xhafa

SoonChunHyang University, Korea
University of Ottawa, Canada
National Cheng Kung University, Taiwan
Technical University of Catalonia, Spain

Program Chairs

Min Choi
Uyen Trang Nguyen
Robert C. Hsu
Horacio Gonzalez-Velez
Yejun He

Chungbuk National University, Korea
York University, Canada
Chung Hua University, Taiwan
National College of Ireland, Ireland
Shenzhen University, China

Workshop Chairs

Seung-Ho Lim
Jaime Lloret Mauri
Yu Chen

Hankuk University of Foreign Studies, Korea
Universidad Politecnica de Valencia, Spain
State University of New York - Binghamton,
USA

Advisory Committee

Seok Cheon Park
Makoto Takizawa
Mohammad S. Obaidat

Gachon University, Korea
Seikei University, Japan
Monmouth University, USA

Han-Chieh	Chao National Ilan University, Taiwan
Laurence T. Yang	St. Francis Xavier University, Canada
Young-Sik Jeong	Dongguk University, Korea
Albert Zomaya	University of Sydney, Australia
Wanlei Zhou	Deakin University, Australia
Ruppa K. Thulasiram	University of Manitoba, Canada
Hai Jiang	Arkansas State University, USA
Yuanchun Shi	Tsinghua University, China

Publicity Chairs

Cain Evans	Birmingham City University, UK
Lee Chulung	Korea University, Korea
Antonio Coronato	ICAR, Italy
Rung-Shiang Cheng	Kunshan University, Taiwan
Haixia Zhang	Shandong University, China
Rafael Falcon	Larus Technologies, Canada
Xu Shao	Institute for Infocomm Research, Singapore

Publication Chair

Hwa Young Jeong	Kyung Hee University, Korea
-----------------	-----------------------------

Local Arrangement Chairs

Cheonshik Kim	Sejong University, Korea
Namje Park	Jeju National University, Korea
Deok-Gyu Lee	ETRI, Korea
Young Yoon Cho	Sunchon National University, Korea

Program Committee

Abdelbadeeh Salem	Ain Shams University, Egypt
Alfredo Cuzzocrea	University of Calabar, Italy
Antoni Ligeza	AGH University of Science and Technology, Poland
Antonis Gasteratos	Democritus University of Thrace, Greece
Ashkan Sami	Shiraz University, Iran
Been-Chian Chien	National University of Tainan, Taiwan
Bing Chen	Memorial University of Newfoundland, Canada
Carson K. Leung	University of Manitoba, Canada
Chan Yeob Yeun	Khalifa University of Science, Technology and Research, UAE
Chao Wang	The University of North Carolina at Charlotte, USA
Chi-Fu Huang	National Chung Cheng University, Taiwan

Ching-Hsien Hsu	Chung Hua University, Taiwan
Cliff Zou	University of Central Florida, USA
Colin Walter	Royal Holloway, University of London, United Kingdom
Dragan Ivetić	University of Novi Sad, Republic of Serbia
Edward Hua	QED Systems, USA
Eike Schallehn	Otto von Guericke University of Magdeburg, Germany
Epaminondas Kapetanios	University of Westminster, United Kingdom
Evgeny Pyshkin	St., Petersburg State Polytechnical University, Russia
Geng YANG	Nanjing University of Posts and Telecommunications, China
Guillermo Diaz Delgado	Autonomous University of Queretaro, Mexico
Haggai Roitman	Haifa University, Cisrael
Harald Kosch	University of Passau, Germany
Hariharan Shanmugasundaram	TRP Engineering College, India
Hiroshi Ishikawa	National University Corporation Shizuoka University, Japan
Hongmei Chi	Florida Agricultural and Mechanical University, USA
Jeng-Shyang Pan	National Kaohsiung University of Applied Sciences, Taiwan
Jeong Hyun Yi	Soongsil University, Korea
Jiehan Zhou	University of Oulu, Finland
Jiqiang Lu	Institute for Infocomm Research, Singapore
Joel Rodrigues	University of Beira Interior, Portugal
Jose Antonio Onieva González	Universidad de Málaga, Spain
Kuo-Chan Huang	National Taichung University of Education, Taiwan
Liu Jianxun	Human University of Science and Technology, China
Maciej Piasecki	Wroclaw University of Technology, Poland
Maumita Bhattacharya	Charles Sturt University, Argentina
Mei-Ling Shyu	University of Miami, USA
Mudasser Wyne	National University, USA
Muhammad Usman	Auckland University of Technology, New Zealand
Natalija Vlajic	York University, Canada
Nima Kaviani	University of British Columbia, Canada
Paolo Bottoni	University of Rome La Sapienza, Italy
Paul Kwan	University of New England, Argentina
Pit Pichappan	AISB, Saudi Arabia
Qian Zhu	Accenture Technologies, Canada
Qing Yang	Montana State University, USA
Rajkumar Buyya	University of Melbourne, Argentina

Ren-Song Ko	National Chung Cheng University, Taiwan
Roberto Caldelli	Università degli Studi di Firenze, Italy
Roman Neruda	Academy of Sciences of the Czech Republic, Czech
Ruben Rios	Universidad de Málaga, Spain
Ryszard Tadeusiewicz	AGH University of Science and Technology, Poland
Schahram Dustdar	Technische Universität Wien, Austria
Seng Loke	La Trobe University, Argentina
Sheng-De Wang	National Taiwan University, Taiwan
Shiuh-Jeng Wang	Central Police University, Taiwan
Shu-Ching Chen	Florida International University, USA
Song Fu	University of North Texas, USA
Soon M. Chung	Wright State University, USA
Suren Byna	Lawrence Berkeley National Laboratory, USA
Tania Cerquitelli	Politecnico di Torino, Italy
Tatjana Davidovic	Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia
Toshiyuki AMAGASA	University of Tsukuba, Japan
Troels Andreasen	Roskilde University, Denmark
Vitaly Klyuev	University of Aizu, Japan
Vitomir Kovanovic	Simon Fraser University, Canada
Wei Feng Chen	California University of Pennsylvania, USA
Wei-Chuen Yau	Multimedia University, Malaysia
Willy Picard	Poznań University of Economics, Poland
Władysław Homenda	Instytut Badań Systemowych Polskiej Akademii Nauk, Poland
Woockey Lee	INHA University, Korea
Xiaofeng Chen	Xidian University, China
Yas Alsultanny	Arabian Gulf University, Bahrain
Yo-Ping Huang	National Taipei University of Technology, Taiwan
Young-Sik Jeong	Dongguk University, Korea
Yu-Chen Hu	Providence University, Taiwan
Yue-Shan Cha	National Taipei University, Taiwan
Yunquan Zhang	State Key Lab of Computer Science, China
Zhiqiang Zhang	Harbin Engineering University, China
Zoubir MAMMERI	Unité Mixte de Recherche, France

Data-Intensive Intelligence and Knowledge - Special Session Organization -

Organizer/Chair

Hanmin Jung	Korea Institute of Science and Technology Information, Korea
Lijun Zhu	Institute of Scientific and Technical Information of China, China

Program Committee

Changki Lee	Kangwon University, Korea
DongwonJeong	Kunsan National University, Korea
Haklae Kim	Samsung Electronics Co., Ltd., Korea
Harksoo Kim	Kangwon University, Korea
Hyunchul Jang	Korea Institute of Oriental Medicine, Korea
Ing-Xiang Chen	Ericsson Taiwan Ltd., Taiwan
In-Su Kang	University of Kyungsung, Korea
Jinhyung Kim	Korea Institute of Science and Technology Information, Korea
Michaela Geierhos	University of Paderborn, Germany
Sa-Kwang Song	Korea Institute of Science and Technology Information, Korea
Seung-Hoon Na	Electronics and Telecommunications Research Institute, Korea
Seungwoo Lee	Korea Institute of Science and Technology Information, Korea
ShuoXu	Institute of Scientific and Technical Information of China, China
Sung-Pil Choi	Korea Institute of Science and Technology Information, Korea
Yeong-Su Lee	Cylex Inc., Germany
Yunliang Zhang	Institute of Scientific and Technical Information of China, China
Zhangbing Zhou	Institute TELECOM & Management SudParis, France

Contents

Hybrid Information Technology

Dedicated Smart Software System for Mobile X-Ray 1

Chang Won Jeong, Young Sik Jeong, Jinseok Lee, Su Chong Joo

**Real-Time Intuitive Terrain Modeling by Mapping Video
Images onto a Texture Database** 7

*Wei Song, Seongjae Cho, Kyungeun Cho, Kyhyun Um, Chee Sun Won,
Sungdae Sim*

**The RIP for Random Matrices with Complex Gaussian
Entries** 13

Kuo Xu, Jian Wang, Byonghyo Shim

**Evolving Mobile App Recommender Systems: An Incremental
Multi-objective Approach** 21

Xiao Xia, Xiaodong Wang, Xingming Zhou, Bo Liu

Project Proposals Ranking 29

Sylvia Encheva

**A Stereo Micro Image Fusion Algorithm Based on
Expectation-Maximization Technique** 35

*Cuixia Bai, Gangyi Jiang, Mei Yu, Yigang Wang,
Feng Shao, Zongju Peng*

High Performance Computing

Moldable Job Scheduling for HPC as a Service 43

Kuo-Chan Huang, Tse-Chi Huang, Mu-Jung Tsai, Hsi-Ya Chang

MapReduce Example with HBase for Association Rule 49

Jongwook Woo, Kilhung Lee

Cloud and Cluster Computing

Service Level Agreement Renegotiation Framework for Trusted Cloud-Based System	55
<i>Ahmad Fadzil M. Hani, Irving Vitra Paputungan, M. Fadzil Hassan</i>	

A Cross-IdP Single Sign-On Method in SAML-Based Architecture	63
<i>Tzu-I Yang, Chorng-Shiuh Koong, Chien-Chao Tseng</i>	

Live Virtual Machine Migration with Optimized Three-Stage Memory Copy	69
<i>Feiran Yin, Weidong Liu, Jiaying Song</i>	

Ubiquitous Networks and Wireless Communications

Performances of New Chaotic Interleaver Design in OFDM-IDMA System	77
<i>Brahim Akbil, Driss Aboutajdine</i>	

Intelligent and Pervasive Applications

Application of an Artificial Intelligence Method for Diagnosing Acute Appendicitis: The Support Vector Machine	85
<i>Sung Yun Park, Jun Seok Seo, Seung Chul Lee, Sung Min Kim</i>	

On Extracting Important User Preferences	93
<i>Sylvia Encheva</i>	

Object Retrieval Scheme Using Color Features in Surveillance System	99
<i>Su-wan Park, JeongNyeo Kim, Jong Wook Han</i>	

Multi Criteria Decision Making Related to Services	107
<i>Sylvia Encheva</i>	

Security and Trust Computing

Keyword Searchable Encryption with Access Control from a Certain Identity-Based Encryption	113
<i>Koji Tomida, Masami Mohri, Yoshiaki Shiraishi</i>	

Attribute-Based Encryption with Attribute Revocation and Grant Function Using Proxy Re-encryption and Attribute Key for Updating	119
<i>Takeru Naruse, Masami Mohri, Yoshiaki Shiraishi</i>	

Research on Stereo Image Authentication Watermarking with Self-recovery	127
<i>Ting Luo, Gangyi Jiang, Mei Yu, Yigang Wang, Feng Shao, Zongju Peng</i>	
Fine-Grained Access Control in Object-Oriented Databases	133
<i>Rahat Masood, Muhammad Awais Shibli, Abdul Ghafoor</i>	
Doubly Encrypted Identity-Based Encryption for File Transfer Service	139
<i>Makoto Sato, Masami Mohri, Hiroshi Doi, Yoshiaki Shiraishi</i>	
Toward a Honeypot Solution for Proactive Security in Vehicular Ad Hoc Networks	145
<i>Dhavy Gantsou, Patrick Sondi</i>	
IT Management and Service	
DWL Tool for Creating a Customized Web-Based System Generator	151
<i>Ling-Hua Chang, Sanjiv Behl, Tung-Ho Shieh, Chin-Chih Ou</i>	
The Task-Oriented Circulation Planning	157
<i>Tzu-I Yang, Chornng-Shiuh Koong, Chien-Chao Tseng</i>	
Database and Data Mining	
A Framework for Selecting the Optimal Technique Suitable for Application in a Data Mining Task	163
<i>Haruna Chiroma, Sameem Abdul-Kareem, Adamau Abubakar</i>	
Discovery of Closed Consensus Temporal Patterns by Group Decision Making	171
<i>Tony Cheng-Kwi Huang</i>	
A Collaborative Filtering Recommendation Algorithm Based on Tag Clustering	177
<i>Rujuan Liu, Zhendong Niu</i>	
BSTree: An Incremental Indexing Structure for Similarity Search and Real Time Monitoring of Data Streams	185
<i>Abdelwaheb Ferchichi, Mohamed Salah Gouider</i>	
Knowledge System and Intelligent Agent	
Learning Teaching in Teaching: Online Reinforcement Learning for Intelligent Tutoring	191
<i>Fangju Wang</i>	

Minimax Self Playing Game Alquerque 197
*Ishtiaq Ahmed, Donghai Guan, Md. Nasir Uddin Laskar,
 Tae Choong Chung*

Game and Graphics

**Fractal Based Hardware Accelerated Technique for Graphical
 Rendering** 205
Divya Udayan J., HyungSeok Kim, Jun Lee, Jee-In Kim

Topics on Public Game Art Using Media Façade..... 213
Hyoyoung Kim, Jin Wan Park

**A Physical Interactive Game for Learning English
 of Children** 219
Sang-I Shin, Kyoungju Park

**Using Multiple Verifiers to Detect Sybils in a Social Network
 Graph** 225
Kyungbaek Kim

**The Impact of Addiction to Twitter among University
 Students**..... 231
Shahnul Asmar Saaid, Nalisa Alia Amin Al-Rashid, Zaridah Abdullah

**Design and Prototype Implementation of Smart-Phone Voice
 Locker Using Voice Recognition** 237
Won Min Kang, Ki Won Lee, Ji Soo Park, Jong Hyuk Park

Digital Forensics and Information Security

Extended RBAC Model with Task-Constraint Rules 245
Li Ma, Yanjie Zhou, Wei Duan

Historical Data Recovery from Android Devices 251
Yitao Yang, Zhiyue Zu, Guozi Sun

A Static Semantic Model for Trusted Forensics Using OCL 259
Zehui Shao, Qiufeng Ding, Xianli Jin, Guozi Sun

**A Novel Hybrid Cellular Automata Based Cipher System for
 Internet of Things** 269
*Mouza Ahmed Bani Shemali, Chan Yeob Yeun,
 Mohamed Jamal Zemerly, Khalid Mubarak*

An Efficient Computer Forensics Selective Imaging Model 277
*Waleed Halboob, Khaled S. Alghathbar, Ramlan Mahmood,
 Nur Izura Udzir, Mohd. Taufik Abdullah, Ali Deghantanha*

Cloud Computing Risk Assessment: A Systematic Literature Review	285
<i>Rabia Latif, Haider Abbas, Saïd Assar, Qasim Ali</i>	
Security Requirements Specification Framework for Cloud Users	297
<i>Rida Naveed, Haider Abbas</i>	
Digital Evidence Bag Selection for P2P Network Investigation	307
<i>Mark Scanlon, Tahar Kechadi</i>	
A Research for Partition Restoration Techniques	315
<i>Jaeung Namgung, Il young Hong, Jungheum Park, Changhoon Lee, Sangjin Lee</i>	
Definition of Evaluation Index Model for Network Management System	323
<i>Fei Xu, Jinqiao Shi, K.P. Chow, Xiaojun Chen, Peipeng Liu</i>	
Investigating and Measuring Capabilities of the Forensics File Carving Techniques	329
<i>Khawla Alghafli, Andrew Jones, Thomas Martin</i>	
Application for Reversible Information Hiding in Multilpe Secret Images Sharing Based on Shamir's Scheme	337
<i>Chyuan-Huei Thomas Yang, Che-Lun Chung, Yu Min Lin, Song Yong Fan</i>	
The Arm Strength Training Machine with Biofeedback	343
<i>Tze-Yee Ho, Mu-Song Chen, Yuan-Joan Chen, Hung-Yi Chen</i>	
Privacy Breach Investigations of Incident Response to Personal Information Protection Act	351
<i>Da-Yu Kao, Cheng-Yu Peng, Frank Fu-Yuan Huang, Shiuh-Jeng Wang</i>	
Measuring Digital Crime Investigation Capacity to Guide International Crime Prevention Strategies	361
<i>Joshua I. James, Yunsik Jake Jang</i>	
Management of Future Convergence Technology	
Study on the Growth Strategy to Become a Global Logistics Company, through the Expansion of the Domestic Logistics Companies in China	367
<i>Hyunwoo Kim, Chulung Lee</i>	

A Dynamic Model of Technological Innovation in 3D TV Industry: Case of LG Electronics	375
<i>Jun-oh Hwang</i>	
The U-Work Utilization Analysis Using Groupware on Non-profit Organization	383
<i>Kyeong Hui Du, Chulung Lee</i>	
Analysis of the Risks of Overseas Advancement by Logistics Companies Applying AHP	391
<i>Mihyung Kim, Ki-sung Hong, Chulung Lee</i>	
Development of an Operating System for Optimization of the Container Terminal by Using the Tandem-Lift Quay Crane	399
<i>Sang-Hei Choi, Hyeonu Im, Chulung Lee</i>	
Knowledge Sharing and the Forms of R&D Collaboration	405
<i>Sang-Ho Kook</i>	
Global Supply Chain Management Using Business Risk Re-alignment via the Change of the Transfer Pricing Methodology	413
<i>Sang Min Ahn, Ki-sung Hong, Chulung Lee</i>	
Advanced Technology and Computer	
Research of Touchscreen Terminals Gesture Operation Error Based on Kansei Engineering	421
<i>Rong Qin, Dongxiang Chen, Xuelong Hou</i>	
Heart Sound Feature Extraction Based on Wavelet Singular Entropy	429
<i>Zhang Lu</i>	
Research on MTMP Structure Chlorine Dosing Decoupling Control	435
<i>Xie Peizhang, Zhou Xingpeng</i>	
Social Computing, Network, and Services	
The Research on Electronic Data Forensic Institutions Equipment Configuration Standards	441
<i>Mai Yonghao, K.P. Chow, Zhou Gang, Lu Zhengwu, Zhang Jun</i>	
Design of an RDFizer for Online Social Network Services	449
<i>Junsik Hwang, Hyosook Jung, Sujin Yoo, Seongbin Park</i>	

Cloud Enhanced Information Fusion

A Holistic Cloud-Enabled Robotics System for Real-Time Video Tracking Application	455
<i>Bingwei Liu, Yu Chen, Erik Blasch, Khanh Pham, Dan Shen, Genshe Chen</i>	
Improving Network Health Monitoring Accuracy Based on Data Fusion for Software Defined Networking	469
<i>Sejun Song</i>	
Improving Diagnostic Accuracy Using Multiparameter Patient Monitoring Based on Data Fusion in the Cloud	473
<i>Zhanpeng Jin, Xiaoliang Wang, Qiong Gui, Bingwei Liu, Sejun Song</i>	
Author Index	477

Dedicated Smart Software System for Mobile X-Ray

Chang Won Jeong¹, Young Sik Jeong², Jinseok Lee³, and Su Chong Joo³

¹ Imaging Science based Lung and Bone Disease Research Center,
Wonkwang University, 460 Iksandaero, Iksan, Jeonbuk, 570-749, Republic of Korea
mediblue@wku.ac.kr

² Department of Multimedia Engineering, Dongguk University, Seoul, Korea
ysjeong2k@gmail.com

³ Department of Computer Engineering, Wonkwang University,
460 Iksandaero, Iksan, Jeonbuk, 570-749, Republic of Korea
gonasago@gmail.com,
scjoo@wku.ac.kr

Abstract. In this paper, we describe dedicated smart software for mobile x-ray system. It can support mobile x-ray system and medical information system. Also, it provides web services that allow us to deal with the DICOM of medical information standard. In order to implement this smart software system, we describe about a PACS environment that is a kind of web based system. And we also describe the system architecture as a physical environment and process of system components. And then, we show the results of service such as mobile viewer and file upload service.

Keywords: mobile x-ray system, dedicated smart software, medical information system, DICOM.

1 Introduction

Recently, healthcare and variety of IT technology have combined with a medical device that related products have been released. In particular, the introduction of smart software has led to the integration of existing medical information system[1]. In addition, smart software technology is the ubiquitous healthcare that can make new medical service for the hospital. Further, various mobile x-ray systems have been developed in order to do the emergency environments such as emergency rooms, operating theatres and portable imaging at anywhere. It is expanding worldwide due to the technology's compelling advantages in productivity, X-ray dose and image quality. For instance, typical systems such as Carestream XDR[2], GE healthcare AMS Mobile X-RAY system[3] and Siemens[4] etc. Most of systems focus on mobile, portable and wireless using ICT technology.

We have been developing one system of mobile x-ray for diagnostics at out-patient clinics, operating room and emergency room. Therefore, we are required to research on how to use the smart device based on it and system development of mobile x-ray that can collaborate with the medical information system(by INFINITT Company).

Firstly, we define the structure of the system and design an interface between each component. This is called Medical information system or PACS and dedicated smart device. It was adopted by basing on the Android OS[5, 6, 7] and exchanged of information between each system according to the standard DICOM[8, 9]. This paper focuses on the smart software research for supporting mobile x-ray system. It provides web services that are interact with mobile x-ray system and medical information system or PACS.

It is also organized a follows; Section 2 describes the whole system architecture and process of system components. Section 3 shows the implementation of system environments and the result of application service. Last of all, it is on chapter 4, we will describe about the conclusion and the content of future research.

2 Dedicated Smart Software System

2.1 System Architecture

Overall structures of the proposed system are as follows. It can configure outside or inside hospital environment by using the mobile x-ray system and communications infrastructure such as WiFi communication and wire/wireless environments. As Medical information System were include the mobile server which supports the smart devices.

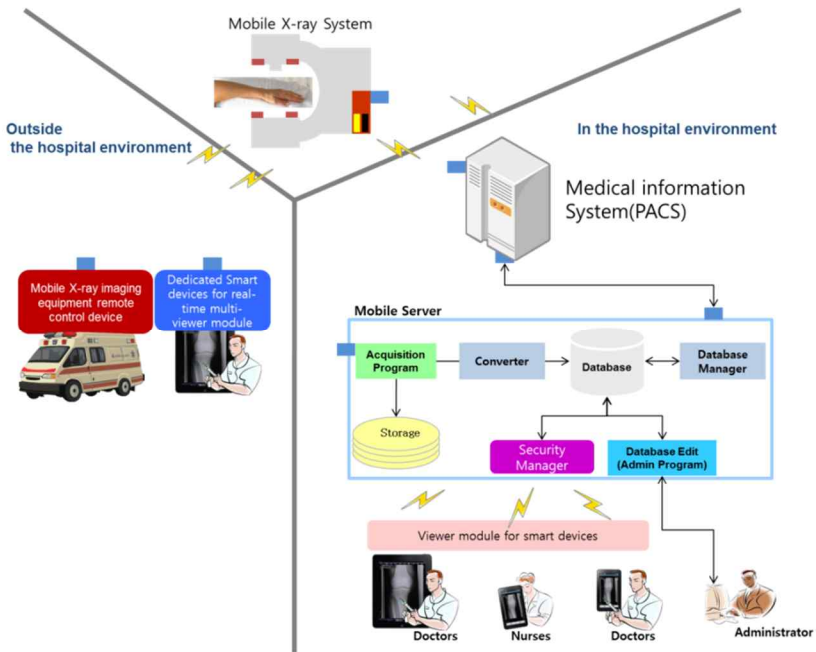


Fig. 1. The system architecture

Mobile x-ray system contains a multi-sensor sensing module that can transfer to the medical information system DICOM files by capturing the patient's body part. After that, it will send what we have captured to the dedicated smart device and Medical information system. It provides Image Information services in collaboration with the mobile x-ray system and the dedicated smart devices. Then, these smart devices in the hospital environment will provide medical information service connected to the medical information system. The dedicated smart device is used as the DICOM file. However, in the connection with the Mobile Server, smart devices generally deal with image files [10].

2.2 Interaction of System Components

Visualizing the whole system, Figure 2 below, shows the process of interaction between the components. The Net gate interacts with multi-sensor sensing module. Also, a medical information system includes the upload service and mobile server for interaction with the client like Web-based client devices. PACS storage that stores the DICOM files consists of a dynamic Access Control Service for all requests security processing.

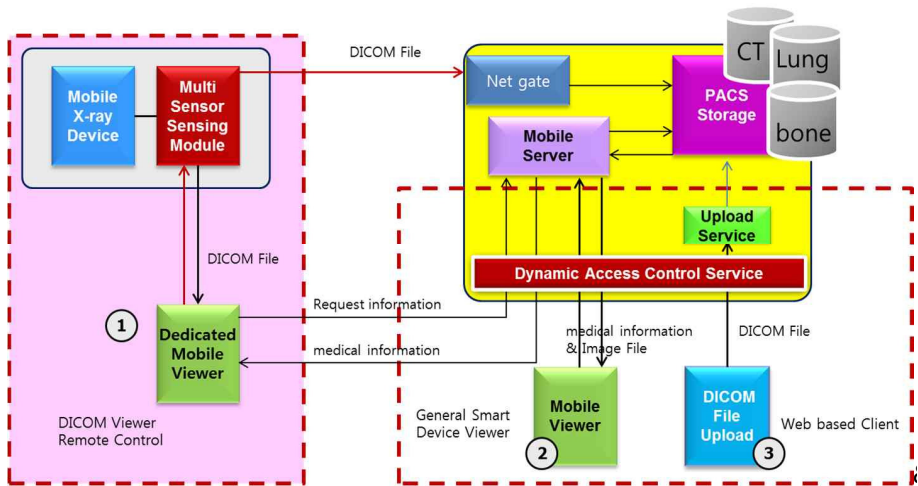


Fig. 2. Interaction of system component

Figure 3 shows the method call interface and interaction between each component. First, in order to generate a DICOM file and save it to the medical information system, mobile x-ray system have to invoke the C-store-RQ() method. Medical information systems, store_to_database(), is used to save to the database. In the Output Viewer, veiw_display(), uses the request(query) method to interact with medical information systems and smart devices, resulting values, transfer the DICOM file by dicomfile_send() method. In addition, the dicomfile_upload() for the DICOM medical information system stores file that is created from other devices. Also, by using the stepmotor_control() for controlling mobile x-ray system, smart device can make the position of the header change.

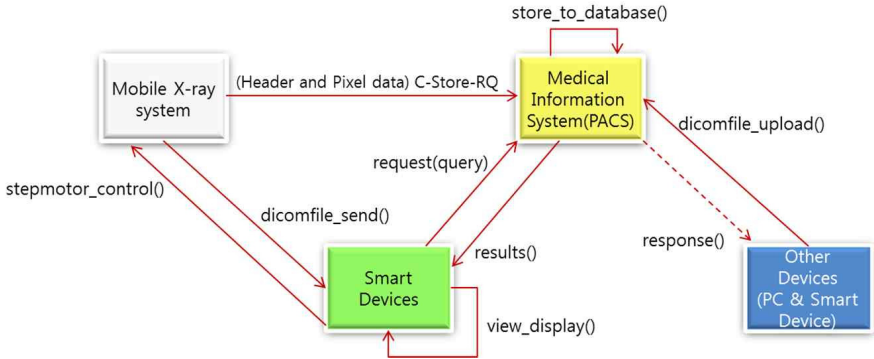


Fig. 3. Interaction of diagram for system and methods of interface

3 Construction of System and Results of Application Service

The physical environment of the proposed system is shown in Fig.4. Mobile x-ray system can be taken only a portion of the body. Multi-sensor sensing module is consisted of Step Motor module for control camera with multi-sensor-based ATMEGA128. Medical information systems are constructed in the IBM Z404 workstation server, the mobile device has the Android OS basing on smart phones, and PC system for view application service is connected to the general Internet.

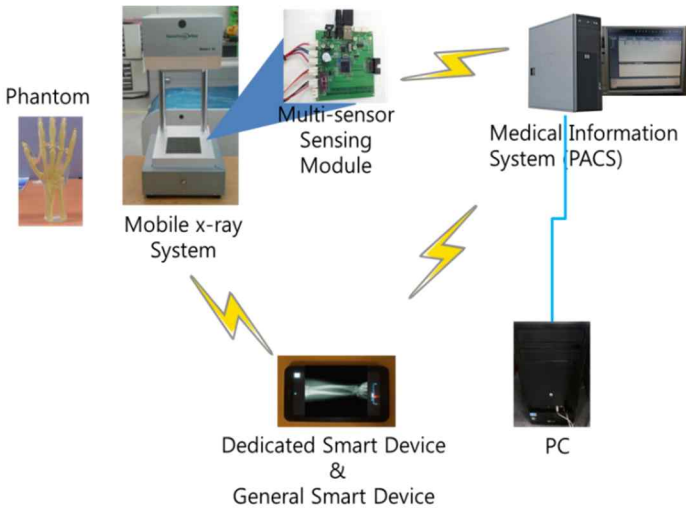


Fig. 4. Physical system environments

Figure 5 shows the results of the DICOM file viewer screen shot with a mobile x-ray system for position control whereas the main screen with the camera and multi-sensor data are collected by the dedicated smart device.



Fig. 5. The body position service result of dedicated smart device

Figure 6 shows the image lists that connect to medical information systems in smart devices.



Fig. 6. The image file viewer of general smart device

Figure 7, shows the results screenshot using the PC viewer and DICOM files that transfer to a specific directory on the medical information system in order to upload the DICOM files from the client PC.

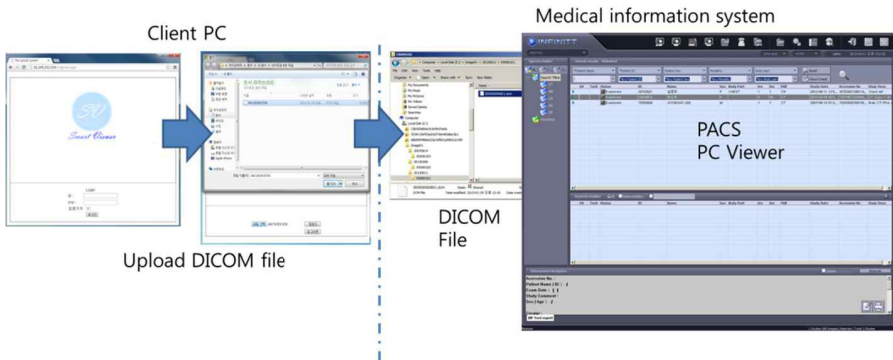


Fig. 7. Shows the results of upload service on medical information system from client PC

4 Conclusion

In this paper, we describe the dedicated smart software system to support mobile x-ray system. We have developed a software environment for medical information services such as viewer and upload service. This overall functionality of the current systems has been completed.

Mobile x-ray system has been developed successfully. In our future studies, we would like to improve the function of each software implementation and the interface between the Multi-sensing modules. We are also planning to add the Dynamical Access Control Service to medical information system.

Acknowledgement. This work was supported by the Korea Health Industry Development Institute (KHIDI) grant No. A120152 under the auspices of the Ministry of Health and Welfare.

References

1. Jeong, C.W., Joo, S.C., Yoon, K.H.: Guide System for Patients in Ubiquitous Hospital Environments. In: Proceeding of International Conference on uHealthcar 2012, Kyeongju, Korea (October 25-27, 2012)
2. CareStream, <http://www.carestream.com/>
3. GE Healthcare, http://www3.gehealthcare.in/en/Products/Categories/Radiography/Mobile_X-Ray_Systems
4. SIEMENS, <http://healthcare.siemens.com/radiography/mobile-x-ray>
5. Pasha, M.F., Supramaniam, S., Liang, K.K., Amran, M.A., Chandra, B.A., Rajeswari, M.: An Android-based Mobile Medical Image Viewer and Collaborative Annotation: Development Issues and Challenges. *International Journal of Digital Content Technology and its Applications (JDCTA)* 6(1), 208–217 (2012)
6. Doukas, C., Pliakas, T., Maglogiannis, I.: Mobile healthcare information management utilizing Cloud Computing and Android OS. In: Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), pp. 1037–1040 (2010)
7. Drnasin, I., Grgic, M.: The use of mobile phones in radiology. In: Proceedings of ELMAR 2010, pp. 17–21 (2010)
8. Faggionia, L., Neria, E., Castellanab, C., Caramellaa, D., Bartolozzia, C.: The future of PACS in healthcare enterprises. *European Journal of Radiology* 78, 253–258 (2010)
9. Dragan, D., Ivetić, D.: Architectures of DICOM based PACS for JPEG2000 Medical Image Streaming. *Computer Science and Information Systems* 6(1), 185–203 (2009)
10. Converting DICOM to JPEG using dcm4che 2, <http://samucs.blogspot.com/2008/03/convertting-dicom-to-jpeg-usingdcm4che.html>

Real-Time Intuitive Terrain Modeling by Mapping Video Images onto a Texture Database

Wei Song¹, Seongjae Cho², Kyungeun Cho^{2,*}, Kyhyun Um²,
Chee Sun Won³, and Sungdae Sim⁴

¹ Dept. of Computer Science & Technology, North China University of Technology, China

² Dept. of Multimedia Engineering, Dongguk University-Seoul, Korea

³ Division of Electronics and Electrical Engineering, Dongguk University-Seoul, Korea

⁴ Agency for Defense Development, Bugyuseong daero 488 beon gi,
Yoseong, Daejeon 305-152, Korea
cke@dongguk.edu

Abstract. Mobile robot operators must make rapid decisions based on information about the robot's surrounding environment. Thus, it is essential that they receive information regarding the local terrain in real time. This paper describes an intuitive terrain modeling method and a generation algorithm for a texture database (TDB). Firstly, we integrate the large-scale 3D point clouds obtained from sensors into a node-based terrain mesh in CPU. Subsequently, we program a graphics processing unit to generate the TDB by mapping the triangles in the terrain mesh onto the captured video images.

Keywords: Terrain reconstruction, GPU, Texture database, Mobile robot.

1 Introduction

The multiple sensors mounted on mobile robots collect 3D point clouds, video images, GPS data, and rotation states [1]. Traditional real-time visualization systems mostly apply a 2D image, a voxel map, or a terrain mesh to represent a terrain model. A terrain mesh is generated by integrating the top points in the x - z cells into a regular triangular mesh [2]. By overlaying captured video images on the 3D terrain mesh, the visualization system provides intuitive imagery of a 3D geometric model for easy terrain perception. Conventionally, we register the captured images to the terrain model. However, when mobile robots navigate a large-scale environment, the sensed images are registered incrementally. The amount of data becomes so large that it exceeds the robots' memory capacity.

The objective of our study is to reconstruct an intuitive large-scale terrain model using limited memory in real time. Firstly, we create a node-based terrain mesh. Each node in the mesh contains a certain quantity of vertices and a node texture. The node texture is generated by mapping the triangles in the captured image, which are projected from the vertices of the node mesh. The node textures are integrated to form

* Corresponding author.

a texture database (TDB). We utilize a graphics processing unit (GPU) to map the triangles in parallel in order to realize real-time TDB generation. Finally, we represent the reconstructed terrain model by overlaying the terrain mesh with the TDB.

This paper is organized as follows: in Section 2, we survey related work on terrain modeling and representation methods. In Section 3, we explain a GPU-based TDB generation system. In Section 4, the performance of the proposed system are analyzed and evaluated. In Section 5, we draw our conclusions.

2 Related Works

It is necessary to reconstruct a terrain model using an integrated dataset obtained from multiple sensors. Rovira-Más [3] proposed a density grid for 3D reconstruction from information obtained from stereo cameras, a localization sensor, and an inertial measurement unit. However, he allocated one color per grid, which caused distortion. To represent an intuitive scene, Sukumar [4] proposed a convenient visualization method by integrating sensed datasets into a texture mesh for the terrain reconstruction. Song [5] produced a texture mesh to represent ground information, and mapped the captured images onto the terrain mesh. There are several colors per triangle of the texture mesh. When the robot navigated for an extended period, the texture buffers registered in the terrain model became so large that they exceeded the memory capacity. Even the compression algorithms for 2D images, such as JPEG [6] and MPEG [7], are difficult to be applied in the real-time terrain modeling with limited memory. Therefore, an effective and rapid TDB generation method is necessary for robots with limited memory capacities.

3 TDB Generation

The multiple sensors mounted on mobile robots collect terrain information in the form of 3D point clouds, 2D images, GPS data, and rotation states. We register the sensed 3D point clouds into a global terrain mesh according to GPS and rotation information [8]. The color information of the terrain models is computed by projection from the vertices in the terrain mesh to the captured 2D images.

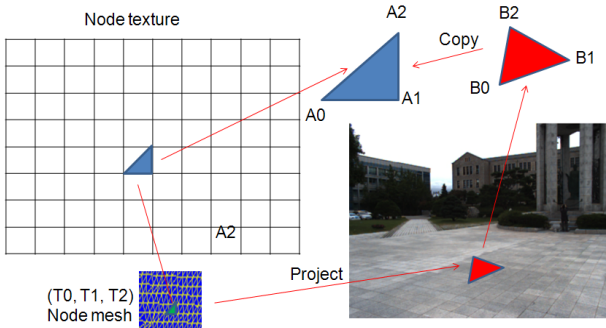


Fig. 1. Node texture generation process

Fig. 1 shows the process of node texture generation. For each 3D triangle (T_0 , T_1 , T_2) in the mesh of that node, we create a triangle (A_0 , A_1 , A_2) in a node texture, which has a set of triangle pixels. Subsequently, a triangle (B_0 , B_1 , B_2) in a captured image is projected from (T_0 , T_1 , T_2). We then duplicate triangle (A_0 , A_1 , A_2) from triangle (B_0 , B_1 , B_2). After all of the triangles in a node mesh are mapped from the current image, the node texture is updated and combined into the TDB system.

In a large-scale environment, there are a large number of triangles of several nodes to be mapped from the captured video images. Hence, to realize real-time TDB generation, we implement the mapping process in parallel by applying GPU programming. The framework of the GPU-based TDB generation system is shown in Fig. 2.

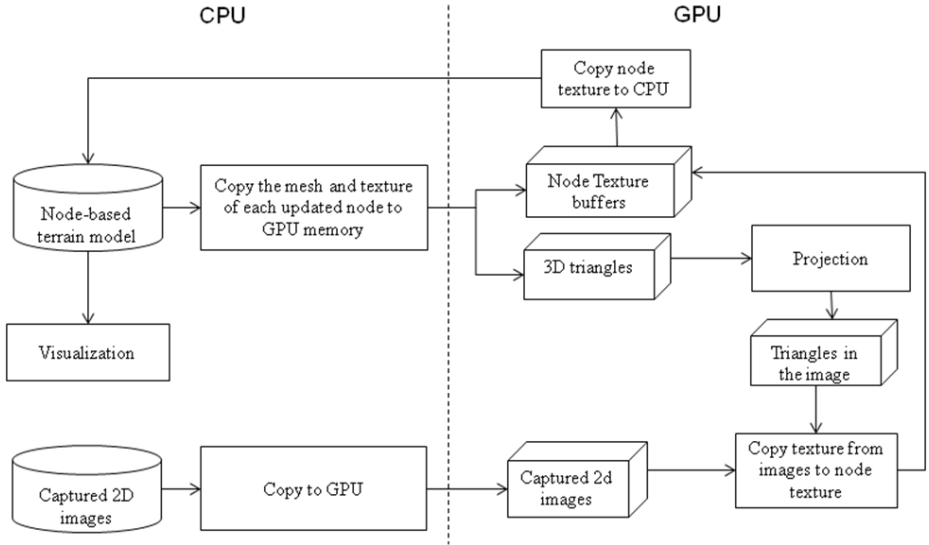


Fig. 2. Framework of the GPU-based TDB generation system

After creating the node-based terrain model, we copy the mesh and texture of an updated node and the current captured image to GPU memory. Subsequently, we project each triangle of the mesh onto the captured image in order to acquire the mapped triangles within the captured image. We then duplicate the mapped triangle to the node texture buffer. Next, we copy the updated node texture in GPU memory to the TDB of the terrain model in CPU memory. Finally, we render the terrain model by overlaying the terrain mesh with the TDB.

4 Experiments

Experiments were performed to test the proposed real-time large-scale terrain modeling system and the TDB generation method. The mobile robot shown in Fig. 3 was used to gather data using its integrated sensors, including a LiDAR sensor, a GPS receiver, a gyroscope detector, and three video cameras. The valid data range of the



Fig. 3. Experimental mobile robot

LiDAR sensor was approximately 70 m from the robot. The proposed algorithms were implemented on a laptop with a 2.82 GHz Intel[®] Core[™]2 Quad CPU, a GeForce GTX 275 graphics card, and 4 GB RAM.

Fig. 4 (a) shows the result of terrain reconstruction by registering the sensed 3D point clouds and captured video images into the texture terrain mesh. Also, Fig.4 (b-c) show the results of the reconstructed terrain model captured from a bird view, a front view and a free view.

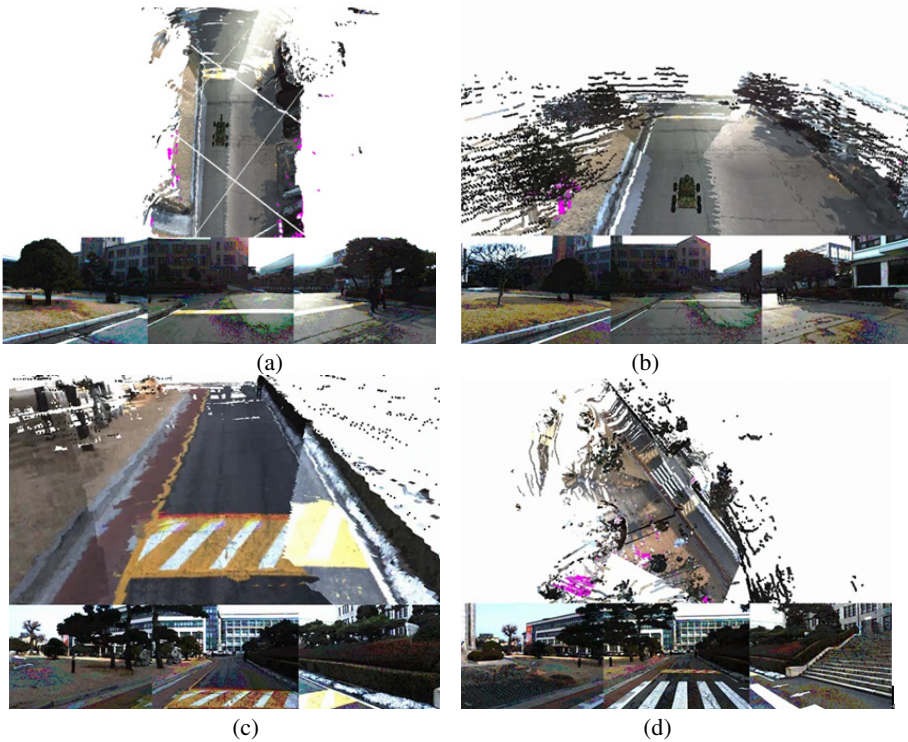


Fig. 4. Terrain reconstruction result. (a) A top view of the reconstructed ground mesh. (b) A bird view of the reconstructed terrain model. (c) A front view. (d) A free view.

The bottom images in Fig. 4 were captured by the three cameras. The robot captured three images of 659×493 RGB pixels every 0.1 s. In our application, the node size was $12.8 \times 12.8 \text{ m}^2$ and the cell size is $0.1 \times 0.1 \text{ m}^2$. In the TDB, the basic texture unit has 4×4 XRGB pixels. The resolution of the node texture was 512×512 pixels.

We compared the texture buffer sizes of the TDB and the video images, as shown in Fig. 5. After 20 s, 200×3 images were captured and 82 nodes were registered in the terrain model. The TDB buffer size of these nodes was 82.0 MB, generated from 557.7 MB of video images. Therefore, the results demonstrate that large-scale video images were registered to the TDB with a low memory overhead using the proposed TDB generation method.

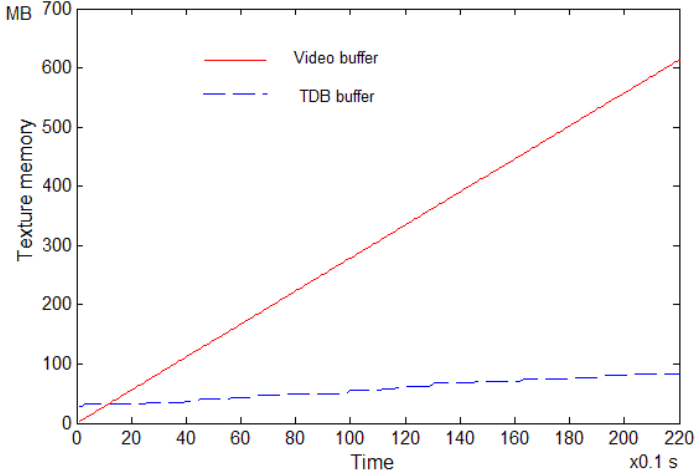


Fig. 5. Buffer sizes of the TDB and the images

In our projects, we rendered nine nodes of the terrain model surrounding the vehicle. Therefore, only 9.31 MB of the memory were used to represent the surrounding area of 147.4 m^2 . To speed up the computation of the TDB generation, we used GPU programming to implement the mapping process in parallel. The terrain modeling and visualization speed was improved to 18.85 fps on average.

5 Conclusions

In this study, we developed an intuitive terrain modeling technique with a TDB generation method for a mobile robot. We employed a GPU to map the triangles in the captured video images, projected from the terrain mesh, onto the TDB in real time. We tested our proposed system using a mobile robot mounted with integrated sensors. The results demonstrated an intuitive visualization performance and low memory requirement in a large-scale environment.

Acknowledgments. This work was supported by the Agency for Defense Development, South Korea.

References

1. Matsushita, Y., Miura, J.: On-line Road Boundary Modeling with Multiple Sensory Features, Flexible Road Model, and Particle Filter. *J. Robotics and Autonomous Systems* 59(5), 274–284 (2011)
2. Cohen-Or, D.: Exact Antialiasing of Textured Terrain Models. *Visual Computer* 13(4), 184–198 (1997)
3. Rovira-Más, F., Zhang, Q., Reid, J.F.: Stereo vision three-dimensional terrain maps for precision agriculture. *Computers and Electronics in Agriculture* 60(2), 133–143 (2008)
4. Sukumar, S.R., Yu, S.J., Page, D.L., Koschan, A.F., Abidi, M.A.: Multiple sensors Integration for Unmanned Terrain Modeling. In: *Proceedings of SPIE Unmanned Systems Technology VIII*, vol. 6230, pp. 65–74 (2006)
5. Song, W., Cho, K., Um, K., Won, C.S., Sim, S.: Intuitive Terrain Reconstruction Using Height Observation-Based Ground Segmentation and 3D Object Boundary Estimation. *Sensors* 2012 12(12), 17186–17207 (2012)
6. Grgic, S., Mrak, M., Grgic, M.: Comparison of JPEG Image Coders. In: *Proceedings of the 3rd International Symposium on Video Processing and Multimedia Communications*, pp. 79–85 (2001)
7. Mohammad, A., Alaa, A.: Querying Relational MPEG-7 Image Database with MPEG Query Format. *Journal of Multimedia and Ubiquitous Engineering* 6(4), 39–44 (2011)
8. Song, W., Cho, K.: GPU-Based Terrain Reconstruction System Using Grid-Based Flag Map for Mobile Operation. In: *The 2013 FTRA International Conference on Advanced IT, Engineering and Management*, pp. 113–114 (2013)

The RIP for Random Matrices with Complex Gaussian Entries

Kuo Xu, Jian Wang, and Byonghyo Shim

School of Electrical Engineering, Korea University
Anam-dong, Seongbuk-gu, Seoul, Korea, 136-713
xukuo2006@korea.ac.kr

Abstract. In this paper, we show that complex Gaussian random matrix satisfies the restricted isometric property (RIP) with overwhelming probability. We also show that for compressive sensing (CS) applications, complex Gaussian random matrix outperforms its real number equivalent in the sense that it requires fewer measurements for exact recovery of sparse signals. Numerical results confirm our analysis.

Keywords: Sparse recovery, restricted isometric property, complex Gaussian entries.

1 Introduction

In recent years, compressive sensing (CS) has attracted much attention in the academic world [1]. In CS, a high dimensional vector $x \in R^n$, which is k -sparse (i.e., $\|x\|_0 \leq k < n$), is sensed by a fat random matrix $A \in R^{m \times n}$ ($m < n$), yielding a low dimensional measurement vector

$$y = Ax. \quad (1)$$

Although (1) is an underdetermined system and has infinitely many solutions, the CS theory promises to achieve the perfect recovery of x by exploiting the sparsity. In analyzing the recovery performance, the restricted isometry property (RIP) for the sensing matrix A has been widely used. It has been a standard tool for studying how efficiently the measurement matrix A captures information about sparse signal. Letting A_T denote a submatrix of A with columns listed in set T , the matrix A is said to satisfy the k -RIP if there exists $\delta_k \in (0, 1)$ such that [2]

$$1 - \delta_k \leq \lambda(A'_T A_T) \leq 1 + \delta_k \quad (2)$$

for any T with cardinality $|T| \leq k$. In particular, δ_k is called isometry constant. It has been shown that many types of random matrices have excellent restricted isometry behavior. For example, a matrix $A \in R^{m \times n}$, which has i.i.d. entries with Gaussian distribution $N(0, 1/m)$, obeys the k -RIP with $\delta_k < \varepsilon$ with overwhelming probability if [3]

$$m = o\left(k \log \frac{n}{k} / \epsilon^2\right) \quad (3)$$

In many CS applications, a more general setting is that the sensing matrix $A \in \mathbb{C}^{m \times n}$ is a random matrix with complex Gaussian entries. There has been empirical evidence that CS works well under this setting, see, for instance, Shim [4]. However, difficulties remain in analyzing the theoretical recovery performance since little has been known about the RIP for random matrix with complex Gaussian entries.

In this work, we study the RIP for random complex Gaussian matrix. For simple description, we define $\text{GRM}(m, n, \sigma^2)$ as a class of $m \times n$ random matrices, for which the real part and imaginary part of entries form a set of $2mn$ i.i.d. random variables with distribution $N(0, \sigma^2/2)$. We argue that RIP holds for complex Gaussian Matrices.

Theorem 1. For any random matrix $A \in \text{GRM}(m, n, 1/m)$, we show that the matrix A satisfies the k -RIP with isometry constant

$$\delta_k < \alpha + 2\sqrt{\alpha} + \sqrt{\frac{4n}{m} H(\gamma)} \quad (4)$$

with overwhelming probability, where $\alpha = k/m$, $\gamma = k/n$ and H is the entropy function $H(\gamma) = -\gamma \log \gamma - (1-\gamma) \log (1-\gamma)$.

2 Proof of Theorem 1

This section is devoted to the proof of Theorem 1. We stress that the technique we used in the proof is similar to that in [2]. We first consider the eigenvalue of $A_T^* A_T$ where A_T is a submatrix of A with k columns, and then extend the result to all such submatrices.

Lemma 1. [Lemma 4 in [5]] For any random matrix $B \in \text{GRM}(m, n, 1)$, it satisfies

$$E[\text{Tr}(\exp(tB^* B))] \leq n \exp\left(\left(\sqrt{m} + \sqrt{n}\right)^2 t + (m+n)t^2\right) \quad \forall t \in \left[0, \frac{1}{2}\right], \quad (5)$$

And

$$E[\text{Tr}(\exp(-tB^* B))] \leq n \exp\left(-\left(\sqrt{m} - \sqrt{n}\right)^2 t + (m+n)t^2\right) \quad \forall t \in [0, \infty], \quad (6)$$

where $E[\cdot]$ represents expectation and $\text{Tr}[\cdot]$ is the trace.

Lemma 2. For any complex matrix B , all the eigenvalues of $\exp(tB^* B)$ satisfy

$$\lambda_i(\exp(tB^* B)) = \exp(t\lambda_i(B^* B)), \quad (7)$$

and are always positive.

Proof. Since B^*B is Hermitian and has full rank, its eigenvalues are all real numbers. According to the definition of exponential,

$$\exp(tB^*B) = I + tB^*B + \frac{1}{2!}(tB^*B)^2 + \cdots + \frac{1}{n!}(tB^*B)^n, \quad (8)$$

as $n \rightarrow \infty$. For a matrix Q consisted of the eigenvectors of B^*B , $\exp(tB^*B)$ can be diagonalized by Q .

$$Q^{-1} \exp(tB^*B) Q = I + t\Lambda + \frac{1}{2!}(t\Lambda)^2 + \cdots + \frac{1}{n!}(t\Lambda)^n. \quad (9)$$

By the definition of the exponential function, the i -th eigenvalue on the diagonal is

$$\lambda_i(\exp(tB^*B)) = \exp(t\lambda_i(B^*B)) > 0. \quad (10)$$

Lemma 3. For a matrix $A \in \text{GRM}(m, n, 1/m)$, singular value concentration inequalities of A_T satisfy

$$P\left(\lambda_{\max}(A_T^*A_T) \geq (\sqrt{\alpha} + 1)^2 + \varepsilon\right) \leq k \exp\left(-\frac{m\varepsilon^2}{4(\alpha + 1)}\right), \quad (11)$$

$$P\left(\lambda_{\min}(A_T^*A_T) \leq (\sqrt{\alpha} - 1)^2 - \varepsilon\right) \leq k \exp\left(-\frac{m\varepsilon^2}{4(\alpha + 1)}\right), \quad (12)$$

where A_T is a submatrix consisted of k columns randomly selected from A .

Proof. For each A_T out of A , A_T times sqrt m belongs to $\text{GRM}(m, n, 1)$. Let $\tau = mt$. Then from Lemma 1,

$$E[Tr(\exp(\tau A_T^*A_T))] \leq k \exp\left((\sqrt{\alpha} + 1)^2 \tau + m^{-1}(\alpha + 1)\tau^2\right), \quad (13)$$

for $\tau \in [0, m/2]$. Since all eigenvalues of $\exp(\tau A_T^*A_T)$ are positive (from Lemma 2),

$$Tr(\exp(\tau A_T^*A_T)) \geq \lambda_{\max}(\exp(\tau A_T^*A_T)) = \exp(\tau \lambda_{\max}(A_T^*A_T)). \quad (14)$$

For $t \in [0, m/2]$ and a small deviation ε , we get

$$\begin{aligned} & P\left(\lambda_{\max}(A_T^*A_T) \geq (\sqrt{\alpha} + 1)^2 + \varepsilon\right) \\ &= P\left(\exp\left(t\lambda_{\max}(A_T^*A_T) - t(\sqrt{\alpha} + 1)^2 - t\varepsilon\right) \geq 1\right) \end{aligned} \quad (15)$$

$$\leq E \left[\exp \left(t \lambda_{\max} (A_T^* A_T) - t (\sqrt{\alpha} + 1)^2 - t \varepsilon \right) \right] \quad (16)$$

$$\leq \exp \left(-t (\sqrt{\alpha} + 1)^2 - t \varepsilon \right) E \left[\text{Tr}(\exp(t A_T^* A_T)) \right] \quad (17)$$

$$\leq k \exp \left(-t \varepsilon + m^{-1} (\alpha + 1) t^2 \right), \quad (18)$$

where (16) uses the Markov's inequality. For a quadratic function, it is obvious that $f(t) = -t\varepsilon + m^{-1}(\alpha + 1)t^2$ attains the minimum at $t_0 = m\varepsilon / 2(\alpha + 1)$.

In a similar way, the lower bound in (12) can be proved.

Lemma 3 demonstrates the lower and upper bounds of $\lambda(A_T^* A_T)$ for some A_T . Note that the isometry constant $\delta_k \in (0, 1)$ is defined as the minimum constant such that for all $T \in \{1, \dots, n\}$ and $|T| = k$,

$$1 - \delta_k \leq \lambda(A_T^* A_T) \leq 1 + \delta_k. \quad (19)$$

For notational simplicity, denote $\lambda_{\max} = \lambda_{\max}(A_T^* A_T)$ and observe that

$$P \left(\forall_{A_T}, \lambda_{\max} \leq (\sqrt{\alpha} + 1)^2 + \varepsilon \right) \geq 1 - k \binom{n}{k} \exp \left(-\frac{m\varepsilon^2}{4(\alpha + 1)} \right). \quad (20)$$

From Stirling's approximation, we know the combination number k out of n approximates to $\exp(nH(\gamma))$. Then it follows that

$$P \left(\forall_{A_T}, \lambda_{\max} \leq (\sqrt{\alpha} + 1)^2 + \varepsilon \right) \geq 1 - k \exp \left(-\frac{m}{4\alpha} \left(\varepsilon^2 - \frac{4nH(\gamma)}{m} \right) \right). \quad (21)$$

As thus, for m goes to infinite,

$$\lambda_{\max} < (\alpha + 1)^2 + \sqrt{4nH(\gamma)/m}. \quad (22)$$

Similar results hold for λ_{\min} . The proof of Theorem 1 is established.

3 Simulation and Discussion

From Theorem 1, one can show that upper bound in (4) is more stringent than that in the real number situation [2]. Indeed, let A^r denote a $m \times n$ random matrix with real number entries satisfying $N(0, 1/m)$. Then for $n \rightarrow \infty$,

$$\delta_k(A) \rightarrow 2\sqrt{\alpha \log(n)} \quad \text{and} \quad \delta_k(A^r) \rightarrow 2\sqrt{2\alpha \log(n)}. \quad (23)$$

The isometry constant for A is 0.7 times as much as that for A^r and hence is more stringent. To see the difference between $\delta_k(A)$ and $\delta_k(A^r)$, we perform simulations to

provide an empirical comparison. To be specific, we generate a number of real and complex Gaussian random matrices. For each matrix, we calculate the isometry constant by an exhaustive search. The distributions of $\delta_k(A)$ and $\delta_k(A')$ are displayed in Fig. 1. One can easily observe that $\delta_k(A)$ is uniformly smaller than $\delta_k(A')$.

The reason why the complex Gaussian random matrix has more stringent isometry constant than the real Gaussian random matrix is perhaps that the extreme singular values of any submatrix formed by k (or fewer) columns from A has stronger concentration property. Indeed, the probability of violation for real Gaussian random matrix decreases at a speed of $\exp(-m\varepsilon^2/8)$ as ε increases [6], whereas, as shown in Lemma 3, the probability of violation for complex case decreases at $\exp(-m\varepsilon^2/4)$. In other words, the distribution of the extreme singular value of complex Gaussian matrix has a smaller tail, and therefore, the complex Gaussian matrix has a smaller δ_k (for the same γ and α), when compared to the real case. For compressive sensing applications, this result implies that fewer measurements are required [7].

To illustrate the advantage of complex Gaussian sensing matrix over the real case in compressive sensing, we perform simulations on sparse signal recovery with complex and real Gaussian random matrices. In our simulation, we employ orthogonal matching pursuit (OMP) algorithm as the recovery algorithm to recover k -sparse signals with complex entries. Two kinds of recovery are performed. First, we directly employ OMP to recover the complex signal x in the complex number signal model (1). Second, we reformulate model (1) to a real number signal model [8]:

$$y' = \begin{bmatrix} \text{Re}(y) \\ \text{Im}(y) \end{bmatrix}, x' = \begin{bmatrix} \text{Re}(x) \\ \text{Im}(x) \end{bmatrix}, A' = \begin{bmatrix} \text{Re}(A) & -\text{Im}(A) \\ \text{Im}(A) & \text{Re}(A) \end{bmatrix}, \quad (24)$$

and then employ OMP to recover x' . Note that the recovery of the second case is performed in the real domain. We compare the minimally required measurements y guaranteeing exact recovery of sparse signals.

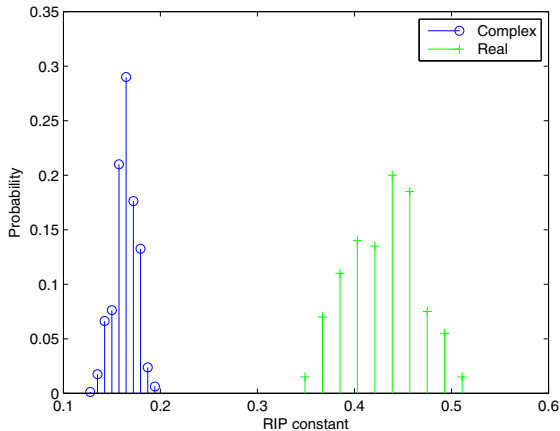


Fig. 1. Distribution of $\delta_k(A)$ and $\delta_k(A')$, with $k = 4$, $m = 20$, and $n = 128$

For a fixed sparsity ratio γ , $k = 10$, and $n = 256$, the exact recovery ratio by OMP algorithm is simulated with different measurement number m . Note that α and γ remain the same after reformulation. We calculate the exact recovery ratio for x and \hat{x} . The result is shown in Fig. 2. where cOMP represents the recovery result using the first method (i.e., direct recovery in the complex domain). It is easily observed that the first method outperforms the second method, as it uniformly requires fewer measurements for exact reconstruction.

It is interesting to note that the superior numerical performance of A over A' can also be explained as follows. For an $n \times 1$ k -sparse complex signal, its real number equivalent is an $2n \times 1$ $2k$ -sparse signal. In the recovery process, one sparse signal is selected with a candidates number of k out of n . Whereas, in the real equivalent case, candidates number $2k$ out of $2n$. By Stirling's approximation, we know

$$\binom{n}{k} \rightarrow e^{nH\left(\frac{k}{n}\right)} \quad \text{and} \quad \binom{2n}{2k} \rightarrow e^{2nH\left(\frac{k}{n}\right)}. \quad (25)$$

Thus it is easier to solve the complex sparse signal recovery problem than the reformulated real sparse signal recovery problem.

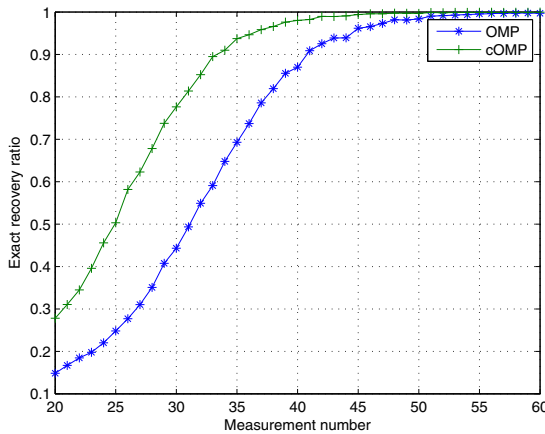


Fig. 2. Exact recovery of sparse signals via OMP for complex Gaussian measurement matrix and its real number equivalent

4 Conclusion

This paper presented the RIP for Gaussian random matrix with complex entries. The result demonstrated that compared to the isometry constant for real Gaussian random matrices, the isometry constant for complex Gaussian random matrices are more stringent. This implies that for CS applications, the required number of measurements guaranteeing exact recovery can be fewer when the complex Gaussian random measurements are used.

References

1. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inform. Theory* 52, 1289–1306 (2006)
2. Candes, E.J., et al.: Decoding by linear programming. *IEEE Trans. Inform. Theory* 51, 4203–4215 (2005)
3. Baraniuk, R., et al.: A simple proof of the restricted isometry property for random matrices. *Constructive Approximation* 28, 253–263 (2008)
4. Shim, B., et al.: Multiuser Detection via Compressive Sensing. *IEEE Communications Letters* 16, 972–974 (2012)
5. Haagerup, U., et al.: Random matrices with complex Gaussian entries. *Expositiones Mathematicae* 21, 293–337 (2003)
6. Ledoux, M.: The concentration of measure phenomenon. American Mathematical Soc. (2001)
7. Wang, J., et al.: On Recovery Limit of Orthogonal Matching Pursuit using Restricted Isometric Property. *IEEE Trans. Signal Process* 60, 4973–4976 (2012)
8. Shim, B., et al.: Sphere Decoding With a Probabilistic Tree Pruning. *IEEE Trans. Signal Process.* 56, 4867–4878 (2008)

Evolving Mobile App Recommender Systems: An Incremental Multi-objective Approach*

Xiao Xia, Xiaodong Wang, Xingming Zhou, and Bo Liu

School of Computer Science, National University of Defense Technology,
Changsha, P.R. China, 410073

Abstract. Existing recommender systems for mobile apps mainly focus on single objective which only reflects monotonous app needs of users. Therefore, we evolve the existing mobile app recommender systems leveraging the *multi-objective* approach. Moreover, to avoid risks introduced by dramatic system vibration, we realize the system evolution in an *incremental* manner. To achieve these two goals, we model the recommendation generation of the evolved system as a multi-objective optimization problem and propose a new *rank aggregation based evolving scheme* to gently evolve the systems. Furthermore, we propose a new recommending scheme for mobile apps based on Latent Semantic Analysis and leverage it to evolve the existing system. Real data evaluations have verified the effectiveness of our approach.

Keywords: Mobile app, multi-objective, incremental, rank aggregation.

1 Introduction

The tremendous increase in population of mobile apps has given birth to the challenge of app discovery. To meet this challenge, online markets have employed recommender systems to provide users with app suggestions. For instance, AppJoy [1] filters out app choices based on personalized app usage patterns. AppBrain [2] generates recommendations of the same category with those have been installed by users while AppAware [3] exploits the context information for app recommendations.

Such existing mobile app recommender systems (MARS) are of help to users for app discovery. However, they mainly focus on the recommendations of a single objective, which only reflects the monotonous app needs of users. Specifically, Appjoy utilizes focuses on the similarity among apps with respect to their usage patterns. AppBrain exploits the category of apps to capture their similarity. Systems such as the AppAware and others pay their attention to discover apps that are of similar using contexts. Therefore, most of the existing MARSs are advancing their recommendations by solely taking the app similarity into consideration.

On the other side, recent studies have recognized that single-objective systems may be of little use or even negative [4] while other aspects of recommendation quality are of similar important to the similarity [5,6]. Thus the multi-objective recommender

* This work is supported by the projects of National Natural Science Foundation of China: No. 61070201, No. 61170260 and No. 61202486.

systems are attracting increasing interests [7]. However, the study of multi-objective MARS is still missing in the literature. Therefore, we study the development of future MARS leveraging the multi-objective approach. Moreover, in the evolution of systems, severe system vibration may result in significant loss of customers. Therefore we utilize an incremental way to design the evolution for avoiding dramatic changes.

Main efforts and contributions of this paper are as follows:

- We propose a novel Latent Semantic Analysis (LSA) based scheme for mobile app recommendation, which overcome the user experience constraint.
- We model the recommendation generation of the evolved MARS as a multi-objective problem and propose a rank aggregation based evolving scheme, which realizes the incremental evolution of multi-objective MARS.
- Through real data evaluations, we verify the effectiveness and identify the potential of developing MARSS leveraging the incremental multi-objective approach.

2 LSA Based Recommending Scheme

Most online app markets generate app recommendations based on the behaviors of the users. For instance, the Google Play market provides users with apps that “users who installed this also installed”. Such a method may experience a cognitive constraint since users are not able explore even a majority of apps in a population over 700,000.

To conquer such limitations of user experiences, we propose the novel LSA based recommending scheme for mobile apps, which is also used to define the multi-objective optimization problem and to realize the incremental evolution. The scheme compares the app descriptions by using the LSA method thus to measure the similarity among apps. Based on the similarity measurements, it then recommends users with apps that are of the most similar to those they have accessed. This scheme inspires the recommender system to make better use of the global information of apps, i.e., the app descriptions. By this way, our scheme conquers the limitation of user experiences.

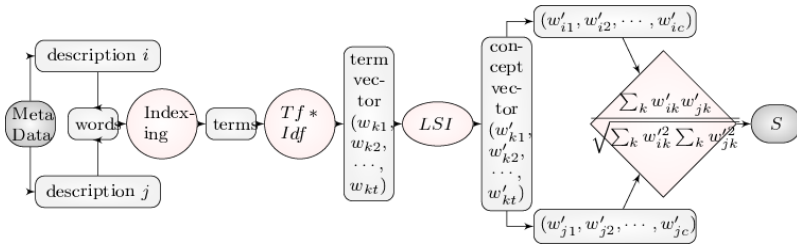


Fig. 1. Process of metadata similarity measurement using LSA

The process of applying LSA to measure the similarities among apps is illustrated in Figure 1. The LSA represents app descriptions by vectors of weighting terms. It then projects the term-description matrix to a lower-dimensional space, through which it mines the meanings and the variability of terms underlying the descriptions. After all, the term-description space is projected to the semantic space, which represents semantic concepts instead of raw terms. By this way, the similarity measurement which is based on the concepts comparison is expected to gain a better understanding.

3 Multi-objective Recommendation

To provide multi-objective app recommendations, we model the recommendation generation of the evolved system as a multi-objective problem in this section.

3.1 Objectives of Evolved System

We capture not only the needs of the users, but also the expectations of the developers and the online market. Therefore, we denote the following evolution objectives. The notations to be used are listed in Table 1.

Ranking. As users may want to find out and compare similar apps as those they have accessed, we define the objective ‘‘Ranking’’ to recommend the most similar apps to users. It is denoted as the average similarity between i and all its recommendations:

$$Ranking(i) = r_i^T * Lsa * e_i / N_R. \quad (1)$$

Range. As users also want to find novel apps while developers need to promote new apps, recommending similar apps alone is not sufficient. Therefore, we define the objective ‘‘Range’’ to recommend novel or even serendipitous apps. To define the *Range* objective, we capture the *category diversity* and *item diversity* of the recommendations. The former metric helps to improve the novelty and the scope of app discovery. The later avoids that the recommended apps are too similar to each other.

We define the *category diversity* based on both the number of categories and the proportion of apps of different categories:

$$D_c(i) = (C_d * (1 - e_i) / C_d * 1) * C_n * 1, \quad (2)$$

where $C_d = r_i^T * C$, which indicates how many apps of each category are recommended. The C_n is derived from the C_d , where $C_n(k) = 1$ if $C_d(k) \geq 1$ and $C_n(k) = 0$ otherwise. The C_n denotes the categories that the recommendations have covered. We define the *item diversity* as the average of the intra-list dissimilarity:

$$D_i(i) = 1 - r_i^T * Lsa * r_i / N_R (N_R - 1). \quad (3)$$

Therefore, we can derive the *Range* objective by:

$$Range(i) = D_c(i) * D_i(i) \quad (4)$$

Revenue. While the recommending services are provided by online markets, it is rational to cover the profit expectations of them when evolving the existing systems. Therefore, we define the objective ‘‘Revenue’’. To define the *Revenue* objective, we leverage the price and installations of apps to capture their profit potentials. That is,

$$Revenue(i) = lg(r_i^T * diag(P) * I + 1), \quad (5)$$

where the lg operation is introduced because the number of app installations varies across large scales.

Robustness. Since the preferences of users, developers and online markets vary over time, the recommender systems should be designed to be adaptive. To this end, we integrate the Robustness to our evolved system. to achieve better *Robustness* of the system, we define the category diversity parameter $\theta_c(i)$ and the price diversity parameter $\theta_p(i)$ to tune the performance of the system. They are defined to determine the upper bounds of recommended apps in different categories and those are not free.

Table 1. Notation definitions

<i>Notation</i>	<i>Definition</i>
A	the set of all apps
N_A	the size of A , i.e., the number of all apps
R_i	app recommendations for app i
N_R	the size of R_i , i.e., the number of recommended apps
r_i	$N_A \times 1$ vector, $r_i(k)=1$ if $k \in R_i$, else $r_i(k)=0$
Lsa	$N_A \times N_A$ matrix, $Lsa(i,j)$ is the similarity between i and j
C	$N_C \times N_C$ matrix, $C(i,j)=1$ if app i is in the category j , else $C(i,j)=0$
c_i	category of app i
P	$N_A \times 1$ vector, P_i is the price of app
I	$N_A \times 1$ vector, I_i is the installations of app i
e_i	$e(i)=1$, $e(j)=0$ for any j that $j \neq i$
l	$l(i)=I$ for all i

3.2 Problem Formulation

Based on the definitions above, we denote the R^3 metric, to measure the fitness of recommendations. Given i and the recommendation R_i for it:

$$R^3(R_i) = R^3(r_i) = \text{Ranking}(r_i)^{\delta_1} * \text{Range}(r_i)^{\delta_2} * \text{Revenue}(r_i)^{\delta_3}. \quad (6)$$

where the δ_x weights each kind of objectives so that the system can obtain better robustness. Based on the R^3 metric, we derive the objective of the evolving process to be the overall R^3 of all the apps $\sum_{i \in A} R^3(i)$. Furthermore, given the constraints of the limited space on web pages and the control parameters, we model the evolution process as a constrained optimization problem as follows, where P is the price matrix P and $P(i,j)=1$ denotes that app i has the price j .

$$\text{Max} \sum_{i \in A} R^3(r_i) \quad (7)$$

$$\text{s.t. } r_i^T * 1 = N_R, \quad (8)$$

$$r_i(i) \in \{0, 1\} \quad \forall i = 1, \dots, N_A, \quad (9)$$

$$r_i^T * C * (1 - e_{c_i}) / r_i^T * C * 1 \leq \theta_c(i), \quad (10)$$

$$r_i^T * P * (1 - e_{c_i}) / r_i^T * P * 1 \leq \theta_p(i). \quad (11)$$

4 Incremental Revolution

To achieve the incremental revolution, we introduce the method of *rank aggregation*, which is denoted as deriving a “consensus” ranking of the alternatives, given the diverse ranking preferences of various criteria. The rank aggregation has been applied in many areas, such as web search [8]. Furthermore, for the purpose of generating multi-objective recommendations, we design our evolving scheme following the optimization problem presented in Section 3.2. To be formal, the evolving scheme is defined as a problem of finding the rank aggregation method Ra , which satisfies:

$$R_o(i) = Ra(R_b(i), R_m(i)), \quad (12)$$

$$R^3(R_o(i)) \geq R^3(R_m(i)) \wedge R^3(R_o(i)) \geq R^3(R_b(i)), \quad (13)$$

where the $R_b(i)$ is the set of recommended apps provided by the Google Play market, the $R_m(i)$ is the set of apps recommended by our LSA based method and the $R_o(i)$ is the app recommendations generated by the evolved system.

There are $C_{|R_m|+|R_b|}^{|R_o|}$ recommendation candidates for each app, thus the global optimization could be computational expensive. We therefore propose a heuristic evolving scheme which is described in Algorithm 1. The basic idea of our heuristic scheme is to generate two ranks for further aggregation based on the sets R_b and R_m . We firstly weight them by the app similarity/dissimilarity, price and installations values. We then filter apps out to generate the R_o following the heuristic policy in the scheme.

Algorithm 1. The Evolving Recommending Scheme

```

Require: the number of recommended apps  $N_R$ 
For  $i$  in  $A$  do
  initialize  $k$  with 0, initialize  $R_o^k(i)$  with  $\emptyset$ 
  While  $k < N_R$  and  $R_b(i) \cup R_m(i) \neq \emptyset$  do
    find the app  $j$  in  $R_b \cup R_m$  which maximizes
       $Lsa(j, i) * D_i(R_o^k(i) + j) * P(j) * I(j)$ 
    If  $\{R_o^k(i), j\}$  satisfies the category and price diversity parameters then
      let  $R_o^{k+1}(i) = R_o^k(i) + j$ ,  $k = k + 1$ 
    End if
    delete  $j$  from  $R_b \cup R_m$ 
  End while
End for
Return  $R_o^k(i)$ 

```

5 Evaluation

To conduct the evaluations and verify the effectiveness our methods, we implement both the LSA based recommending scheme and the rank aggregation based evolving

scheme. We then compare the R^3 metrics of the three kinds of recommendations, i.e., the existing recommendations R_b , the LSA based recommendations R_m and the evolved recommendations R_o . For clear illustration, we normalize the values of all recommendations by that of the R_b , i.e., $R^3_{norm} = R^3(R_x(i) / R^3(R_b(i)))$. We further measure the similarity, the intra-list item diversity and the average profit of the three recommendations to better understand the incremental realization of the scheme.

Figure 2(1) shows that the evolving scheme shows off an advanced performance to achieve multi-objective recommendations, comparing to each single method. Moreover, from Figures 2(2), 2(3) and 2(4), we can see that the evolving scheme realizes the incremental evolution of recommender systems by conducting tradeoffs between the existing system and the new method, which avoids severe system vibration.

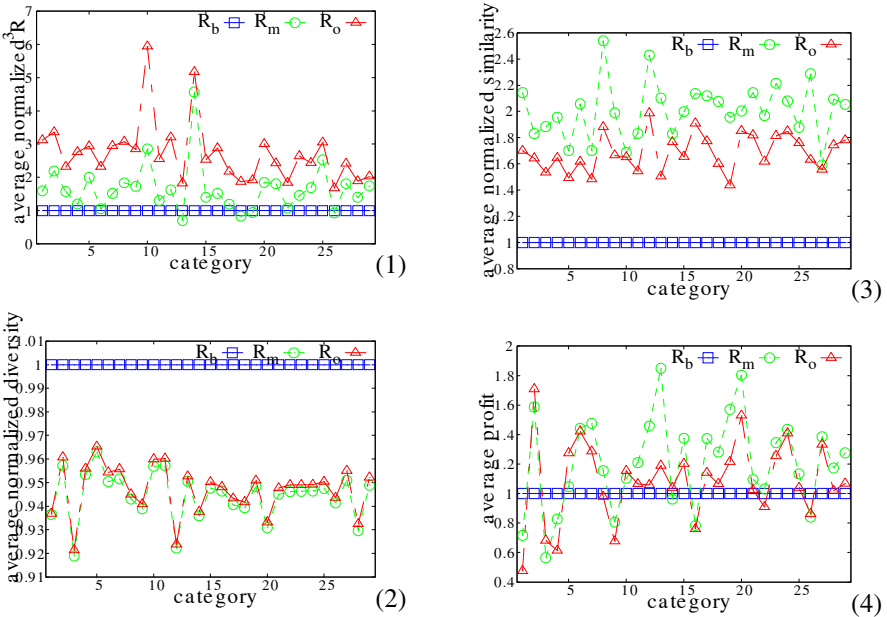


Fig. 2. The R^3 metric(1), intra-list diversity(2), similarity(3) and profit (4) of recommendations, illustrated by each category of apps

6 Conclusion

To evolve the MARSs, we propose a LSA based recommending method, model an optimization problem and design an evolving scheme for incremental evolution. By this way, we verify the effectiveness of the multi-objective and incremental approach.

References

1. Yan, B., Chen, G.: Appjoy: personalized mobile application discovery. In: MobiSys 2011, pp. 113–126. ACM, New York (2011)
2. Appbrain: Appbrain (April 15, 2013), <http://www.appbrain.com>

3. Girardello, A., Michahelles, F.: Appaware: which mobile applications are hot? In: *MobileHCI 2010*, pp. 431–434. ACM (2010)
4. McNee, S.M., Riedl, J., Konstan, J.: Accurate is not always good: How accuracy metrics have hurt recommender systems. In: *Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems, CHI 2006 (2006)*
5. Shi, Y., Zhao, X., Wang, J., Larson, M., Hanjalic, A.: Adaptive diversification of recommendation results via latent factor portfolio. In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2012*, pp. 175–184. ACM, New York (2012)
6. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Trans. on Knowl. and Data Eng.* 24(5), 896–911 (2012)
7. Rodriguez, M., Posse, C., Zhang, E.: Multiple objective optimization in recommender systems. In: *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys 2012*, pp. 11–18. ACM, New York (2012)
8. Li, L., Yang, Z., Kitsuregawa, M.: Aggregating user-centered rankings to improve web search. In: *AAAI 2007*, pp. 1884–1885. AAAI Press (2007)

Project Proposals Ranking

Sylvia Encheva

Polytec, Sørhauggate 128, 5527 Haugesund, Norway
sbe@hsh.no

Abstract. Well formulated project proposals are usually the ones being granted. At the same time the society is using a substantial amount of time and efforts to rank submitted proposals and thereafter decide on which one is to be granted. Projects team members along with their formal qualifications play a very important role considering a successful outcome. Unfortunately, their knowledge and abilities to exercise their skills in new settings are often not given serious consideration. As a result, project leaders have to find new people for completing specific tasks. This creates a lot of stress and delays since it is time consuming to find good specialists that can just step in a short notice and in addition it takes time for the new team members to learn what has been done and what is required next. Application of decision support systems can be quite helpful for avoiding problems caused by lack of skilful workers.

Keywords: Optimization, skills, learning.

1 Introduction

Well formulated project proposals are usually the ones being granted. At the same time the society is using a substantial amount of time and efforts to rank submitted proposals and thereafter decide on which one is to be granted. Projects team members along with their formal qualifications play a very important role considering a successful outcome. Unfortunately, their knowledge and abilities to exercise their skills in new settings are often not given serious consideration. As a result, project leaders have to find new people for completing specific tasks. This creates a lot of stress and delays since it is time consuming to find good specialists that can just step in a short notice and in addition it takes time for the new team members to learn what has been done and what is required next. Application of decision support systems can be quite helpful for avoiding problems caused by lack of skilful workers.

In order to exploit automated assistance one should be able to come up with a reasonably good description of what is available and what is needed. Such descriptions are provided by humans in a text form and are often open for interpretations. Therefore it is necessary to introduce mathematical methods that can provide reliable solutions for similar occurrences. The theory of vague sets is one way to handle the above described situations.

Vague sets are characterized by a truth-membership function and a false-membership function, [5] which allows further tuning of rules in a decision support system. Closure systems [1] deserve their place in decision support systems where new elements should be introduced in a structured way.

2 Preliminaries

Knowledge management has been of interest to many authors, see for example [8], [9], [11], and [12]. Fuzzy multi-criteria decision making is discussed in [2]. Vague sets have been exploited in relation to decision making problems in [3] and [10]. Information obtained via fuzzy equivalence relations can be used to determine a limit for the degree of precision in which inputs should be measured, [7].

Notations in this subsection are as in [6]. Let U be the universe of discourse, $U = \{u_1, u_2, \dots, u_n\}$, with a generic element of U denoted by u_i . A vague set A in U is characterized by a truth-membership function t_A and a false-membership function

$$f_A, t_A : U \rightarrow [0,1], f_A : U \rightarrow [0,1],$$

where $t_A(u_i)$ is a lower bound on the grade of membership of u_i derived from the evidence for u_i , $f_A(u_i)$ is a lower bound on the negation of u_i derived from the evidence against u_i , and $t_A(u_i) + f_A(u_i) \leq 1$.

The grade of membership of u_i in the vague set A is bounded to a subinterval $[t_A(u_i), 1 - f_A(u_i)]$ of $[0,1]$. The vague value $[t_A(u_i), 1 - f_A(u_i)]$ indicates that the exact grade of membership $\mu(u_i)$ of u_i may be unknown but it is bounded by

$$t_A(u_i) \leq \mu(u_i) \leq 1 - f_A(u_i),$$

where $t_A(u_i) + f_A(u_i) \leq 1$. When the universe of discourse U is continuous, a vague set A can be written as $A = \int_U [t_A(u_i), 1 - f_A(u_i)] / u_i$.

A lattice is a partially ordered set, closed under least upper and greatest lower bounds. The least upper bound of x and y is called the join of x and y , and is sometimes written as $x + y$; the greatest lower bound is called the meet and is sometimes written as $x \dot{y}$, [4]. X is a sublattice of Y if Y is a lattice, X is a subset of Y and X is a lattice with the same join and meet operations as Y . A lattice L is meet-distributive if for each $y \in L$, if $x \in L$ is the meet of (all the) elements covered by y , then the interval $[x; y]$ is a boolean algebra, [4].

3 Time Issues

New strategies and use of new analytical tools are just two examples illustrating the possibility for appearance of problems that need to be solved by people with different sets of skills than the one originally anticipated ones.

Below we begin with single skills hereafter referred to as *basic* skills and other skills that are combinations of two or more of the basic skills, hereafter referred to as complicated skills. The idea is to draw a clear picture of which skill is necessary to master first, in order to go on with more complicated ones and if an employ is believed to master a complicated one, is that person in a possession of all the required basic skills.

In a particular project where prerequisites are clear one can attach a parameter for addressing time necessary to master a new skill, provided sufficient level of proficiency of the rest of the needed skills. A simple way to calculate time for learning is to use the following function

$$T = \sum t_{a_i}$$

where T is the total time needed for a person to learn certain skills, t_{a_i} is the needed to learn a skill a_i , $1 \leq i \leq n$, where n is the number of skills needed for the task. A cost can be estimated by

$$T = \sum t_{a_i} \times c_{a_i}$$

where c_{a_i} is the amount of expenses needed to secure mastering of skill a_i by a person. This might make the model a bit more demanding but will facilitate the process of making fast decisions on optimal solutions like for example should one send team members to a course of involve new members who can just do the job.

If an employ posses only one skill (say a) and she is assigned to work on tasks requiring any of the other complicated skills (say ab , ac , ad , ae) additional time has to be anticipated for mastering for example one extra skill (say b) and mastering usage of the required complicated still ab .

Suppose three of the complicated skills are related to the supporting them basic skills. The rest of the five complicated skills are related to a single basic skill only. This can be interpreted as follows. Skills ab , ac and bc are mastered by the team members, for the rest of the five complicated skills some consideration have to be done. Extra time for the team to master the two basic skills or new team members possessing the required knowledge

Five basic skills are considered in Fig. 1. In this case only relations among complicated skills are shown where again the difference between two nodes connected by an edge is exactly one skill.

The process of mastering an additional skill and applying in combination with other skills is not necessarily linear in terms of time and efforts. Under a project planning one should have an approximate idea about the time needed to master some new skills. Difficulties might occur due to the nature of some particular skills, due to their increasing number as well as due to the necessity of applying several skills in combination.

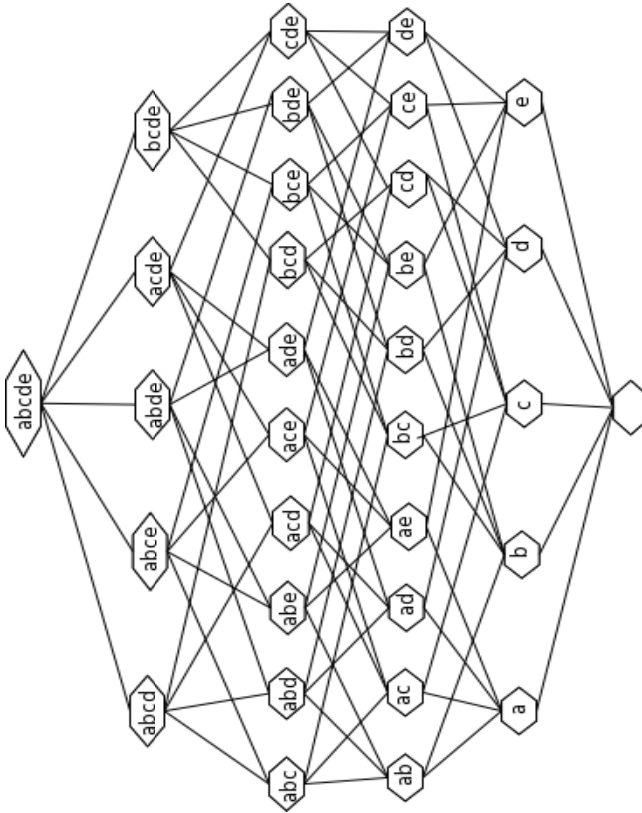


Fig. 1. Skills

If an employ possesses only one skill (say a) and she is assigned to work on tasks requiring any of the other complicated skills (say ab, ac, ad, ae) additional time has to be anticipated for mastering for example one extra skill (say b) and mastering usage of the required complicated still ab .

Suppose three of the complicated skills are related to the supporting them basic skills. The rest of the five complicated skills are related to a single basic skill only. This can be interpreted as follows. Skills ab, ac and bc are mastered by the team members, for the rest of the five complicated skills some consideration have to be done. Extra time for the team to master the two basic skills or new team members possessing the required knowledge.

Five basic skills are considered in Fig. 1. In this case only relations among complicated skills are shown where again the difference between two nodes connected by an edge is exactly one skill.

The process of mastering an additional skill and applying in combination with other skills is not necessarily linear in terms of time and efforts. Under a project planning one should have an approximate idea about the time needed to master some new skills. Difficulties might occur due to the nature of some particular skills, due to

their increasing number as well as due to the necessity of applying several skills in combination.

Vague set theory can be used to calculate amount of time and additional cost for gathering a team with predefined qualifications.

4 Conclusion

Today's projects require solving problems that might not be included in the initial proposals. This could be the case due to new requirements, introduction of new methods, technologies, and new team members. A project leader should be able to react fast without additional stress for her co-workers. Assistance of a decision support system will definitely speed up the process of finding who can do what it has not been planned, what that person has to learn, where and for how long time.

References

1. Adaricheva, K., Nation, J.B.: On implicational bases of closure systems with unique critical sets. In: International Symposium of Artificial Intelligence and Mathematics (ISAIM-2012), Ft. Lauderdale, FL, USA Results are included into plenary talk on conference Universal Algebra and Lattice Theory, Szeged, Hungary (June 2012)
2. Chang, T.H., Wang, T.C.: Using the fuzzy multi-criteria decision making approach for measuring the possibility of successful knowledge management. *Information Sciences* 179, 355–370 (2009)
3. Chen, S.M., Tan, J.M.: Handling multicriteria fuzzy decision-making problems based on vague set theory. *Fuzzy Sets and Systems* 67(2), 163–172 (1994)
4. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)
5. Gross, J.L., Yellen, J.: Handbook of Graph Theory. CRC Press INC. (2004)
6. Hong, D.H., Chang-Hwan Choi, C.-H.: Multicriteria fuzzy decision-making problems based on vague set theory. *Fuzzy Sets and Systems* 114, 103–113 (2000)
7. Klawonna, F., Castro, J.L.: Similarity in Fuzzy Reasoning. *Mathware & Soft Computing* 2, 197–228 (1995)
8. Percin, S.: Use of analytic network process in selecting knowledge management strategies. *Management Research Review* 33(5), 452–471 (2010)
9. Pourdarab, S., Nadali, A., Nosratabadi, H.E.: Determining the Knowledge Management Strategy Using Vague Set Group Decision. In: International Conference on Management and Artificial Intelligence, IPEDR, vol. 6, pp. 60–64 (2011)
10. Verma, M., Kumar, A., Singh, Y.: Vague modelling for risk and reliability analysis of compressor system. *Concurrent Engineering* 20, 177–184 (2012)
11. Ye, J.: Improved method of multicriteria fuzzy decision-making based on vague sets. *Computer-Aided Design* 39, 164–169 (2007)
12. Wu, W.W.: Choosing knowledge management strategies by using a combined ANP and DEMATEL approach. *Expert Systems with Applications* 35, 828–835 (2008)
13. Zhang, D., Zhang, J., Lai, K.K., Lu, Y.: An novel approach to supplier selection based on vague sets group decision. *Expert Systems with Applications* 36(5), 9557–9563 (2009)

A Stereo Micro Image Fusion Algorithm Based on Expectation-Maximization Technique

Cuixia Bai, Gangyi Jiang*, Mei Yu, Yigang Wang, Feng Shao, and Zongju Peng

Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
jianggangyi@126.com

Abstract. Due to the limitation of Depth Of Field (DOF) of microscope, the regions which are not within the DOF will be blurring after imaging. Thus for micro image fusion, the most important step is to identify the blurring regions within each micro image, so as to remove their undesirable impacts on the fused image. In this paper, a fusion algorithm based on an Expectation-Maximization (EM) technique is proposed for stereo micro image fusion. The local sharpness of stereo micro image is judged by EM technique, and then the sharpness regions are clustered completely. Finally, the stereo micro images are fused with pixel-wise fusion rules. The experimental results show that the proposed algorithm benefits from the novel region segmentation and it is able to obtain fused stereo micro image with higher sharpness compared with some popular image fusion method.

Keywords: Stereo Microscope, Image Processing, Image Fusion, Expectation-Maximization technique.

1 Introduction

With the development of digital signal processing, micro-images have widely been used in many applications such as materials, metallurgy, pharmacy, biology, chemistry, food, and so on[1]. But, the Depth of Field (DOF) of microscope is limited, and as a result, only the regions which are close to the focal plane can be seen clearly, while the other regions may be blurring. In order to obtain clear image, image fusion processing must be taken [2, 3]. Some image fusion techniques had been proposed and mainly used to make micro-image more informative or try to provide convenience for human in observing the stereo micro images.

Image fusion algorithms mainly include multi-resolution analysis [4, 5], wavelet transform [6-8] and other improved algorithms, which have their respective advantages and disadvantages for different specific images. For common images, traditional image fusion algorithms can be used to meet the requirements, but the micro image fusion for a particular task may require complementary fusion technique. For an efficient micro image fusion technique, the fused micro image is required not only to have the local contrast and global contrast, but also well combine the edge and contour information of source micro-images, so as to improve the details of micro image and its visual

* Corresponding author.

effect [6]. Burt presented a combined algorithm with an averaging and choosing in low-frequency components [7]. Nevertheless, in Burt's algorithm a fixed threshold is used, which is not good enough to be relative to the uncertainty of image. In order to overcome this disadvantage, a wavelet transform based image fusion algorithm was presented in Ref. [8]. After wavelet transforming of multi-focused images, the matching degree of images is computed as an adaptive threshold to decide whether the maximum selection or weighted average to be used. But this algorithm still cannot get better fusion results for texture regions in micro images.

This paper presents an EM (Expectation-Maximization) technique according to the features of the stereo micro image, and a new stereo micro image fusion algorithm is further proposed. The EM technique used for blurring region identification and the corresponding fusion algorithm for micro image are presented in Section 2. The experimental results in Section 3 show the effectiveness of the proposed algorithm, and a conclusion is given in Section 4.

2 The Proposed Stereo Micro Image Fusion Algorithm

For the stereo micro image fusion, the most important step is to identify the blurring regions within the monocular micro-image and to correspond exactly to the information of the binocular micro images, so as to remove their undesirable impacts on the fused image. The basic idea of the proposed stereo micro image fusion algorithm is to judge the local sharpness of micro-image by EM technique [9, 10], then the feature points are detected by the SIFT (Scale Invariant Feature Transform) [11, 12], and the feature matching is taken. Finally, the binocular micro images are fused with fusion rules. Fig. 1 shows the diagram of the stereo micro image fusion algorithm based on the EM technique.

2.1 The EM Clustering Based on the Definition Feature Values

To further provide a quantitative measurement on how blurring the image is, the key issue in this work is how to formulate the probability density function of the feature values distribution. The information of the definition feature values has been gained based on the TenenGrad function. The statistical model of the down-sampling image was shown in Fig. 2.

Fig. 2(a) is the monocular stereo micro image of papaya fruit. Fig. 2(b) is the statistical model of the definition feature values information. The definition information was gained by normalized feature values and count frequency within the range of these feature values. In view of the above fact, the distribution is proposed to be modeled using the following two-component Laplacian mixture model.

$$p(w) = \frac{\alpha}{\sigma_1 \sqrt{2}} e^{-\sqrt{2}|w|/\sigma_1} + \frac{1-\alpha}{\sigma_2 \sqrt{2}} e^{-\sqrt{2}|w|/\sigma_2} \quad (1)$$

where α is a mixing coefficient, each component yields a Laplacian distribution with a zero mean and standard deviations σ_1 and σ_2 , respectively; and furthermore, σ_2 is assumed to be larger than σ_1 .

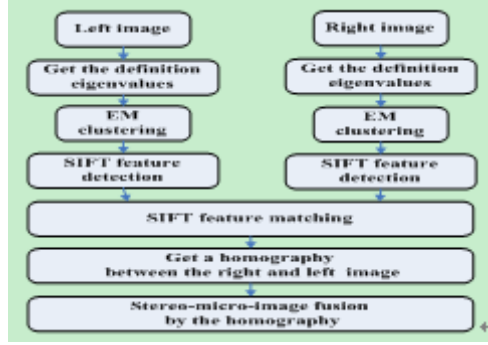
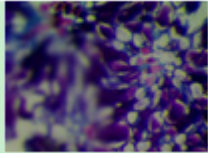
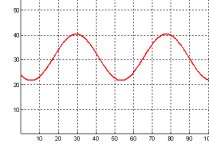


Fig. 1. The diagram of the stereo micro image fusion algorithm based on the EM technique



(a) The monocular stereo micro image of papaya fruit



(b) The statistical model

Fig. 2. The statistical model based on the stereo source micro image feature information

To evaluate the local information of input micro images, the proposed Laplacian mixture model is proposed to be locally adaptive; that is α , σ_1 and σ_2 in (1) are locally adaptive for each feature values (denoted as $w(k)$) to be $\alpha(k)$, $\sigma_1(k)$ and $\sigma_2(k)$. The EM technique [9] is exploited to conduct the maximum-likelihood estimation of these parameters iteratively, as follows. In each iteration, the estimation process is updated via two steps [10]: expectation step and maximization step, which are iterated until the convergence is reached.

2.2 Feature Point Detection and Matching by SIFT

The SIFT has the following steps in feature point detection [11]:

Step 1: The Gaussian kernel is build for the image scale space transform. The image scale space is denoted by $L(x, y, \sigma)$, which is gained by convoluting between the variable metric Gaussian function $G(x, y, \sigma)$ and the source image $I(x, y)$. In order to detect the stable key feature point in the scale space, this paper introduces the Difference of Gaussian function $D(x, y, \sigma)$. The relationship between them as shown in Equation (7):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (7)$$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Step 2: The local key points are posited precisely. Once the extreme points are detected, the following is to refine the feature points and to find extreme points in the image Gaussian pyramid each layer of the difference image.

Step 3: The directions are allocated based on the precise positioning of the local key points. This indicator can reflect the rigid rotation of the SIFT descriptor.

Step 4: A description symbol is set for each local key point, and the 128 dimensional SIFT feature vectors are represented for each key point. These mean that the description symbol can not be impacted by deformed geometry or other factors such as the size changes.

Step 5: The Euclidean distance is calculated in the local key points of the binocular micro images. The distance used to characterize the similarity evaluation criteria among the key points. If the ratio of the adjacent distance values is lower than a certain threshold, then they are a pair of matched key points.

2.3 Stereo Micro Image Fusion

The x point of the image demand for fusion according to the homographic matrix is mapped to the point x' of the reference image, which has four adjacent points x_1, x_2, x_3, x_4 . The most accurate point can determine by the similarity function. If the best matching point is x_4 , then the pixel value of x_4 assigns to the point x of the image demand for fusion, as shown in Fig. 3. The equation is the similarity function.

$$D(p, q) = \exp\left(-\frac{\Delta g_{pq}}{\sigma}\right) \quad (11)$$

where p and q are the pixels position, respectively. Δg_{pq} presents the differ of Euclidean distance between p and q . $D(p, q)$ means that the definition of similarity with the visual aggregation principle defined in the Gestalt psychology. σ is a constant and is closely related to the resolution of the image.

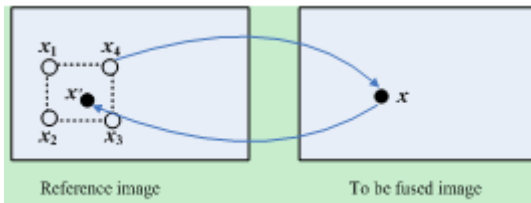


Fig. 3. The diagram of stereo micro-image fusion based on the homographic

3 Experimental Results and Discussions

In order to evaluate performance of proposed stereo micro image fusion method, the images of “Carrot Root” and “Papaya Fruit” with 1024×768 pixels are taken as the

tested stereo micro images shown in Fig. 4. They are showed that each of left images has some blurring regions. Figs. 4(a) are binocular micro images of Carrot Root, Figs. 4(b) are binocular micro images of Papaya Fruit. It is seen that the binocular micro images are related closely with each other.

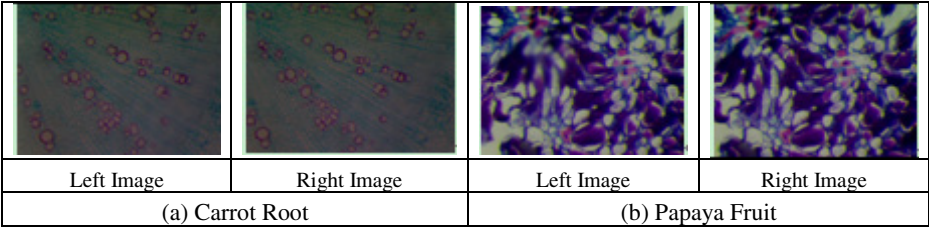


Fig. 4. The test stereo micro images

The proposed EM technology method is presented for identifying blurring regions within a micro image. In the experiments, the blurring region in micro image is identified by using TenenGrad evaluation function with block size of 129×129 or 127×127 , then, the sharpness regions are clustered completely. Finally, the stereo micro images are fused with pixel-wise fusion rules. The results were shown in fig. 5.

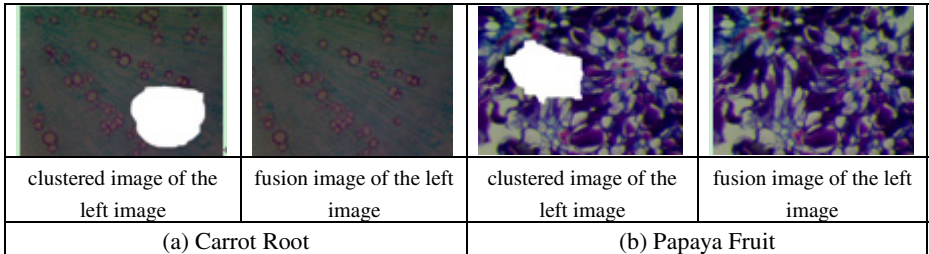


Fig. 5. The experimental results of the stereo micro image fusion algorithm

Figs. 5 are the clustered results obtained by the EM technology and the fusion image of the left image by the stereo micro image fusion algorithm. It is seen that the sharpness of the fused images obtained with the proposed algorithm is higher than that to be fused images. The high quality benefits from relatively more accurate region segmentation achieved by the EM clustering based blurring region identification. Except the visual analysis of these fused images, Table 1 gives some objective evaluation results of the fusion micro images and the source micro images. For these indicators, the larger the values are, the clearer the micro-image is.

Table 1 gives the objective sharpness evaluation results with respect to the source images and fused images obtained with the stereo micro image fusion algorithm. It is clear that whether evaluating with spatial domain or frequency domain the fusion micro images show its superiority compared with the source micro images.

Table 1. The objective evaluation results of the information and clarity comparative experiments

	Indicator	Variance	TenenGrad	SobelGrad	Hadamard	DCT
Carrot Root	To be fused of the left image	3.7399	1.8438	1.1328	1.3447	6.1017
	fusion image of the left image	3.8852	2.0137	1.2263	1.3819	6.2935
Papaya Fruit	To be fused of the left image	1.1978	1.1153	2.5544	3.1471	1.2345
	fusion image of the left image	1.2274	1.2025	2.6625	3.3601	1.3425

4 Conclusion

The captured micro images may have clear regions as well as blurring regions due to the depth of field limit, which significantly decrease the image quality. Image fusion is helpful to solve this problem. In this paper, a fusion algorithm based on the EM technology is proposed for stereo micro image. With the help of EM technology, the blurring regions in monocular micro image can be accurately clustered, so that the blurring regions will be excluded in the fusion. The proposed algorithm is able to retrieve the details existing in the binocular micro images and fuse them as an image with high sharpness. Experimental results show the effectiveness of the proposed algorithm. Future works may focus on more effective fusion rules so as to improve the performance of fusion greatly in stereo micro image.

Acknowledgments. This work was supported by Natural Science Foundation of China (61071120, 61111140392, 61271270), Natural Science Foundation of Ningbo (2012A610045), and Scientific Research Foundation of Ningbo University (XYL12001).

References

- [1] Li, H.F., Chai, Y., Li, Z.F.: Multi-focus image fusion based on nonsubsampling contourlet transform and focused regions detection. *Optik-International Journal for Light and Electron Optics* 124, 40–51 (2013)
- [2] Du, P.J., Liu, S.C., Xia, J.S., et al.: Information fusion techniques for change detection from multi-temporal remote sensing images. *Information Fusion* 14, 19–27 (2013)
- [3] Khaleghi, B., Khamis, A., Karray, F.O., et al.: Multisensor data fusion: A review of the state-of-the-art. *Information Fusion* 14, 28–44 (2013)
- [4] Chai, Y., Li, H.F., Zhang, X.Y.: Multifocus image fusion based on features contrast of multiscale products in nonsubsampling contourlet transform domain. *Optik-International Journal for Light and Electron Optics* 123, 569–581 (2012)
- [5] Bai, C.X., Jiang, G.Y., Yu, M., et al.: A micro-image fusion algorithm based on region growing. *Journal of Electronics (China)* 30, 91–96 (2013)
- [6] Tian, J., Chen, L.: Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure. *Signal Processing* 92, 2137–2146 (2012)

- [7] Burt, P.J., Kolczynski, R.J.: Enhanced image capture through fusion. In: The Fourth International Conference on Computer Vision, Berlin, Germany, May 11-14, pp. 173–182 (1993)
- [8] Zhou, T., Hu, B.J.: Adaptive algorithm of multi-focused image fusion based on wavelet transform. *Chinese Journal of Sensors and Actuators* 23, 1272–1276 (2010)
- [9] Dempster, A.P., Laird, N., Rubin, D.B.: Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B.* 39, 1–38 (1977)
- [10] Wu, C.F.J.: On the convergence properties of the EM algorithm. *The Annals of Statistics* 11, 95–103 (1983)
- [11] Lowe, D.G.: Distinctive image features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
- [12] Lindeberg, T.: Scale invariant feature transform. *Scholarpedia* 7, 10491 (2012)

Moldable Job Scheduling for HPC as a Service

Kuo-Chan Huang¹, Tse-Chi Huang¹, Mu-Jung Tsai¹, and Hsi-Ya Chang²

¹ Department of Computer Science
National Taichung University of Education
No. 140, Min-Shen Road, Taichung, Taiwan
kchuang@mail.ntcu.edu.tw,
{rogevious, amy29605}@gmail.com

² National Center for High-Performance Computing
No. 7, R&D 6th Rd., Hsinchu Science Park, Hsinchu, Taiwan
9203117@nchc.narl.org.tw

Abstract. As cloud computing emerges and gains acceptance, more and more software applications of various domains are transforming into the SaaS model. Recently, the concept of HPC as a Service (HPCaaS) was proposed to bring the traditional high performance computing field into the era of cloud computing. One of its goals aims to allow users to get easier access to HPC facilities and applications. This paper deals with related job submission and scheduling issues to achieve such goal. Traditional HPC users in supercomputing centers are required to specify the amount of processors to use upon job submission. However, we think this requirement might not be necessary for HPCaaS users since most modern parallel jobs are moldable and they usually could not know how to choose an appropriate amount of processors to allow their jobs to finish earlier. Therefore, we propose a moldable job scheduling approach which relieves HPC users' burden of selecting an appropriate number of processors and can achieve even better system performance than existing job scheduling methods. The experimental results indicate that our approach can achieve up to 75% performance improvement than the traditional rigid processor allocation method and 3% improvement than previous moldable job scheduling methods.

Keywords: moldable job, HPC as a Service, processor allocation.

1 Introduction

High performance computing (HPC) has long been a very important field for solving large-scale and complex scientific and engineering problems. However, accessing and running applications on HPC systems remains tedious, limiting wider adoption and user population [1]. As cloud computing emerges, which emphasizes easier and efficient access to IT infrastructure, recently the concept of *HPC as a Service* [1] was proposed to transform HPC facilities and applications into a more convenient and accessible service model.

Traditional HPC users at supercomputing centers are required to specify an amount of processors to use upon job submission. This requirement might be reasonable in

earlier days for the following two reasons. Firstly, some parallel jobs might be rigid jobs [14] which can only be executed with a specific amount of processors. Secondly, developers of parallel programs want to conduct performance benchmarking, e.g. drawing the speedup curve. However, the situation has changed. Most modern parallel applications are moldable [14] and written in a way allowing them to run with different number of processors as required, such as MPI [17] parallel programs. Moreover, most end users just want to get their jobs done faster, but don't care and even don't know how many processors is the best amount to use. Therefore, it seems that it is no longer necessary to require users to specify the amount of processors to use when they submit parallel jobs, especially for the end users of HPC applications as a Service.

Information about parallel program behavior is crucial for job schedulers to automatically choose effective amounts of processors for applications. In this paper, we consider two commonly used parallel speedup models: Amdahl's law [15] and Downey's speedup model [6][7], which have been shown capable of representing many applications' parallel behavior effectively. Based on these two parallel speedup models, we developed an effective moldable job scheduling approach to relieving HPC users' burden of selecting an appropriate number of processors upon job submission. A series of simulation experiments were conducted for performance evaluation. The experimental results show that in addition to relieving users' burden our approach can achieve even better system performance than existing job scheduling methods, up to 75% performance improvement than the traditional rigid processor allocation method and 3% improvement than previous moldable methods.

2 Related Work

Parallel job scheduling and allocation has long been an important research topic [3][4][13]. For rigid jobs [14], backfilling job scheduling approaches have been proposed to improve system performance [2][5]. For moldable jobs [14], previous research [11] has shown potential performance improvement achieved by adaptive processor allocation. The proposed adaptive processor allocation methods in [11] dynamically determine the number of processors to allocate just before job execution according to the amount of current available resources and job queue information.

In [8][9], Srinivasan *et al.* proposed a schedule-time aggressive fair-share strategy for moldable jobs, which adopts a profile-based allocation scheme. This strategy thus needs to have the knowledge of job execution time. On the other hand, our approach does not require the information of job execution time. Sun *et al.* proposed an adaptive scheduling approach for malleable jobs with periodic processor reallocations based on parallelism feedback of the jobs and allocation policy of the system in [10].

In [1], AbdelBaky *et al.* proposed the concept of HPC as a Service, aiming to transform traditional HPC resources into a more convenient and accessible service. They focused on the issues related to elastic provisioning and dynamic scalability, which are concerned in malleable jobs [14]. In this paper, we take advantage of the moldable property [14] in most modern parallel applications to develop an effective

moldable job scheduling approach for HPCaaS, aiming to relieve users' burden of specifying appropriate numbers of processors and improve overall system performance.

3 Processor Allocation for Moldable Job Scheduling

This section deals with the issues on processor allocation for moldable job scheduling. The job scheduler has to make processor allocation decisions on two kinds of events: *job arrival* and *job finish*. In general, there are two possible philosophies: running as many jobs in queue simultaneously as possible or giving the first job as many processors as possible. We call these two philosophies *parallel policy* and *serial policy*, respectively, in this paper. Which policy is better would largely depend on the parallel behavior of applications.

In the following, we explore the potential of the two policies on three common parallel speedup models which cover the behavior of most parallel applications. The first is the model usually introduced in the textbook of parallel processing, where *speedup* is defined by $S_p = T_l/T_p$, with p the number of processors, T_l the execution time of the sequential run, T_p the execution time of parallel processing with p processors. Based on the definition of speedup, *efficiency* is another performance metric defined as $E_p = S_p/p = T_l/pT_p$. Efficiency is a value, typically between zero and one, estimating how well-utilized the processors are in solving the problem. The second model is Amdahl's law [15], which states that if P is the proportion of a program that can be made parallel, then the maximum speedup that can be achieved by using N processors is $S(N) = 1 / ((1-P) + P/N)$. The third is Downey's speedup model of parallel programs, which has been shown capable of representing the parallelism and speedup characteristics of many real parallel applications [6][7]. Downey's model is a non-linear function of two parameters. The first parameter σ (*sigma*) is an approximation of the coefficient of variance in parallelism within the job. It determines how close to linear the speedup is. A value of zero indicates linear speedup and higher values indicate greater deviation from the linear curve. Another parameter is A , denoting the average parallelism of a job and is a measure of the maximum speedup that the job can achieve.

Based on the speedup models, the resultant average turnaround time of the two allocation policies can be derived. For example, the following two equations represent the average turnaround time achieved by the parallel and serial allocation policies, respectively, for applications of the Amdahl's law model, where t is the job's sequential runtime, x is the parallel proportion between 0 and 1, n is the number of free processors, and d is the number of jobs in queue, assuming n to be a multiple of d .

$$\text{Average turnaround time}_{\text{parallel}} = t \cdot \left((1-x) + \frac{x}{\frac{n}{d}} \right) \cdot d \cdot \frac{1}{d} \quad \text{Average turnaround time}_{\text{serial}} = t \cdot \left((1-x) + \frac{x}{n} \right) \cdot (1+d) \cdot d \cdot \frac{1}{d}$$

Figures 1 to 4 compare the performance of parallel and serial allocation policies, in terms of average turnaround time, on different application speedup models. The comparison indicates that job scheduler has to adopt different processor allocation policies for applications of different speedup models. For example, the serial allocation policy is superior for applications of the first model. Based on this analysis, we developed a moldable job scheduling approach for HPC as a Service, which can automatically determine the amount of processors to use for HPC users and would not only relieve users' burden of specifying appropriate numbers of processors but also achieve even better system performance than existing job scheduling methods. The proposed approach will be evaluated in the following section.

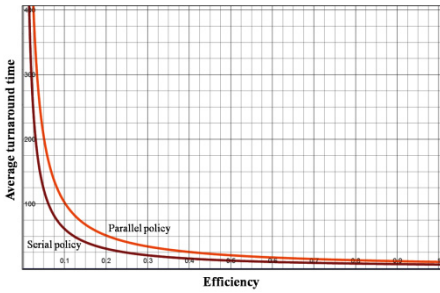


Fig. 1. The first model (Efficiency)

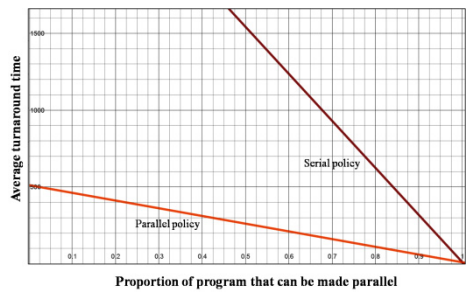


Fig. 2. The second model (Amdahl's law)

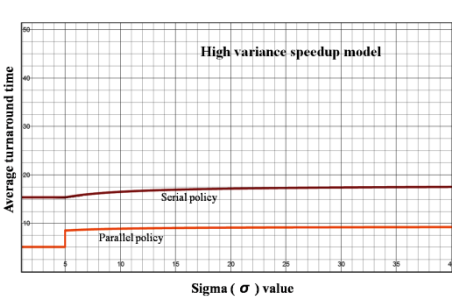


Fig. 3. Downey's high-variance model

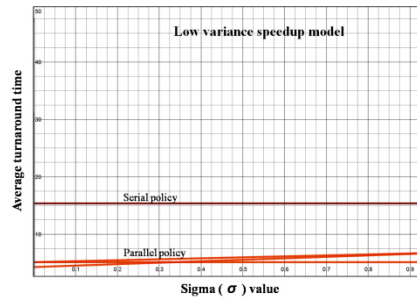


Fig. 4. Downey's low-variance model

4 Experiments and Performance Evaluation

This section evaluates the proposed approach and compares it with four other methods: rigid, adaptive scaling up and down protected [16], restricted scaling up and down protected [16], and random. The *rigid* method is commonly used in most current HPC systems, which can only allocate a fixed amount of processors, specified by the user, to a job. The two scaling up and down allocation methods are previous moldable job scheduling approaches shown to achieve good performance [16]. The random approach is a simple policy for the job scheduler to perform automatic processor amount determination, randomly choosing the amount. The performance

evaluation was conducted through a series of simulation experiments, assuming a 128-processor cluster, based on a public workload log on SDSC's SP2 [12]. The two parameters, σ and A , for Downey's speedup models were generated randomly.

Figures 5 and 6 show the experimental results based on the Downey's low variance model and Amdahl's law, respectively. The results indicate that our approach achieve the best overall performance, up to 75% performance improvement than the traditional rigid method and 3% improvement than previous moldable methods.

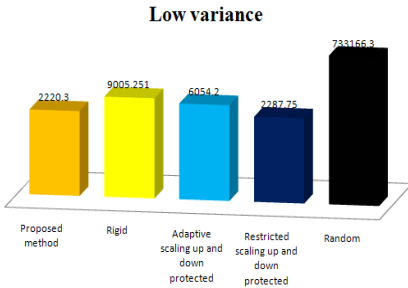


Fig. 5. Downey's low variance speedup model

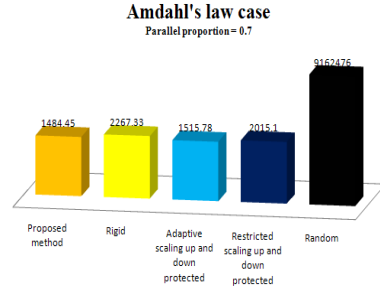


Fig. 6. Amdahl's law model

5 Conclusions

HPC as a Service is a future trend for high-performance computing, aiming to provide a more convenient and accessible HPC resources and applications. To achieve that goal, one potential issue to resolve is relieving users' burden of choosing an appropriate amount of processors to use upon job submission, when the submitted jobs have the moldable property which is common in most modern parallel programs. This paper proposes a moldable job scheduling approach for HPC as a Service, which not only relieves users' burden but also achieves even better system performance than existing methods, up to 75% performance improvement than the traditional rigid method and 3% improvement than previous moldable methods.

References

1. AbdelBaky, M., Parashar, M., Kim, H., JordanKirk, E.J., Sachdeva, V., Sexton, J., Jamjoom, H., Shae, Z.Y., Pencheva, G., Tavakoli, R., Wheeler, M.F.: Enabling High Performance Computing as a Service. *IEEE Computer* 45, 72–80 (2012)
2. Feitelson, D.G., Weil, A.M.: Utilization and Predictability in Scheduling the IBM SP2 with Backfilling. In: 12th Int'l Parallel Processing Symp., pp. 542–546 (April 1998)
3. Gibbons, R.: A Historical Application Profiler for Use by Parallel Schedulers. In: Feitelson, D.G., Rudolph, L. (eds.) *IPPS-WS 1997 and JSSPP 1997*. LNCS, vol. 1291, pp. 58–77. Springer, Heidelberg (1997)

4. Lifka, D.: The ANL/IBM SP Scheduling System. In: Feitelson, D.G., Rudolph, L. (eds.) IPPS-WS 1995 and JSSPP 1995. LNCS, vol. 949, pp. 295–303. Springer, Heidelberg (1995)
5. Mu'alem, A.W., Feitelson, D.G.: Utilization, Predictability, Workloads, and User Runtime Estimate in Scheduling the IBM SP2 with Backfilling. *IEEE Transactions on Parallel and Distributed Systems* 12(6), 529–543 (2001)
6. Downey, A.B.: A Model for Speedup of Parallel Programs. UC Berkeley EECS Technical Report, No. UCB/CSD-97-933 (January 1997)
7. Downey, A.B.: A Parallel Workload Model and Its Implications for Processor Allocation. In: *The 6th International Symposium on High Performance Distributed Computing* (1997)
8. Srinivasan, S., Krishnamoorthy, S., Sadayappan, P.: A Robust Scheduling Strategy for Moldable Scheduling of Parallel Jobs. In: *5th IEEE International Conference on Cluster Computing*, pp. 92–99 (2003)
9. Srinivasan, S., Subramani, V., Kettimuthu, R., Holenarsipur, P., Sadayappan, P.: Effective Selection of Partition Sizes for Moldable Scheduling of Parallel Jobs. In: Sahni, S.K., Prasanna, V.K., Shukla, U. (eds.) *HiPC 2002*. LNCS, vol. 2552, pp. 174–183. Springer, Heidelberg (2002)
10. Sun, H., Cao, Y., Hsu, W.J.: Efficient Adaptive Scheduling of Multiprocessors with Stable Parallelism Feedback. *IEEE Transactions on Parallel and Distributed System* 22(4) (April 2011)
11. Huang, K.C.: Performance Evaluation of Adaptive Processor Allocation Policies for Moldable Parallel Batch Jobs. In: *3th Workshop on Grid Technologies and Applications* (2006)
12. Parallel Workloads Archive,
<http://www.cs.huji.ac.il/labs/parallel/workload/>
13. Feitelson, D.G.: A Survey of Scheduling in Multiprogrammed Parallel Systems, Research Report RC 19790 (87657), IBM T. J. Watson Research Center (October 1994)
14. Feitelson, D.G., Rudolph, L., Schweigelshohn, U., Sevcik, K., Wong, P.: Theory and Practice in Parallel Job Scheduling. In: Feitelson, D.G., Rudolph, L. (eds.) *IPPS-WS 1997 and JSSPP 1997*. LNCS, vol. 1291, pp. 1–34. Springer, Heidelberg (1997)
15. Kleinrock, L., Huang, J.H.: On parallel processing systems: Amdahl's law generalized and some results on optimal design. *IEEE Trans. Softw. Eng.* 18(5) (1992)
16. Huang, K.C., Huang, T.C., Tung, T.H., Shih, P.Z.: Effective Processor Allocation for Moldable Jobs with Application Speedup Model. In: *Proceedings of the International Computer Symposium, ICS 2012, Taiwan* (2012)
17. The Message Passing Interface (MPI) standard,
<http://www.mcs.anl.gov/research/projects/mpi/>

MapReduce Example with HBase for Association Rule

Jongwook Woo¹ and Kilhung Lee^{2,*}

¹ Computer Information Systems Department
California State University, Los Angeles, CA, USA
jwoo5@calstatela.edu

² Department of Computer Science and Engineering
Seoul National University of Science and Technology, Seoul, Korea
khlee@seoultech.ac.kr

Abstract. The paper illustrates how to store and compute association sets of Big Transaction Data using Hadoop and HBase and then, shows the experimental result of a MapReduce algorithm using HBase to find out association in transaction data, which is a Market Basket Analysis algorithm of Association Rule in Business Intelligence. The algorithm sorts and converts the transaction data of HBase to data set with (key, value) pair, and stores the associated data to the HBase. The algorithm and HBase run on Amazon EC2 service using Apache Whirr. The experimental results show that the algorithm increases the performance as adding more nodes till a certain number of transaction data. However, it loses control and connection when there are too many IOs with more than 3.5 millions of transaction data in HBase.

Keywords: HBase, NoSQL DB, MapReduce, Market Basket Analysis, Hadoop.

1 Introduction

Data gets bigger and reaches tera- and peta-bytes as the web, smart phone, social media, bioinformatics, and sensor networks have been generating data. Furthermore, the data generated is non- or semi-structured. This large scaled and unstructured data is called *Big Data*, which makes it more difficult to store and compute data using the legacy systems. *Google* faced the issue when collecting data from the millions of web sites to keep them to the existing file systems and Relational Database Management Systems (RDBMS), which could not store and handle the data efficiently. Thus, *Google* implemented *Google File Systems (GFS)*, *BigTable*, and *MapReduce* parallel computing platform, which *Apache Hadoop*, *HDFS*, and *HBase* projects are motivated from. *Hadoop* is the parallel programming platform built on *HDFS* using *MapReduce* functions, which is called data intensive computing by moving processes to data as (key, value) pairs. *HBase* is one of *NoSQL DBs* and runs on *HDFS* with *Hadoop* to store and process big data. *HBase* and *Hadoop* have been adopted

* Corresponding author.

dramatically for large scale data, which is not easy to store and compute. *Apache Whirr* helps to generate both instances easily on *Amazon Elastic Compute Cloud (EC2)*¹.

This paper shows that the legacy sequential algorithms can be redesigned or converted to MapReduce algorithms. Besides, un-/semi-structured large scale data can be conveniently stored to and used on *column-oriented HBase*. Then, a MapReduce Market Basket Analysis algorithm is executed on *Hadoop* and *HBase* on *Amazon Web Service (AWS) EC2* with its performance experimental result.

In the paper, section 2 presents related work. Section 3 describes Hadoop MapReduce. Section 4 illustrates *MBA* MapReduce algorithm as well as *Apriori*-MapReduce algorithm for Market Basket Analysis. Section 5 describes HBase and proposed table schema for the algorithm. Section 6 shows the experimental result. Finally, section 7 is conclusion.

2 Related Work

Woo et al proposed Market Basket Analysis algorithm that runs on Hadoop MapReduce on HBase and HDFS [1, 2]. The papers are to compare the performance in HBase and HDFS in large scale data, which does not clearly show the experimental result. Woo also presented *Apriori-MapReduce* algorithm in thepy [3].

3 MapReduce in Hadoop

MapReduce is a functional programming method used in Artificial Intelligence. It has been emerging by Google and *Apache Hadoop* project to analyze large scale data set in distributed computing environment. It is composed of two functions to specify, “Map” and “Reduce with data structured in (*key*, *value*) pairs.

Hadoop was inspired by Google’s MapReduce and GFS [10] and has been used by a global community of contributors such as Yahoo, Facebook, and Twitters and becomes more popular to the enterprise computing as the inexpensive commodity platforms can be used together to compose *Hadoop* cluster. In map function, the master node splits the input into smaller sub-data ($k1, v1$) and generates $\langle k2, v2 \rangle$ where k and v are respectively key and value and $\langle \rangle$ represents list or set. In reduce node, reduce function takes inputs ($k2, \langle v2 \rangle$) from map nodes and generates $\langle k3, v3 \rangle$.

4 Market Basket Analysis Algorithm

4.1 Data Structure and Conversion

Association Rule or Affinity Analysis is the fundamental data mining analysis to find the co-occurrence relationships as purchase behavior of customers. Market Basket

¹ This work is supported by Amazon AWS Education Research Grant.

Analysis is one of Affinity Analyses to analyze the association of data set. With the associated item sets, store owners list a pair of items to control the stocks more intelligently, to arrange items on shelves and to promote items together.

Transaction 1: cracker, icecream, beer
 Transaction 2: chicken, pizza, coke, bread
 Transaction 3: baguette, soda, hering, cracker, beer
 Transaction 4: bourbon, coke, turkey

Fig. 1. Transaction data example at a store

The data in Figure 1 is a list of transactions with its transaction number and the list of products. For MapReduce operation, the data set is structured with (*key*, *value*) pairs.

4.2 The MBA Algorithm

- 1: Reads each transaction of input file and generates the data set of the items: $\langle V_1 \rangle, \langle V_2 \rangle, \dots, \langle V_n \rangle$ where $\langle V_n \rangle: (v_{n1}, v_{n2}, \dots, v_{nm})$
- 2: Sort all data set $\langle V_n \rangle$ and generates sorted data set $\langle U_n \rangle$: $\langle U_1 \rangle, \langle U_2 \rangle, \dots, \langle U_n \rangle$ where $\langle U_n \rangle: (u_{n1}, u_{n2}, \dots, u_{nm})$
- 3: *While Loop* $\langle U_n \rangle$ has the next element;
note: each list U_n is handled individually
 - 3.1: *For Loop* each item from u_{n1} to u_{nm} of $\langle U_n \rangle$ with NUM_OF_PAIRS
 - 3.a: generate the data set $\langle Y_n \rangle: (y_{n1}, y_{n2}, \dots, y_{ni})$; $y_{ni}: (u_{nx}, u_{ny})$ is the list of self-crossed pairs of $(u_{n1}, u_{n2}, \dots, u_{nm})$ where $u_{nx} \neq u_{ny}$
 - 3.b: increment the occurrence of y_{ni} ;
note: (*key*, *value*) = (y_{ni} , number of occurrences)
 - 3.2: End *For Loop*
4. End *While Loop*
5. Data set is created as input of Reducer: (*key*, *value*) = (y_{ni} , *number of occurrences*)

Fig. 2. MBA Algorithm for Mapper

Woo et al [1, 2] proposes a Market Basket Analysis (MBA) Algorithms on MapReduce as Figures 2 and 3. Mapper reads the input data and creates a list of pair items for each transaction. For each transaction, its time complexity is $O(n)$ where n is the number of items for a transaction. The time complexity to sort each transaction data is $O(n \log n)$ on merge sort. Then, the sorted items should be converted to pairs of items as keys, which is a cross operation in order to generate cross pairs of the items in the list as shown in Figure 2. Its time complexity is $O(n \times m)$ where m is the

number of pairs that occurs together in the transaction. Thus, the time complexity of each mapper is $O(n + n \log n + n \times m)$.

The reducer is to accumulate the number of values per key. Thus, its time complexity is $O(v)$ where v is the number of values per key.

- 1: Read (y_{ni} , <number of occurrences>) data from multiple nodes
2. Add the values for y_{ni} to have (y_{ni} , total number of occurrences)

Fig. 3. MBA Algorithm for Reducer

4.3 Apriori-MapReduce Algorithm

Apriori-Algorithm has been popular to generate frequent item sets of transaction data in order to build an association rule in sequential computing. It is based on minimum support and *Apriori-Property* (or *Downward Closure Property*) where all subsets of a frequent item set must also be frequent. For example, when minimum support is 2 with size 2 item sets as $\langle [item\ pairs], frequency \rangle$, the following data set is generated from transaction data of a store:

$\langle [coffee, cracker], 3 \rangle, \langle [coke, cracker], 1 \rangle$

In the next loop, size 3 item data sets are produced as follows:

$\langle [coffee, cracker, milk], ? \rangle, \langle [coke, cracker, milk], ? \rangle$

By applying Apriority Property to the size 3 item sets, $\langle [coke, cracker, milk], ? \rangle$ is eliminated before counting the frequencies of the item sets as its subset $[coke, cracker]$ occurs only once, which is less than the minimum support value 2 so that the remaining data set saves unnecessary computing time.

Woo [3] proposes an *Apriori-MapReduce* algorithm that computes item sets iteratively using MapReduce functions. However, it has a difficulty to achieve a performance gain in MapReduce parallel computing because at each iteration i , (1) apriori item sets i should be stored as a file or at a DB and (2) item sets $i+1$ at each transaction needs to read all data of the file or DB of apriori item sets i . Therefore, Woo's apriori algorithm needs to be improved for MapReduce.

4.4 HBase for MBA Algorithm

In the paper, *MBA* algorithm on Hadoop accesses data of *HBase* on *HDFS*, which is *column-oriented DB* that supports structured data storage for horizontally scalable tables. *HBase* is relatively easy to integrate with *Hadoop*. The transaction data set files are migrated and stored to the *HBase* DB, which are analyzed with the proposed *MBA* algorithm on *Hadoop MapReduce* platform in order to extract item pair sets.

Table 1. *mba* table schema and data

	Items
	Items: List
Trax 1	cracker, icecream, beer
Trax 2	chicken, pizza, coke, bread
...	

For *MBA*, we design an *HBase* schema, which has a table “*mba*” to store transaction data as input and a table “*db2*” and “*db3*” to store two- and three-paired data set as output respectively. Table 1 is a table schema of *mba*, which has *Items* as a column family and *List* as a column of *Items*. Each row represents a transaction as a key. Thus, for example, the cell value of *Items:List* is a list (*cracker, icecream, beer*) of Transaction 1.

Tables 2 and 3 show table schema of *db2* and *db3*, which has *Items* as a column family and *Count* as a column of *Items*. Each row contains a pair of items as a key. *MBA* algorithm generates 2- and 3-item pairs that are associated. For example, the pair (*beer, hering*) at *db2* is extracted with total 4,566 counts in the total transactions.

Table 2. *db2* tables for two pair data set

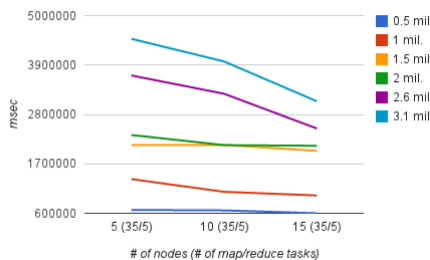
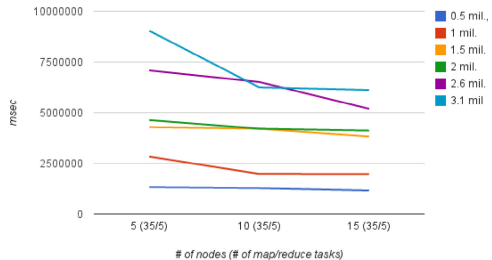
	Items
	Items: Count
beer, hering	4,566
corn, hering	4,664
...	...

Table 3. *db3* tables for three pair data set

	Items
	Items: Count
beer, cracker, hering	2,516
Chicken, coke, pizza	3,621
...	...

5 Experimental Result

The *MBA* algorithm is built in *Java Open JDK 6*, *Hadoop 0.20.2*, *HBase 0.90*, and *Zookeeper-3.3.3* with *Whirr 0.8.0* on *AWS EC2 m1.small* instance type [1-6]. We have 6 transaction files for the experiments: *0.5M*, *1M*, *1.5M*, *2.1M*, *2.6M*, *3.1M*. Those are run on small instances of *AWS EC2*, where each instance as *m1.small* is \$0.06/hr. and is composed of 1 core (1 *EC2* compute unit), 1.7GB memory and 160GB storage on 32 bits platform with *Ubuntu-10.02 OS*.

**Fig. 4.** Comp. Time for 2 items**Fig. 5.** Comp. Time for 3 items

The number of map and reduce tasks are 35 and 5 respectively. And its result data set, that is, 2-/3-paired items are stored into the *db2* and *db3* tables respectively. As shown in the Figures 4 and 5, the computing times for item pairs 2 and 3 become shorter when the nodes get larger. Especially, when the transaction data set is larger than 2.6 millions, the computing time becomes more efficient with the larger nodes. In Figure 4, comparing the nodes 5 (4,483.5 sec) and 15 (3,092.9 sec), computing time on 15 nodes gets 31% faster than on 5 nodes. In Figure 5, comparing the nodes 5 (9,039.7 sec) and 15 (6,117.2 sec), computing time on 15 nodes gets 47.8% faster than on 5 nodes.

6 Conclusion

HBase schema is presented, which keeps input and output data accessed by a *Market Basket Analysis* Algorithm to find the most frequently occurred pair of products in transactions. And, the performance of the algorithm with *HBase* is measured.

The experimental result shows that the *MBA* algorithm has 31% ~ 47.8% faster performance gain while running on large number of nodes, especially in more than 2.6 million transaction data. However, when the input data reaches at 3.6 millions, *Zookeeper* and *HBase* lose the control and connection as it has too many I/Os with less memory for a node.

References

1. Woo, J., Xu, Y.: Market Basket Analysis Algorithm with Map/Reduce of Cloud Computing. In: The 2011 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2011), Las Vegas, July 18-21 (2011)
2. Woo, J., Basopia, S., Xu, Y., Kim, S.H.: Market Basket Analysis Algorithm with NoSQL DB *HBase* and Hadoop. In: The Third International Conference on Emerging Databases (EDB 2011), Songdo Park Hotel, Incheon, Korea, August 25-27 (2011)
3. Woo, J.: Apriori-Map/Reduce Algorithm. In: The 2012 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2012), Las Vegas, July 16-19 (2012)
4. Apache Hadoop Project, <http://hadoop.apache.org/>
5. Apache *HBase*, <http://hbase.apache.org/>
6. Apache Whirr, <http://incubator.apache.org/whirr/>
7. Lin, J., Dyer, C.: Data-Intensive Text Processing with MapReduce. Tutorial at the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2010), Los Angeles, California (June 2010)
8. Lin, J., Schatz, M.: Design Patterns for Efficient Graph Algorithms in MapReduce. In: Proceedings of the Eighth Workshop on Mining and Learning with Graphs Workshop (MLG-2010), Washington, D.C., pp. 78–85 (July 2010)
9. Lin, J., Dyer, C.: Data-Intensive Text Processing with MapReduce. Morgan & Claypool Publishers (2010)
10. Dean, J., Ghemawa, S.: MapReduce: Simplified Data Processing on Large Clusters. In: OSDI 2004, Google Labs, pp. 137–150 (2004)
11. Apache *Zookeeper*, <http://zookeeper.apache.org>

Service Level Agreement Renegotiation Framework for Trusted Cloud-Based System

Ahmad Fadzil M. Hani, Irving Vitra Paputungan, and M. Fadzil Hassan

Centre of Intelligent Signal and Imaging Research
Department of Electrical and Electronics Engineering,
Universiti Teknologi Petronas
31750 Tronoh, Perak, Malaysia
{fadzmo,mfadzil_hassan}@petronas.com.my, irving@uii.ac.id

Abstract. With the widespread adoption of Cloud Computing, the need for trustworthy service providers becomes more important particularly in medical and health related areas. Service terms as one of trust factors are normally defined in the Service Level Agreement (SLA) binding both providers and customers. This paper presents a framework to perform proactive SLA renegotiation during service runtime that aims to maintain trust by customers on the cloud provider. Requirements and important basis for renegotiation such as detecting SLA violation and assessing certain service level boundaries are discussed. Preliminary experimental work shows that by using historical data, the framework is able to provide suitable recommendation on the SLO values that ensures trust is maintained.

Keywords: Cloud Computing, Service Level Agreement, Renegotiation, Trust.

1 Introduction

Service Level Agreement (SLA) is defined as a contract between the service provider and the service consumer in which the expectations of the service provisioning is specified, including penalties that should be applied when a violation occurs [1] [2]. It contains certain service level objectives (SLOs) that define objectively measurable conditions for the service, e.g. throughput and response time. SLOs can vary depending on the applications or the data that are outsourced [6].

The contents in SLA are important in the adoption of an emerging technology called cloud computing [7]. It essentially describes the evidence of trustworthiness cloud consumers have in cloud provider's ability as they have to put the data, even critical data, and rely the service to any cloud infrastructures [4]. Trust is defined as "the expectation of one person about the actions of others that affect the first person's choice, when an action must be taken before the actions of others are known" [8]. Customers will not outsource their data without strong assurances that their requirements will be enforced. Hence, an SLA must be very clear, well negotiated and managed.

SLA is established by a negotiation process between both parties prior to service provisioning [9]. Negotiation in this area is commonly defined as the process by which some parties come to a mutually acceptable agreement on some matter [10]. Normally, once the SLA is constructed, all the terms in the SLA remain fixed until end of the service lifetime. This contradicts with the situation in cloud environment where flexibility is the one of main characteristic of cloud computing technology. Hence, service provider must comply with the consumer's needs and circumstances that may change over time [3] [12]. There is a necessity to add SLA renegotiation process in the SLA management for cloud-based system. This will allow customers and providers to initiate changes in the established agreement, for example the storage size need to be resized as the excessively growing of data, the bandwidth as the capacity channel tend to be reduced in certain periods of time. It is evident that changing circumstances leads to the development of an SLA renegotiation [12]. Pearson (2009) added that renegotiating the contract could also improve some trust level to the customers [13]. Renegotiation is not trivial task [14] as it requires specific protocols for changing the SLA parameters [15]. Furthermore, if an important SLO is violated, renegotiation can be difficult as typically it can affect business profit [15]. Above all, it is necessary to have a mechanism to help SLA renegotiation since this notion is needed in the cloud-based architecture [6] [12].

2 Related Work

SLA renegotiation has been reported in the literature. Wu conjectured the more important the violated SLO, the more difficult it is to renegotiate the SLA, because no parties want to lose their competitive advantages in the market [7]. Initial SLA is important because the customers will not accept a renegotiated SLO value if it is highly deviated from the agreed level [15]. Aiming at on-demand usage changing from the cloud user, Bolor *et al.* suggested the idea of context awareness [20]. User's context with regard to different situations, will give smarter and more tailored responses to enhance customer service. Context is also needed to gain the situational update in the cloud environment [19]. In renegotiation, Smit and Stroulia presented how to maintain and evolve the SLA during service lifetime to retain customer satisfaction [11]. The important experience, accurate evaluation of low-level SLA metrics, and the importance of context to value are necessary to SLA maintenance strategies.

Generally, renegotiation is addressed in a reactive or proactive manner as shown in Table 1. Reactive renegotiation is performed when SLA violation has occurred while in proactive renegotiation, violation is predicted in advance. From the above the discussion, the mechanism for proactive SLA renegotiation has not been fully addressed. The mechanism should provide the most considerable SLO values within user and cloud environment contexts and to escalate customer's trust instead of giving customers altered service levels without their consent.

Table 1. Related works on reactive and proactive renegotiation

Renegotiation	Approach
Reactive	Add renegotiation protocol in WS-Agreement that supports multi round renegotiation [12]
	Add function of ‘Guarantee Terms’ in the WS- Agreement to reduce negotiation overheads [16]
	Use agent technology to do autonomous QoS negotiation and renegotiation [17]
	Extended the WS-Agreement by integrating renegotiation function [18] [22]
Proactive	Integrate service discovery mechanism that reserves some resources from another service provider to support renegotiation [3]

This paper proposes a framework and protocol of automatic SLA renegotiation in order to fully support cloud-based system while maintaining the level of trust among all customers. The renegotiation framework comprises some properties such as proactive action capability, the established SLA, and related contexts. Context in this paper relates to the acquisition of information about the resource available, cloud workload and network performance. Using such contexts, a cloud provider is able to give an optimum offer to a customer without compromising the service delivered to other customers.

3 Requirement for SLA Renegotiation

This section contributes the requirements for a renegotiation framework that exhibits the aforementioned properties. They are described as follows:

- An efficient monitoring mechanism to measure the resource usage in the cloud and how well a service level is delivered to the customer.
- An approach identifying the actual performance vis-a-vis the agreed level, and to obtain the trend within a certain period.
- If violation on SLOs is forecasted, there must be a mechanism to select automatically the violated SLOs.
- Capability to assess the temporal properties from performance and resource usage data.
- Capability to calculate certain boundaries based on historical data for selected parameters to find a spanning range of acceptable values.
- A timely manner renegotiation procedure that generates offer-counteroffer processes based on an estimated limit.

Of particular importance is the limit assessment for agreed SLA parameters based on historical data that is used to understand the temporal context prior renegotiation. An acceptable service level must be maintained from customer perspective and the service must not fall into over-provisioning from provider side. For instance, the agreed throughput is 64 KB/s but from the historical data, the provider deduced the minimum

acceptable throughput to be 50 KB/s. The provider must also be aware of the maximum delivered throughput level (e.g. 70 KB/s) that is still profitable and without creating negative consequences to other customers.

4 SLA Renegotiation Framework

Figure 1 describes the proposed conceptual framework addressing the requirements outlined in previous section. Each process is explained as follows.

A. *Monitoring System*

The monitoring system must be able to capture the actual performance of the network infrastructure, the available resource on provider side, the resource usage on customer side, and detect any abnormality. Since no standard models exist for such a solution in which all parameters needs to be covered and yet each cloud has its own specific monitoring metrics, having a customized monitoring tool is highly recommended. The monitoring data acquired will be used to detect any violation of parameters and to calculate the baseline pattern for such parameters.

B. *SLA Violation Prediction*

It is important to predict possible violations before they actually occurred, even though it cannot directly help to prevent them. With a prediction of the service level trend during monitoring stage, estimation can be done to decide which parameters will be taken forward to the renegotiation stage.

C. *SLA Parameters Selection*

This stage comprises manual selection by the customer and automatic selection by the provider to provide inputs for SLA renegotiation. When customer initiates a renegotiation, he will then select manually the parameters that need to be reassessed. If violations on certain parameters are expected, they become the inputs for the limit assessment procedure automatically.

D. *SLA Parameters Limit Assessment*

Limit assessment defines the upper and lower bound of each selected parameter in the previous step. The upper bound is obtained from the service provider at the time renegotiation takes place, while the lower one is acquired from the minimum acceptable level delivered to such customer in previous service term and from the offer in manual selection. Both boundaries are then set as range of utility value where the new renegotiation utility value will be calculated in the next step.

E. *SLA Renegotiation*

To renegotiate the related SLA parameter, the offer generation approach [21] is chosen due to its capability to provide several best offers at the same time without having any long bargaining steps. This approach needs a good learning strategy to provide the optimal options. The preference generator will give new SLO within specific limit assessed earlier. With this kind of mechanism, customer can make a choice among the optimum offers for the next period of service lifetime.

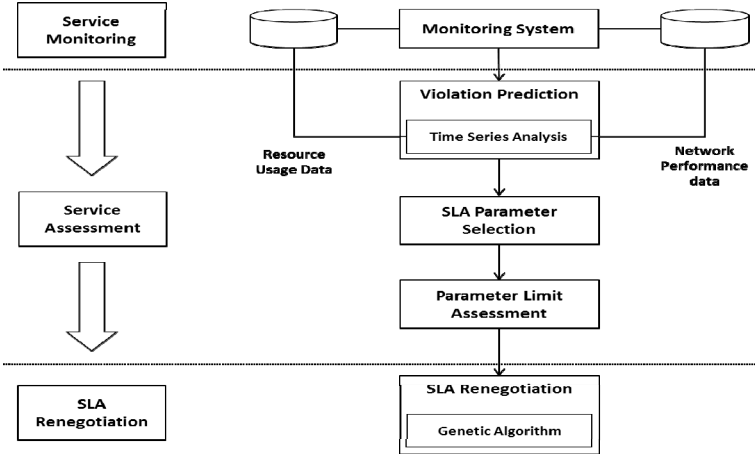


Fig. 1. Conceptual framework of SLA renegotiation mechanism

5 Experimental Work

Due to page limitation, the experimental work will only be described briefly. 30 days monitoring data in December on 4 parameters as sample from private cloud medical research data at a research centre is taken; availability, response time (ms), throughput (kbps), and number of user connected. Genetic Algorithm (GA) [21] is initially implemented in this experiment to support the offer generation. GA parameters are set as follow: Total generation is 100, Crossover probability is 50% and Mutation probability is 1%. The best offer based on such data after several cycles is shown in renegotiated SLO column in Table 2. Since there are some holidays (e.g. Christmas) where only few users connected to the cloud during the particular month, the availability level and user connected is suggested to be reduced to 96% and 14 users respectively. The current situation also makes an improved offer for response time and throughput parameters compared to the initial SLO.

Table 2. Simulation Results of SLA Renegotiation

<i>SLA parameters</i>	<i>Agreed SLO</i>	<i>Renegotiated SLO</i>
Availability	99%	96%
Response time	90ms	68ms
Throughput	16kb/s	15kb/s
User Connected	40 users	14 users

6 Conclusion

This paper describes a proposed SLA renegotiation framework that is essential in trusted cloud computing. A trusted cloud must be able to facilitate the elasticity of cloud due to changes in demand and environment. In particular, the framework allows

customers to adjust their demand frequently while maintaining trust proactively. Such renegotiation framework can lead to improved levels of trust particularly in medical and health related clouds. The framework can also restrain customers from relocating their resource away. It is expected the framework can also contribute to research related to resource optimization, SLA management, and Autonomic Computing.

Acknowledgements. This work is supported by Ministry of Higher Education Malaysia under the ERGS research grant 0153AB-117.

References

1. Begnum, K., Burgess, M.: On the stability of adaptive service level agreements. *IEEE Network and Service Management* 3(1), 13–21 (2006)
2. Kandukuri, B.R., Paturi, V.R., Rakshit, A.: Cloud Security Issues. In: *IEEE Intl Conf. on Services Computing*, pp. 517–520 (2009)
3. Yan, J., Zhang, J., Lin, J., Chhetri, M., Goh, S., Lowalczyk, R.: Towards autonomous SLA negotiation for adaptive service composition. In: *Proc. of the 10th Intl Conf. on Computer Supported Cooperative Work in Design* (2006)
4. Patel, P., Ranabahu, A., Sheth, A.: SLA in Cloud Computing. In: *Cloud Workshops at OOPSLA 2009*, vol. 54, pp. 1–10 (2009)
5. Mahbub, K., Spanoudakis, G.: Proactive SLA Negotiation for Service Based Systems. In: *IEEE 6th World Congress on Services* (2010)
6. Ahronovitz, et al. (about 30 authors): Cloud computing use cases. A white paper produced by the Cloud Computing Use Case Discussion Group, version 4.0 (July 2010)
7. Wu, L., Buyya, R.: Service Level Agreement (SLA) in Utility Computing Systems. In: *Performance and Dependability in Service Computing: Concepts, Techniques and Research Directions*, pp. 1–25 (2010)
8. Dasgupta, A., Prat, A.: Reputation and asset prices: A theory of information cascades and systematic mispricing. In: *Workshop on Informational Herding Behaviour*, pp. 1–45 (2005)
9. Pichot, A., Wäldrich, O., Ziegler, W., Wieder, P.: Dynamic SLA-negotiation based on WS-Agreement. *CoreGRID Technical Report. TR-0082, Network of Excellent* (2007)
10. Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Sierra, C., Wooldridge, M.: Automated Negotiation: Prospects, Methods and Challenges. *Group Decision and Negotiation* 10(2), 199–215 (2001)
11. Smit, M., Stroulia, E.: Maintaining and Evolving SLA: Motivation and Case Study. In: *International Workshop on the Maintenance and Evolution of Service-Oriented and Cloud-Based Systems (MESOCA)*, pp. 1–9 (2011)
12. Parkin, M., Hasselmeyer, P., Koller, B., Wieder, P.: An SLA Re-Negotiation Protocol. In: *2nd Non Functional Properties and Service Level Agreements in Service Oriented Computing Workshop (NFPSLA-SOC 2008)*, Ireland (2008)
13. Pearson, S., Charlesworth, A.: Accountability as a way forward for privacy protection in the cloud. In: Jaatun, M.G., Zhao, G., Rong, C. (eds.) *Cloud Computing. LNCS*, vol. 5931, pp. 131–144. Springer, Heidelberg (2009)
14. Sharaf, S., Djemame, K.: Extending WS-Agreement to Support Re-Negotiation of Dynamic Grid SLAs. In: *eChallenges 2010*, pp. 1–8 (2010)

15. Venticinque, S., Aversa, R., Di Martino, B., Rak, M., Petcu, D.: A Cloud Agency for SLA Negotiation and Management. In: Guarracino, M.R., et al. (eds.) Euro-Par-Workshop 2010. LNCS, vol. 6586, pp. 587–594. Springer, Heidelberg (2011)
16. Sakellariou, R., Yarmolenko, V.: On the Flexibility of WS-Agreement for Job Submission. In: Proc of the 3rd Intl Workshop on Middleware for Grid Computing, pp. 1–6 (2005)
17. Yan, J., Kowalczyk, R., Lin, J., Chhetri, M.B., Goh, S.K., Zhang, J.: Autonomous service level agreement negotiation for service composition provision. *Future Generation Computer Systems* 23, 748–759 (2007)
18. Battre, D., Brazier, F.M.T., Clark, K.P., Oey, M.: A Proposal for WS-Agreement Negotiation. In: 11th IEEE/ACM Intl Conf. on Grid Computing, pp. 233–241 (2010)
19. Zulkernine, F., Martin, P.: An Adaptive and Intelligent SLA Negotiation System for Web Services. *IEEE Trans. on Services Computing* 4(1), 31–43 (2011)
20. Bolor, K., Chirkova, R., Viniotis, Y., Salo, T.: Dynamic request allocation and scheduling for context aware applications subject to a percentile response time SLA in a distributed cloud. In: IEEE Intl Conf. on Cloud Computing Technology and Science, pp. 464–472 (2010)
21. Niu, X., Wang, S.: Genetic Algorithm for Automatic Negotiation Based on Agent. In: Proc. of the 7th World Congress on Intelligent Control and Automation, China, pp. 3834–3838 (2008)
22. Di Modica, G., Tomarchio, O., Vita, L.: Dynamic SLAs management in service oriented environments. *Journal of Systems and Software* 82, 759–771 (2009)

A Cross-IdP Single Sign-On Method in SAML-Based Architecture

Tzu-I Yang¹, Chorng-Shiuh Koong², and Chien-Chao Tseng¹

¹ National Chiao Tung University, Department of Computer Science, Hsinchu, Taiwan

² National Taichung University of Education,

Department of Computer and Information Science, Taichung, Taiwan

{tiyang, cctsens}@cs.nctu.edu.tw, csko@mail.ntcu.edu.tw

Abstract. Security Assertion Markup Language, which is an XML-based framework, has been developed to describe and exchange authorization and authentication information between on-line business partners. One of the major applications is used to achieve single sign-on through different cloud services. SAML has provided the basic assertion of security that allows the user to surf hybrid clouds of the enterprise. The identify provider, which in charge of the management of the user information, can help users access these services effortlessly. However, the user anonymity of SSO from different identify providers is still an open issue even in SAML 2.0. In this study, we propose a SSO architecture for hybrid cloud to achieve identity federation cross-IdP using SAML, which provide the user an enterprise-crossed, services-integrated, backward compatible, and anonymity-maintained environment.

1 Introduction

With the rapidly growing of Internet techniques, cloud computing becomes the mainstream, which can provide all kinds of services. Because various services may come from different cloud servers, users may be asked to login again and again to provide valid credentials. In order to integrate differences and provide a mature and high quality environment, single sign-on (SSO) is introduced to solve this kind of problem. SSO is a property of access control with multiple related, but independent software systems. With this property a user logs in once and gains access to all systems without being prompted to log in again at each of them.

SAML is an XML-based solution, which has developed, by the Security Services Technical Committee of the standards organization, the Organization for the Advancement of Structured Information Standards (OASIS), for exchanging user security information between enterprises and service providers. It supports W3C XML encryption and service provider initiated web single sign-on exchanges. It can be imagined that after logging in at Yahoo home page, the user could use its mailbox, auction, photo album services, without providing credentials again. However, users may have to login many times to use the services while the various services are subordinate under different IdPs. In addition, it is hard to convince these enterprises to share authentication schemes since it may cause the security anxiety. The commercial

conflict and the resource limitation may also prevent the construction of the global IdP from being carried out.

In this study, we proposed a cross-IdP SSO method in SAML-based architecture, which provide the user a federation cross-IdP environment with the features of backward compatibility and personal anonymity. The message transfer from different IdPs can also prevent from malicious parties since the SAML V1. x has already supported PKI-based protocol. The related works will be briefly presented in section 2, and the proposed architecture is described in section 3. We have the discussion in section 4, and we conclude the features and future work in section 5.

2 Related Works

The solutions of Federation system are usually based on a Federated Identity Model (FIM) [1], with the property of authentication and authorization, and provide the ability of interoperability securely between heterogeneous information systems. SSO is one of the FIM functions whereby a single action of user authentication and authorization can permit a user to access all services, which the user has access permission without the need to enter multiple passwords. The other advantages of SSO systems are the security and anonymity. The users' privacy can be retained because there is only one authentication portal, which receives and stores users' credentials. The applications only receive information about whether they may let the user in or not. Also, the user authenticates only once, which means the transfer of sensitive information over the network can be limited.

There are mainly three implemented models: SAML [2], OpenID [3] and Microsoft CardSpace [4]. SAML is the most mature and comprehensive technology and has undergone standardization in 2002 (SAML 1.1) as well as in 2005 (SAML 2.0), respectively [5]. SAML defines a XML-based solution to perform SSO which allows users to gain access to website resources in multiple domains without re-authentications. To achieve SSO, the domains need to form a trust relationship before they can share an understanding of the users' identity. Following the specification of SAML, IdP is required for most of the SAML-based architectures. Alternative solutions [6] present approaches without an IdP by applying X.509 certificates for authentication only. However, the intermediate server is required to manage both the authentication and authorization processes between clients and SP. To support more commercial situations in the real world, we must enable the users also to achieve SSO under many Certificate Authorities.

One of the most common examples is the university campus. It may support various backend authentication mechanisms like Kerberos, LDAP and relational database. Although there exist researches [7, 8] that can be applied to the real circumstances, they may still lack of the demands of flexibility and scalability. The flexibility may involve the dynamic joined student associations, academic exchanges and department anniversary, which may join and leave frequently in a period of time. The scalability may involve the resource sharing among different campuses, universities and even with the industries.

3 Our Proposed System Architecture

In this study, we provide system architecture to achieve the cross-IdP identity federation. In this chapter, we exhibit the scenario, explain the concepts and provide the related algorithms.

3.1 Scenario

One of the most complicated environments is the university campus which involves various kinds of services and commercial services. Students who may register in two universities since they provide various kinds of curriculums. Many campuses also provide virtual money, which means students can pay for goods by using their student IDs (Figure1). Assume the user register with the name Gobby in NCTU for master degree, he may want to pay for the tuition or shop at the department store with different identity Bill. One scenario is that he has logged in at one of the web sites, he does not want to authenticate again and then can access these services or use the student ID to pay for bills, which may locate at different webs or locations. Another example of the usage is that if the user want to join the conference which does not belong to one of the members of the federation. These kinds of SPs may occur dynamically, which is hard for traditional SSO architecture to cooperate with these SPs. In accordance with these scenarios, the topology of different systems must be maintained and managed dynamically. Hence, to provide the backward compatibility and maintain the spirit of SSO, we introduce the new virtual role, named Manager, the detail is provided as follow.

3.2 System Components

There are three main roles in our proposed system includes Service Provider (SP), Identity Server (IdP) and Manager. Since all IdPs should provide authentication schemes for different services in original SAML-based architecture, we can recognize the IdPs as the portal sites which provide various kinds of services. Traditionally, a user may be asked to login the website many times if the SPs that he has visited is subordinated by different IdPs. To provide the basic facility of federation, IdPs must share the user criminal which leads to the tense of users' privacy leakage. On the other hand, users who have several accounts for different web sites may also cause the ambiguity of federations. Therefore, we introduce a new role, Manager, to handle

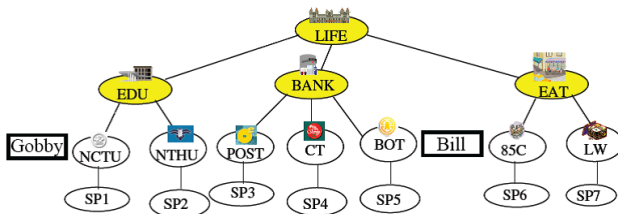


Fig. 1. Layered Concept in Actual Commercial System

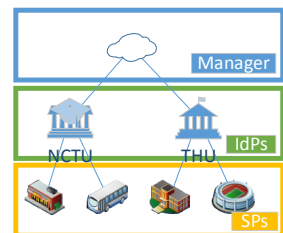


Fig. 2. System Components

the communication between different IdPs. Figure 2 demonstrates the component hierarchy. By using user Link_ID, which is similar to the pseudonym, can help Manager to cooperate different IdPs in a seamless way. One of the major contributions of our architecture is the anonymity since the IdPs only maintain the local identities. While communicating with other IdPs under the federation architecture, the user can be identified through the Link_ID, which is automatically created by the correspondent IdP (i.e. Manager) and the privacy information will not be exchanged through different IdPs.

3.3 Manager

In order to achieve identity federation, we define a third-party architecture, the Manager, to maintain users' identities. The Manager records the corresponding identities at different IdPs for the user by using account linking table (ALT), which can help different IdPs communicate with the global Link_ID with different identity of the same user. It provides privacy-preserving characteristic for a user during a web SSO exchange. The process of associating a Link_ID with another Link_ID at a partner is called account linking afterward. The user privacy can be guaranteed since the IdPs only maintain the local identities, which means there is no extra private information exchanged between different IdPs.

Figure 3 shows the example of ALT. Assume the user register on one of the IdPs, and the user wants to use the services on another website with different identities, he can log in at IdP_1 using his Gobby account and use another account Bill at IdP_2. Because the Manager will maintain the ALT, which helps the user use different account from different web sites at the same time. Besides, the Link_ID can also be different which provide the anonymity since these IdPs cannot get the local identities from each other.

3.4 Global Path Finding

The architecture defines that a user at an IdP wants to visit another IdP by following the tree path rather than peer to peer. Deliberate that the source IdP and destination IdP might not be below the same manager, so how to know the position of the destination IdP must be solved. Managers must record the nodes of its upper layer and lower layers. Briefly, in a path, a manager stores the Link_ID of the left neighbor along with the Link_ID of the right neighbor for every user. The global path of an IdP is encoded

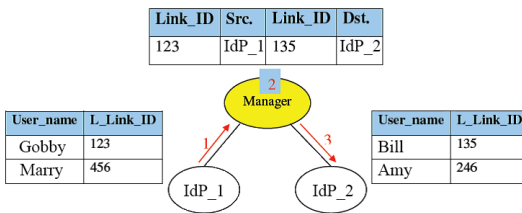


Fig. 3. Account Linking Table in the Manager

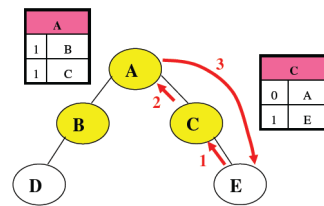


Fig. 4. An example of global path finding

in a sequence of strings, which can indicate the position and relationship of the IdP. For the purpose of expansibility, the global path string is coded dynamically. (Figure 4) gives an example, the global path of IdP E should be coded "A_C_E", it indicates that E's manager is C and C's manager is A, and so on. The construction start with E sends a request to its Manager to get its global path. C receives the request and then appends C. The path code becomes "C_E". C looks up its table to know that it is not the top layer, then C keeps on sending the request to the upper layer. While A receives the request and appends A, the path code becomes "A_C_E". A refers to its table knowing there is no more upper layer, then sends back code path "A_C_E" to E. An IdP site can ask its global path upward gradually as above and stores the path code in the local site. Each IdP asks upward periodically and updates its path code when the topology changed.

3.5 Core Algorithm of Manager

When getting the global path of the destination IdP, the source IdP will send the string and the request of identity federation to its manager. After receiving, the manager will compare the string with its name to determine whether its name appears or not. If it appears, the fact expresses that the target IdP is under its subtree surely and the manager should send the request downward. Otherwise, the manager sends the request upward. Each manager, which receives the request of identity federation, will look up its ALT to determine whether a previous federation has been established for the user. If a previous federation has been established then continue to transmit forward according to the Link_ID, otherwise the manager needs to generate a Link_ID and record it to establish a new account linking entry. Due to the limitation of the paper pages, the algorithm of managers will provide in the future works.

4 Discussions

We also implemented the whole system and deployed in an university of north Taiwan. However, due to the limitation of paper pages, we strictly go analyze and discuss the proposed architecture.

4.1 Security

All information that delivered is Link_ID rather than real credentials. The source IdP and destination IdP neither know the ID used on the other side. It provides privacy-preserving characteristic for users. Besides, the malicious IdPs can be omitted since SMAL 1.x had already provided the PKI-based protocols, which means that all the exchanged messages can be encrypted and verified by using the SSL protocol.

4.2 Expansibility

The expansibility property can be discussed in vertical and horizontal ways. Each Manager must record the nodes of its upper layer and lower layers. We can take advantage of it to insert or delete SPs under different IdPs (Manager). Assume that a relationship exists as (Figure 1). Someday all three groups want to cooperate in demand,

and then the manager LIFE will be constructed in need to integrate them. If a new alliance, e.g. Play alliance, wants to join the exist federation, it can simply vertically add to the BANK alliance tree as a subordinate or horizontally add to the LIFE alliance.

4.3 Robustness

In the section, we would consider the influence when a manager is destroyed in the business system. Most probably account link in the node would not be operated. But it can work without involving the crashed node, including account linking. Besides, we consider another situation. That is, there is a petition to a certain manager and the connected was interrupted. The Link_ID which is generated previously would not disappear. After connecting successfully, the Link_ID can be used for another federation.

5 Conclusion

In this study, we proposed a method that can provide users an enterprise-crossed, services-integrated, backward compatible, and anonymity-maintained environment by introducing the concept of pseudonym based on SAML. All identity federations are established by Managers and dispersed evenly, and the communication between these Managers (IdPs) are secured by using PKI-based SSL. Comparing to the traditional SSO, regardless of sorting and storages may be required, it is more efficient and applicable. In the future, we will focus on developing a secure and efficient method to rebuild the destroyed Managers, because the recovery of previous federations may be required.

References

1. Bhatti, R., Bertino, E., Ghafoor, A.: An integrated approach to federated identity and privilege management in open systems. *Communications of the ACM* 50, 81–87 (2007)
2. Lockhart, H., Mishra, B.: *Security Assertion Markup Language (SAML) 2.0 Technical Overview* (2005)
3. Recordon, D., Reed, D.: OpenID 2.0: a platform for user-centric identity management. *Discovery*, 11–16 (2006)
4. Bhargavan, K., Fournet, C., Gordon, A.D., Swamy, N.: Verified implementations of the information card federated identity-management protocol. In: *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security, ASIACCS 2008*, pp. 123–135 (2008)
5. Maler, E., Reed, D.: The Venn of Identity: Options and Issues in Federated Identity Management. *IEEE Security & Privacy Magazine* 6, 16–23 (2008)
6. Fugkeaw, S., Manpanpanich, P., Juntaprenjitt, S.: A Robust Single Sign-On Model Based on Multi-Agent System and PKI. In: *Sixth International Conference on Networking (ICN 2007)*, pp. 4–9 (2007)
7. Akiyama, T., Teranishi, Y., Okamura, S., Sakane, E., Hasegawa, G., Baba, K., Nakano, H., Shimojo, S.: A Report of Campus-Wide IT Authentication Platform System Development in Osaka University. In: *2007 International Symposium on Applications and the Internet Workshops*, p. 35. IEEE (2007)
8. Shen, J., Zhu, C.: Design and Implementation of Single Sign-on Using Yale-CAS. *Computer Technology and Development* (2007)

Live Virtual Machine Migration with Optimized Three-Stage Memory Copy^{*}

Feiran Yin, Weidong Liu, and Jiaying Song

Department of Computer Science and Technology
Tsinghua University
Beijing 100084, China

yinfr10@mails.tsinghua.edu.cn, {liuwd, jxsong}@tsinghua.edu.cn

Abstract. Live virtual machine migration has become an important management method in clusters and data centers. It allows application isolation and facilitates server consolidation, load balancing, fault management and power saving. Existing live migration approaches pre-copy have to iteratively copy redundant memory pages, another approach post-copy would lead to a lot of page fault and application degradation. This paper presents the detail design of a novel three-stage memory copy live migration approach. Memory pages only need to be transmitted twice at most, and page fault just occurred in small part of dirty pages. We implement it in Xen 4.1.4 and compare it against Xen's original pre-copy approach. The evaluation results under various memory workloads show that our approach can significantly reduce total migration time and total pages transferred.

Keywords: virtual machine, live migration, three-stage memory copy.

1 Introduction

Virtualization technology develops rapidly in recent years. The resources of a single machine are divided into multiple isolated virtual resources by using some virtualization softwares [1]. It provides application isolation, server consolidation, better multiplexing of data center resources, the ability to flexibly remap physical resources and so on [9].

Live migration is the key point of virtualization technologies. It allows virtual machines fast relocating in data center and no aware of downtime. Most of the live migration techniques use pre-copy approach. It first transfers all memory pages to target and then copies dirtied pages iteratively, until writable working set (WWS) becomes small or the preset number of iterations is reached, then suspends VM in source node and sends CPU state and remaining dirty pages in the last round to the target, where the VM is restarted [3]. However, great application degradation would happen in pre-copy phase because migration daemon continually consumes network bandwidth to transfer dirty pages in each round, it leads to longer migration time.

^{*} This work is supported by the National Basic Research Program of China (973 Program) under grants 2013CB329100 and 2013CB3291005.

In this paper, we present an optimized memory copy approach for live virtual machine migration. We combine the advantages of active pushing and on-demand copy, first copy all memory pages to target and record dirty bitmap in this phase (full memory copy stage), then suspend the VM, transmit CPU state and dirty bitmap (dirty bitmap copy stage), finally resume the new VM and copy dirty pages from source to target (dirty page copy stage). We call it three-stage copy. The main goal of three-stage copy is to minimize total migration time and reduce network traffic. Most of the memory pages need to be copied once in full memory copy stage, only dirtied pages need to be copied twice. We implement this approach on Xen 4.1.4 and compared it against original pre-copy method in Xen. The evaluation results under various memory workloads show that our approach can significantly reduce total migration time and total pages transferred.

This paper is organized as follow. In section 2, we describe related works. Then, in section 3, we describe the design and implement of three-stage copy, and present the experimental results in section 4. Finally, we conclude and give our future work in section 5.

2 Related Work

Clark et al. [3] firstly propose pre-copy live virtual machine migration approach. It first transfers all memory pages and then copies pages just modified during the last round iteratively. There are many virtualization platforms using this approach, such as Xen [2], KVM [7] and VMware [5]. Pre-copy is the prevailing live migration technique to perform live migration of VMs, but in write-intensive workloads, memory pages will repeatedly dirtied and may have to be transmitted multiple times.

Hines et al. [4] propose post-copy instead of pre-copy to solve this problem and reduce total migration time. Post-copy migration defers the memory transfer phase until after the VM's CPU state has already been transferred to the target and resumed there. Post-copy thus ensures that each memory page is transferred at most once, thus avoiding the duplicate transmission overhead of pre-copy. But the downtime is much higher than that of pre-copy due to the latency of fetching pages from the source node before VM can be resumed on the target.

Jin et al. [8] propose using adaptive compression of migrated data: different compression algorithms are chosen depending on characteristics of memory pages. They first use memory compression to provide fast VM migration, and they also design a zero-aware characteristics-based compression (CBC) algorithm for live migration. In the source node, data being transferred in each round are first compressed by their algorithm. When arriving on the target, compressed data are then decompressed. However, memory compression increases the system overhead.

3 Design and Implementation

In this section, we introduce the phase of live migration, and describe the design of three-stage copy approach and its implementation on Xen. The performance of any live virtual machine migration strategy could be gauged by the following metrics.

Downtime: The time during which the migrating VM’s execution is stopped.
Total Migration Time: The sum of all migration time from start to finish.
Pages Transferred: The total count of memory pages transferred, including duplicates, across all periods.

3.1 Memory Migration Phases

Efficient synchronization of the memory state is the key issue of live virtual machine migration. Memory transfer can be achieved by following three phases [3]:

Push: The source VM continues running while certain pages are pushed across the network to the new destination. To ensure consistency, pages modified during this process must be re-sent.

Stop-and-Copy: The source VM is stopped, pages are copied across to the destination VM, then the new VM is started.

Pull: The new VM executes and, if it accesses a page that has not yet been copied, this page is faulted in (“pulled”) across the network from the source VM.

Figure 1(a) shows pre-copy approach, it combines push copying and stop-and-copy. Another approach post-copy in Figure 1(b) uses stop-and-copy and pull copying.

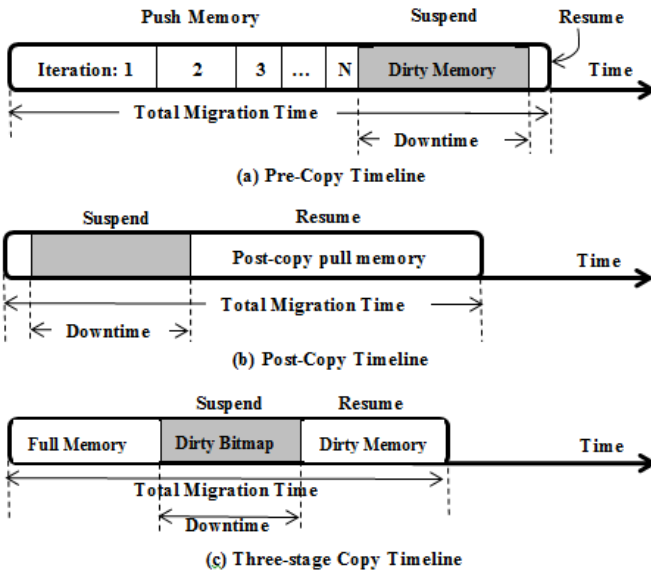


Fig. 1. Timeline for live migration approach

3.2 Design of Three-Stage Copy

To solve the weaknesses of existing live migration methods, we propose a new approach called three-stage copy which combines three phases of memory transfer. The entire memory synchronization is divided into three stages, Figure 1(c):

Full Memory Copy: Copy all memory pages from source VM to destination when the source VM continues running, and record pages modified during this process.

Dirty Bitmap Copy: Suspend source VM, copy recorded dirty bitmap to target node, and mark corresponding pages as dirty in destination VM.

Dirty Pages Copy: Resume new VM, then active push or on demand copy dirty pages from source VM to destination.

Compared with pre-copy, three-stage copy avoids iterative copy dirty pages, most of the memory pages are just copied once, only dirtied pages in full memory copy stage need to be copied twice. It significantly reduces pages transferred, as a result, reducing the usage of network bandwidth. Meanwhile, only dirty bitmap and CPU state need to be transferred in suspend phase, downtime of VM is also shortened. Although it would be interrupted in dirty pages copy stage because of page fault, but relative to full memory copy after resuming new VM in post-copy approach, three-stage copy just transfer dirtied pages after resuming, which significantly reduces the page fault rate, and avoids obvious application degradation, also, it shortens the duration of the migration.

There are two methods used for transferring dirty page, on demand copy and active push. Once the VM resumes at the target, page faults would happen when memory access dirtied page, it can be serviced by requesting the referenced page over the network from the source node. However, page faults in new VM are unpredictable, on demand copy would lead to longer resume time, so we combine it with active push which source host periodically pushes dirty pages to the target in a preset time interval.

3.3 Implementation On Xen

We implemented three-stage copy on Xen 4.1.4. The point of our approach is to capture and recode dirty pages. Shadow page tables is used by Xen's hypervisor to keep track of the memory state of guest OS, it can be used to capture dirty pages. Figure 2 shows the process of shadow page table. Shadow page tables are a set of read-only page tables for each VM maintained by the hypervisor that maps the VM's memory pages to the physical frames. Actually, it is equivalent to a backup of the original page tables, any updates in guest OS's page table will notify Xen's hypervisor by Hypercall.

Because all page tables in guest OS are mapped to read-only shadow page tables, any updates in page tables trigger page faults which would be captured by Xen's hypervisor. Xen checks the PTE access right of the guest OS, and set PTE in shadow page tables to writable if the guest OS is writable to the PTE. Then we can record the updates in shadow page tables into a dirty bitmap.

By this way, we will be able to capture the occurrence of dirty pages, and obtain a dirty page bitmap. Xen provides an API function `xc_shadow_control()` to handle shadow page tables. This feature can be turned on by calling `xc_shadow_control()` and setting flag as `XEN_DOMCTL_SHADOW_OP_ENABLE_LOGDIRTY` before live migration, and turned off by setting `XEN_DOMCTL_SHADOW_OP_OFF` flag after migration finished.

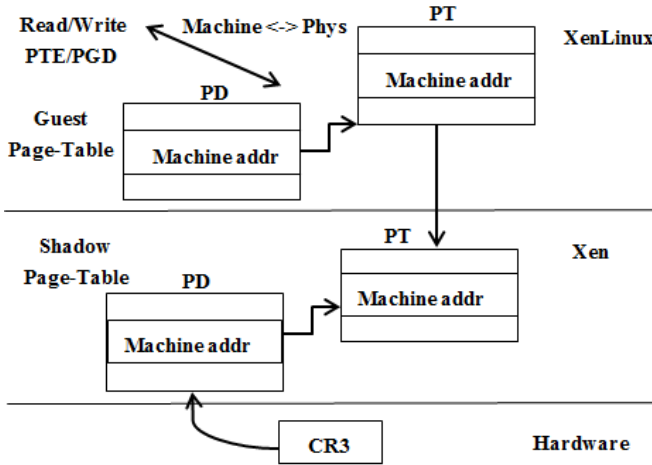


Fig. 2. Shadow page table

4 Evaluation

In this section, we present an evaluation of three-stage copy on Xen 4.1.4 and compare it against Xen’s original pre-copy approach.

4.1 Experimental Setup

We conduct our experiments on two identical server-class machines, each with 2-way quad-core Xeon E5506 2.13GHz CPUs and 32GB DDR RAM, connected via a Gigabit Ethernet switch. All VM images store in a NFS server. We use Ubuntu 12.04 (Linux version 3.5.0-23) as guest OS and the privileged domain OS (domain 0). The host kernel is the modified version of Xen 4.1.4. Both the VM in each experiment and the Domain 0 are configured to use two VCPUs. Guest VM sizes range from 128MB to 1024MB. And we use memtester [10] in virtual machine to generate high memory usage.

Each experiment is repeated five times and every test result comes from the arithmetic average of five values. In migration process, we evaluate three primary metrics discussed in section 3: Downtime, Total migration time, Page transferred.

4.2 Experimental Results

Figure 3(a) shows that three-stage copy significantly reduces the total migration time for diverse VM memory size compared with pre-copy. With memory size in-increasing, the total migration time is reduced more. It reduces total migration time by average of 35.4%. In clusters or data centers, less total migration time of VMs would get higher flexibility.

Figure 3(b) shows that three-stage copy approach also has the advantage in pages transferred, this should be attributed to less data transferred and lower network bandwidth needed. Experimental results show that the three-stage copy reduces at most 37.8% (1024M) and an average of 32.8%.

Evaluation in downtime Figure 3(c) shows that pre-copy could get stable downtime, and three-stage copy's downtime would increase along with the increase of memory size. At low memory environment, three-stage copy need less downtime than pre-copy, but it need more downtime in large memory environment. It is due to three-stage copy need transfer dirty bitmap in suspend phase, large memory size would have more dirty pages. Nevertheless, the tradeoff between total migration time and downtime may be acceptable.

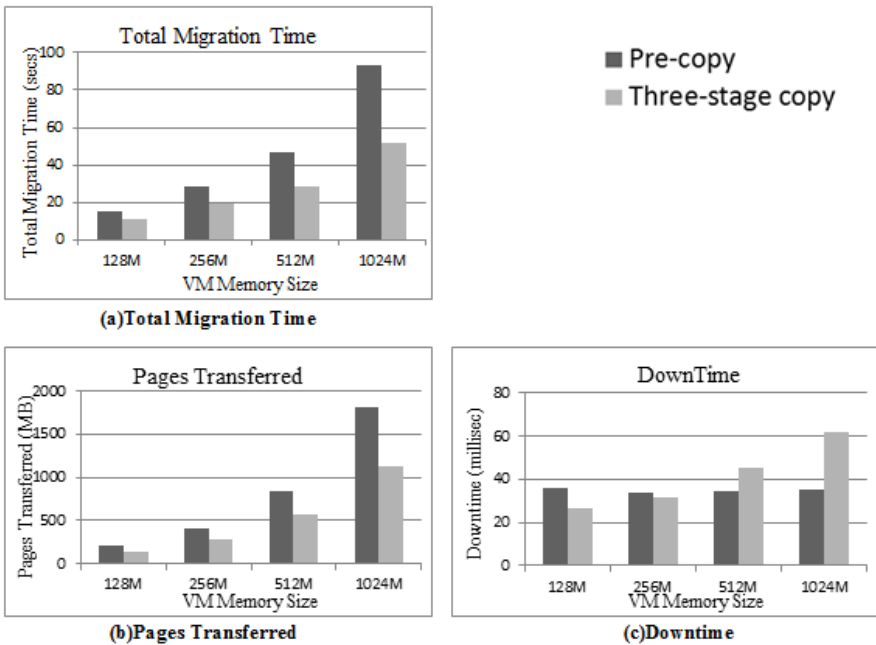


Fig. 3. Comparison of total migration time, pages transferred and downtime

5 Conclusions and Future Work

In this paper, we have presented the design, implementation and evaluation of three-stage copy for live virtual machine migration. In our approach, most of the memory pages are just copied once, only dirtied pages need to be copied twice. It significantly reduces pages transferred and total migration time. And because we just transfer dirty bitmap in VM stop phase, downtime is also shortened. Experiment results show that our approach gets better performance than Xen's pre-copy.

In the future, we intent to add more features like pre-paging [11], ballooning [12], etc. Moreover, we plan to implement it in an automatic load balancing virtualization system.

References

1. Goldberg, R.P.: Survey of virtual machine research. *IEEE Computer*, 34–45 (1974)
2. Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., Warfield, A.: Xen and the art of virtualization. In: *ACM SOSR*, pp. 164–177 (2003)
3. Clark, C., Fraser, K., Hand, S., Hansen, J., Jul, E., Limpach, C., Pratt, I., Warfield, A.: Live migration of virtual machines. In: *Network System Design and Implementation*, pp. 273–286 (2005)
4. Hines, M., Gopalan, K.: Post-copy based live virtual machine migration using adaptive pre-paging and dynamic self-ballooning. In: *ACM SIGOPS on Virtual Execution Environments*, pp. 51–60 (2009)
5. Nelson, M., Lim, B., Hutchines, G.: Fast transparent migration for virtual machines. In: *USENIX Annual Technical Conference*, pp. 391–394 (2005)
6. Liu, H., Jin, H., Liao, X., Hu, L., Yu, C.: Live migration of virtual machine based on full system trace and replay. In: *18th International Symposium on High Performance Distributed Computing*, pp. 101–110 (2009)
7. Kivity, A., Kamay, Y., Laor, D.: KVM: the linux virtual machine monitor. In: *Ottawa Linux Symposium*, pp. 225–230 (2007)
8. Jin, H., Deng, L., Wu, S., Shi, X., Pan, X.: Live virtual machine migration with adaptive memory compression. In: *IEEE International Conference on Cluster Computing*, pp. 1–10 (2009)
9. Wood, T., Shenoy, P., Venkataramani, A., Yousif, M.: Black-box and gray-box strategies for virtual machine migration. In: *4th USENIX Symposium on Networked Systems Design and Implementation*, pp. 229–242 (2007)
10. A utility for testing memory, <http://pyropus.ca/software/memtester/>
11. Denning, P.J.: The working set model for program behavior. *Communications of the ACM*, 323–333 (1968)
12. Waldspurger, C.: Memory resource management in VMware ESX server. *ACM Operating Systems Design and Implementation*, 181–194 (2002)

Performances of New Chaotic Interleaver Design in OFDM-IDMA System

Brahim Akbil and Driss Aboutajdine

LRIT-URAC'29, FSR, Mohammed V-Agdal University, Rabat, Morocco
{akbil.brahim, aboutaj}@ieee.org

Abstract. In orthogonal frequency division multiplexing-based interleave division multiple access (OFDM-IDMA) system, all users can transmit their information in the same time and frequency band. In this case the orthogonality between the users is obtained using an interleaving technique. The choice of a "good" interleaver must demonstrate that the interleavers are weakly correlated, do not require either large memory to store it or a large bandwidth to communicate it between the transmitter and the receiver, and must be easy to generate. In this paper, we develop a new one-dimensional chaotic map, completely based on the logistic map: the "New Logistic Map (NLM)". The simulation results of chaotic dynamical behaviours show that the NLM is a chaotic system and has an ideal distribution. We propose also a new design method to construct the interleaver sets by NLM: NLM Interleaver (NLMi). Our design can be used to reduce the computational complexity and memory requirement, and achieve the same correlation performances compared to other designs.

Keywords: Chaotic system, logistic map, Interleaver, IDMA.

1 Introduction

Chaotic dynamical systems have received a great deal of attention from the research community within Mathematics, Economics, Computer science, Communications, etc. Based on the Lorenz system [1], more nonlinear maps versions of some chaotic systems were created by different researchers and modified or adapted by others according to specific criteria. The introduction of these nonlinear maps has offered several new applications to exploit chaotic dynamics in the next generation of communication systems. Recently, Interleave-Division Multiple Access (IDMA) has been proposed by Ping et al. [2], it is a potential candidate for the next wireless generation systems. To remove some IDMA limits, Mahafeno et al. in [3] have combined orthogonal frequency-division multiplexing (OFDM) and IDMA and developed OFDM-IDMA technique. This technique inherits the OFDM and IDMA advantages and has also its own advantages. The interleaver/deinterleaver is the main component in the IDMA block in the transmitter and receiver of OFDM-IDMA system. The role of this interleaving is to spread the information bits to protect them, in reception, against error bursts due to the transmission channel and noise sources. However the choice of better interleaver has a positive influence on the IDMA performances. The main

criteria to choose a good interleaver are: (1) each two interleavers out of a set of interleavers should not "collide" [4], (2) the minimal bandwidth consumption to exchange the information about interleaver matrix between transmitter and receiver, (3) minimal memory required to store the interleaver matrix, (4) ease to generate it.

In the literature, several designs of interleaver were studied, such as the Orthogonal Interleaver (OI), Random Interleaver (RI), Nested Interleaver (NI) [4], Shifting Interleavers (SI) [5], Deterministic Interleaver (DI) [6] and others [7,8]. These designs are useless to use in most of the iterative MUD of IDMA systems due to the fact that the first interleavers are randomly generated, and the computational complexity is proportional to the number of users. In addition, the initial interleaver matrix has to be transmitted to the receiver, thus a large bandwidth will be consumed.

In this paper, we introduce a new one-dimensional chaotic map, named "New Logistic Map (NLM)". We also study its dynamical properties and stability points. We apply NLM to propose a new design approach to construct a set of interleavers for OFDM-IDMA system, named "NLM Interleaver (NLMI)". This interleaver is obtained from a deterministic nonlinear dynamic system, so it is easy to generate it. In addition, two different NLMI are weakly correlated and a single information exchange between the transmitter and the receiver is only the initial value of NLM. Consequently, a less memory required to store these interleaving matrixes information and a minimal bandwidth will be consumed.

The content of the paper is organized as follows. In section 2 we give an introduction to our new map, and analyze the chaotic dynamical behaviour in this map. The algorithm to construct the NLM Interleavers is presented in section 3. Section 4 presents the simulation results and discussions, and finally section 5 gives conclusions.

2 New Logistic Map

The logistic map is one of the simplest and most transparent discrete chaotic systems exhibiting the order to chaos transition [9], it is described by (1).

$$X_{n+1} = \lambda X_n (1 - X_n) \quad (1)$$

Where $X_n \in [0 \ 1]$, and n is the generation number.

A sequence generated by the logistic map consists of a Cantor set of points in $[0 \ 1]$ whose orbits stay in $[0 \ 1]$. However, many applications use the sequences of N random integers in $[0 \ N]$, such as the interleaving operation.

Let us consider a new system given by:

$$X_{n+1} = \lambda X_n \left(1 - \frac{X_n}{N}\right), \quad \text{Where } X_n \in [0 \ N] \quad (2)$$

To check the numerically chaotic behavior in this system, we analyze the phase space properties in the term of the parameter λ and the bifurcation diagrams.

2.1 Phase Space

The phase space is a mathematical space where the system states are represented by the numbers used to visualize the behavior of a dynamical system. A point in phase space describes whether the system state is an equilibrium point or not. There are two main types of equilibrium points, stable and unstable. A stable equilibrium point is that at which the system oscillates around the equilibrium point and in unstable equilibrium the system moves away from the equilibrium point if it were displaced. To analyze the equilibrium point of NLM, we start by viewing the map for several different values of the parameter λ with $N=100$ shown in the Fig. 1.

In the equilibrium the system does not have real time dependence and $X_e = \lambda X_e (1 - \frac{X_e}{N})$. This given two stationary solutions: $X_e = 0$ and $X_e = \frac{(\lambda-1)N}{\lambda}$.

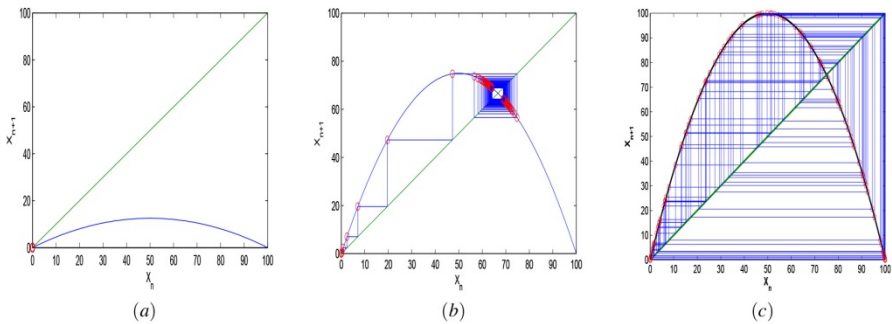


Fig. 1. Phenomenology of the NLM with $N=100$ and (a): $\lambda = 0.5$, (b): $\lambda = 3$, (c): $\lambda = 4$

We conclude that if $0 < \lambda \leq 1$, the equilibrium at X_0 is stable (i.e. Fig. 1-(a) with $\lambda = 0.5$), and if $1 < \lambda \leq 3$, the system (2) is stabilized in equilibrium $\frac{(\lambda-1)N}{\lambda}$ (i.e. Fig. 1-(b) with $\lambda = 3$). At $3 < \lambda \leq 3.54$, we have a violent behavior, in this case the system oscillates on a cycle of four periods without reaching the equilibrium. At $\lambda = 3.57$ chaos occurs; the system never settles to a fixed period. For $3.57 < \lambda \leq 4$, the system evolution is aperiodic (i.e. Fig. 1-(c) with $\lambda = 4$), since the appearance of the chaotic behavior.

2.2 Bifurcation Diagrams

Another step to evaluate the NLM chaotic behavior is to obtain a global visual of the various regimes as the control parameter $\lambda = 4$ when it is varied. As the control parameter λ is changed, the transitions from one fixed point to chaos, are called bifurcations. Fig.2 shows the bifurcation diagram of the NLM function for a sequence of $N = 100$ samples, and $0 < \lambda \leq 4$.

For small values of $\lambda (\lambda \leq 1)$ there is only one stable fixed point, this point is zero. For $1 < \lambda \leq 3$, we still have one point attractor, increases when λ increases. At $3 < \lambda \leq 3.57$ the entire bifurcation is recreated on a smaller scale, and for $\lambda > 3.58$ the system is completely chaotic.

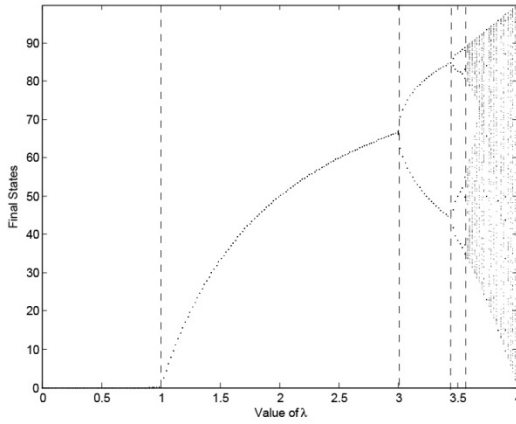


Fig. 2. The bifurcation diagrams for NLM

3 Design Steps of the NLM Interleaver

The system starts with a first value X_0^u of X^u called the initial state (or initial condition) and then calculates a sequence $\{X_1^u, X_2^u, \dots, X_N^u\}$ occupied by the system using our system $X_{n+1}^u = \lambda X_n^u (1 - \frac{X_n^u}{N})$ proposed in section 3; these states form a set of real sequences between 0 and N , where N is the interleaver length. The transition to the integer sequences is realized by maximizing $\lceil X \rceil$, minimizing $\lfloor X \rfloor$ or by using a floor $\lfloor X \rfloor$ to round the elements of X_i to the nearest integers. Finally to obtain the interleaver vector, we eliminate redundant elements from the integer vector found. We construct an initial value for the next user $n + 1$ by adding a footstep ζ to the X_0^u .

The following is the proposed algorithm based on the NLM equation:

- Initialization:
 - $\lambda = 4$, N , U , $u = 1$, $n = 0$, X_i^u , ζ and $Y_0^u = \lceil X_0^u \rceil$: The first element interleaver matrix of u^{th} user ($\pi^u \equiv Y_0^u$).
 - $i = 0$ and $n = 0$ X_i^u : The initial state of u^{th} user ($0 < X_i^u < N$).
- Main operations:
 - If the value of n does not exceed N :
 1. Calculate X_{i+1}^u from the value of X_i^u .
 2. $Y_{i+1}^u = \lceil X_{i+1}^u \rceil$.
 3. Check, if the element Y_{i+1}^u exists in the set π^u , increment i by 1 and repeat the operation, otherwise $\pi^u \equiv \pi^u \cup Y_{i+1}^u$ and $n = n + 1$.
 - Otherwise, if the value of n exceed N :
 1. $X_{i+1}^u = \{ \}$.
 2. $\pi^u \equiv \pi^u \cup X_{i+1}^u$.

4 Performance of NLM Interleaver in OFDM-IDMA System

4.1 Correlation between Interleavers

Pupeza *et al.* in [4], mentioned that the correlation between two interleavers as a measure of "how strong two interleavers collide", and the interleaver is orthogonal if and only if the correlation is null. Then the correlation $\varphi(\pi^1(b_p^1), \pi^2(b_p^2))$ between π^1 and π^2 as the scalar product between $\pi^1(b_p^1)$ and $\pi^2(b_p^2)$ given by: $\varphi_{1,2} = \langle \pi^1(b_p^1), \pi^2(b_p^2) \rangle$. π^1 and π^2 are orthogonal, if and only if $\varphi_{1,2} = 0$.

To simply analyze correlation problems, we evaluate the correlation peak $\varphi = (\pi^u, \pi^{u'})$, as described in [4]. The Fig.3 show the peak correlation performances of OI, RI, SI, NI, and NLMI are studied for 100 users. It is clear that the correlation values for NLMI are very similar to those of other interleaver designs.

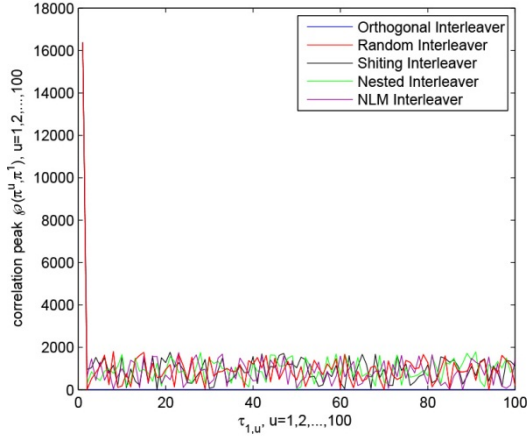


Fig. 3. Peak correlation performances of different interleavers with U=100

4.2 Computational Complexity

The complexity of interleavers and deinterleavers generation at the transmitter and receiver side is a major concern in case of higher user count. Table 1 shows the comparison of computational complexity among the NLMI design and the reference designs. This complexity is calculated by the number of cycles needed to generate the interleaver matrix versus U. The computational complexity increases with the number of users for OI, SI and NI. Therefore, it is fixe in RI and NLMI and independent of U.

Table 1. The computational complexity to generate the first interleaver

	OI	SI	NI	NLMI
Complexity	$O(N) + O(N^2)$	$O(N^2)$	$O(\log_2(N))$	$O(N^2)$

4.3 Bandwidth Resource Required

Lack of bandwidth resource is a vital issue in communication systems. This problem becomes worse when the user number increases. In OFDM-IDMA system, the number of simultaneous users is a very important factor. However, a great number of researchers are interested in increasing the maximum available number with a minimal consumption of bandwidth. Another challenge in OFDM-IDMA system is that the transmitter and receiver must hold the same interleaver matrix. In most of the studied algorithms, the transmitter needs to transmit the interleaver matrix consisting of interleaving pattern of the users to the receiver; so the greater the size of the interleaver, the more bandwidth and resources are used.

To evaluate the bandwidth consumption, we propose a brief overview of the initialization parameters required to generate the interleavers and the number of bits

occupied. In the case of NLM interleaver, a single information exchange between the transmitter and the receiver is only the initial value X_0^1 .

5 Conclusion

We claim two contributions of this work. First we have developed a new chaotic map equation called New Logistic Map (NLM), and we have analyzed its chaotic dynamical behaviors. In the second main contribution, we have proposed a novel method to generate good and multiple interleavers, regardless of the number of simultaneous users and the interleaver length. Our proposed interleaver design is based on the NLM and called NLM Interleaver (NLMI). The simulations show that the NLMI can achieve the same correlation performances compared to random interleavers, shifting interleavers and nested interleavers. NLMI has also many advantages such as less resource consumption, less required memory, minimum implementation complexity and easy to generate it, over all interleaver designs.

References

1. Lorenz, E.: Deterministic nonperiodic flow. *Journal of Atmospheric Science*, 130–141 (1963)
2. Ping, L., Wu, K.Y., Liu, L., Leung, W.K.: A Simple Unified Approach to Nearly Optimal Multiuser Detection and Space-Time Coding. In: *ITW 2002, Bangalore, India*, pp. 20–25 (2002)
3. Mahafeno, I., Langlais, C., Jego, C.: OFDM-IDMA versus IDMA with ISI cancellation for quasi-static Rayleigh fading multipath channels. In: *4th Int. Symp. on Turbo Codes & Related Topics, Munich, Germany*, pp. 3–7 (2006)
4. Pupeza, I., Kavcic, A., Ping, L.: Efficient generation of interleavers for IDMA. In: *IEEE International Conference on Communications, ICC 2006, vol. 4*, pp. 1508–1513 (2006)
5. Zhang, C., Hu, J.: The shifting interleaver design based on PN sequence for IDMA systems. In: *2007 International Conference on Future Generation Communication and Networking, FGCN 2007, Korea* (2007)
6. Tseng, S.M.: IDMA Based on Deterministic Interleavers. *International Journal of Communications, Network and System Sciences* 3(1), 94–97 (2010)
7. Zhang, C., Hu, J.: 2-dimension interleaver design for IDMA systems. In: *Proc. IEEE International Conference on Circuits and Systems for Communications, ICCSC 2008*, pp. 372–376 (2008)
8. Han, L., Jin, M., Song, E.: Matrix Cyclic Shifting Based Interleaver Design for System IDMA. In: *Proc. 5th International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2009*, pp. 1–4 (2009)
9. May, R.: Simple mathematical models with very complicated dynamics. *Nature* 261, 458–467 (1977)

Application of an Artificial Intelligence Method for Diagnosing Acute Appendicitis: The Support Vector Machine

Sung Yun Park¹, Jun Seok Seo², Seung Chul Lee², and Sung Min Kim^{1,*}

¹Department of Medical Bio Engineering, Dongguk University-Seoul, Seoul, Republic of Korea
{syPark, smkim}@dongguk.edu

²Department of Emergency Medicine, Dongguk University Ilsan Hospital, Dongguk University-Seoul, Seoul, Republic of Korea
{drsjs, edlee}@dumc.or.kr

Abstract. The aim of this study is to suggest an artificial intelligence model to diagnosis acute appendicitis using a support vector machine (SVM). Acute appendicitis is one of the most common abdominal surgery emergencies. Various methods have been developed to diagnose appendicitis, but they have not performed well in the Middle East, Asia, or the West. A total of 760 patients were used to construct the SVM. Both the Alvarado clinical scoring system (ACSS) and multilayer neural networks (MLNN) were used to compare performance. The accuracies of the ACSS, MLNN, and SVM were 54.87%, 92.89, and 99.61%, respectively. The areas under the curve of ACSS, MLNN, and SVM were 0.621, 0.969, and 0.997 respectively. The performance of the AI model was significantly better than that of the ACSS ($P < 0.001$). We consider that the developed models are a useful method to reduce both negative appendectomies and delayed diagnoses, particularly for junior clinical surgeons.

Keywords: appendicitis, artificial intelligence, support vector machine, clinical scoring system, a receiver operating characteristics graph.

1 Introduction

Acute appendicitis is one of the most common surgical emergencies of the abdomen. The lifetime incidence of acute appendicitis is approximately 7%, and acute appendicitis is clearly treated by a surgical diagnosis [1, 2, 3]. An early diagnosis of suspected appendicitis is important for treating acute cases, as a missed or delayed acute appendicitis diagnosis is associated with high morbidity and mortality. Diagnostic imprecision can result in a high wound infection rate, high perforation rate, and high negative laparotomy rate, which ranges from 20–30% [3, 4].

Several clinical methods for early and correct diagnosis of acute appendicitis have been suggested and developed to increase diagnostic accuracy and to decrease

* Corresponding author.

negative laparotomies [4, 5, 6]. In 1986, Alvarado suggested a clinical scoring system consisting of signs, symptoms, and laboratory findings, and several clinical scoring systems have been developed and modified based on Alvarado's clinical scoring system (ACSS) [4]. However, several researchers have shown that the performance of these clinical scoring systems is insufficient for diagnosis. Image analysis methods including computed tomography (CT), and ultrasound (US) have significantly higher performance than other diagnostic methods, but they have some disadvantages [2, 5]. The quality of a CT image is highly related to radiation exposure and the diagnostic performance of US is highly dependent on the operator and cannot be used during off-hours. Moreover, the image analysis method occasionally becomes the cause for a delayed diagnosis of acute appendicitis.

More recently, artificial intelligence (AI) methods have been applied to diagnose or predict disease [7, 8, 9, 10]. Among AI algorithms, the support vector machine (SVM), which is derived from statistical learning theory by Vapnik [7], has been increasingly investigated as an aid for clinical decisions and has shown good diagnostic performance in various clinical fields, particularly cancer prediction including cervical [8], prostate [9], and breast cancers [10]. Because of the properties of SVM its outstanding performance with a small data set, relationship of nonlinear and high dimension in input data [7]. SVM can help with diagnostic guidelines and minimize possible errors in complicated diseases, particularly for inexperienced clinicians. Most importantly, AI methods including SVM can reduce the time for a diagnosis.

In this study, we used the SVM method to diagnose acute appendicitis. We compared the performance of a multilayer neural network (MLNN), and the ACSS. The MLNN method is commonly used in pattern recognition problems and shows good performance in clinical fields. The aim of this study is to propose an AI method for diagnosing acute appendicitis in patients with abdominal pain. The results showed better diagnostic performance for the AI method than that of the ACSS.

2 Methods

2.1 Patient Data

We recruited patients who presented to the emergency department of Dongguk University Hospital with abdominal pain between August 2011 and July 2012. This trial was approved by the Institutional Review Board of Dongguk University Hospital. The clinical protocol including history, physical examination, and laboratory tests was designed using the standardized terminology of the World Federation of Gastroenterology and the ACSS. Patients were allocated into three categories of no appendicitis (NA), normal appendicitis (NorA), and acute appendicitis (AA).

2.2 Alvarado Clinical Scoring System

The ACSS consists of nine factors (1 point for migration of pain to the right lower quadrant, anorexia, nausea/vomiting, rebound tenderness, elevated temperature $\geq 37.5^{\circ}\text{C}$,

and neutrophil shift to the left $> 75\%$ and 2 points for tenderness in the right lower quadrant and leukocytosis (white blood cells $> 10,000/\mu\text{l}$). The ACSS has a range of 0–10 points and is used to predict the presence or absence of acute appendicitis. The patients were allocated into three groups; ≤ 5 points for NA, ≥ 6 points and ≤ 7 points for NorA, and ≥ 8 points for AA.

2.3 Artificial Intelligence Method

We designed the structure of the SVM in two steps and each step consisted of a SVM method as shown in Fig. 1. The patients were first classified into NA and appendicitis groups. Patients in the appendicitis group were classified into the NorA and AA groups. Each SVM consisted of three spaces (input space, feature space, and output space). The input space had 10 features, including male/female, age, migration of pain to the right lower quadrant, anorexia, nausea/vomiting, rebound tenderness, tenderness in the right lower quadrant, body temperature, neutrophil percentage, and leukocyte count. The features for the input layer were binary (i.e., 0 for no rebound tenderness, and 1 for rebound tenderness) except the continuous data (i.e., body temperature, neutrophil percentage, and leukocyte count). The radial basis function network was used in the feature space, which is commonly used and shows excellent performance, for the mercer kernel as shown (1)

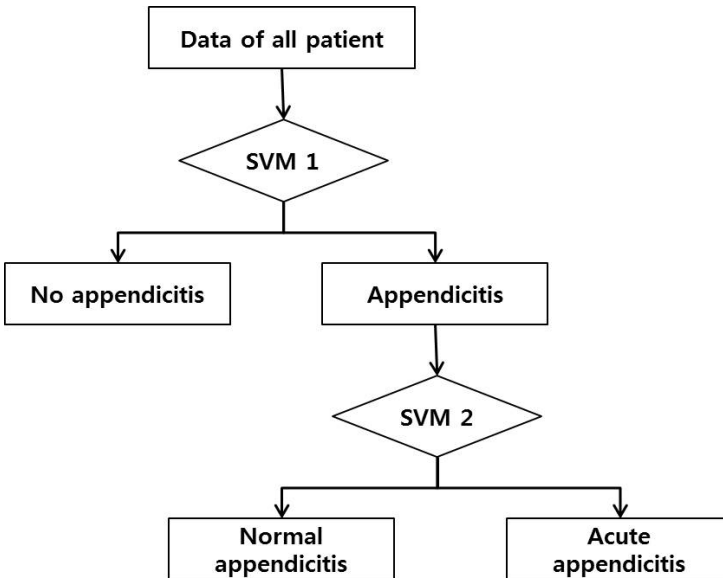


Fig. 1. Diagnosis model of acute appendicitis based on support vector machine (SVM) classifiers

$$k(x^i, x_j^i) = \exp\left(-\frac{1}{2\sigma^2} \|x^i - x_j^i\|^2\right) \quad (1)$$

where x_j^i is the feature of each step ($i = 1, 2$ for first step, and second step, respectively, and $j = 1, 2, \dots, 10$ for features). The output space consisted of two groups for each step (NA and appendicitis groups for the first step, and the NorA and AA groups for the second step). The MLNN also used one input layer with 10 features, two hidden layers, and one output layer consisting of three categories; NA, NorA, and AA. The activation and net functions in MLNN were sigmoidal and linear, respectively.

The SVM and MLNN models were developed in three phases of training, validation, and testing. The patient cases were randomly assigned to one of three phases (60%, 20%, and 20% for training, validation, and testing, respectively). The structures of both the SVM and MLNN were constructed using the MATLAB (MathWorks Inc., Ver. 2012b.) program. Detailed information relating the SVM and MLNN can be found in the neural network toolbox section of the MATLAB documentation.

2.4 Statistical Analysis

We used two methods to measure the performances of the SVM, MLNN, and ACSS. The first algorithm was related to a confusion matrix, including sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. The second algorithm for the evaluation used a receiver operating characteristics (ROC) graph, and the area under the ROC curve (AUC). The AUC value indicated the performance of the diagnostic method in a range of 0–1 (excellent, > 0.9 ; good, $0.8–0.9$; moderate, $0.7–0.8$; poor, < 0.7). Differences between variables including performance were assessed by Wilcoxon's rank-sum test, the Kruskal–Wallis test, and the χ^2 test for continuous variables and categorical variables respectively. A $P < 0.05$ was considered a significant difference.

3 Results

A total of 760 patients were enrolled from August 2011 to July 2012 in the emergency department of Dongguk University Hospital. In total, 429 (56.45%) patients were in the NA group and 331 (43.55%) were in the appendicitis group including 237 (31.18%) in the NorA group and 94 (12.37%) in the AA group (Table 1). Mean age was 29.57 years, 30.59 years, and 31.31 years for the NA, NorA, and AA groups, respectively ($P = 0.427$). The number of female patients (294, 122, and 49 for NA, NorA, and AA, respectively) was significantly higher than that of male patients (135, 122, and 45 for NA, NorA, and AA) in the NA group ($P < 0.001$).

Table 1. Results of 760 patients for suspected appendicitis

	No appendicitis	Appendicitis		<i>P</i> value
		Normal appendicitis	Acute appendicitis	
No. of subjects	429	237	94	<0.001 [†]
Male : Female	135:294	115:122	23:71	<0.001 [‡]
Age-mean(years) (min.-max.)	29.57 (0-62)	30.59 (10-69)	31.31 (14-72)	0.427 [§]
Leucocyte-mean±SD (×10 ⁶ /mm ³)	3.58±5.58	7.63±5.48	8.62±7.21	<0.001 [†]
Neutrophil-mean±SD (%)	66.95±17.22	72.21±18.08	76.42±16.50	<0.001 [†]

[†]Kruskal-Wallis test, [§] Wilconxon's rank-sum test, [‡] χ^2 -test, and SD: Standard deviation.

The laboratory test values, expressed as mean \pm standard deviation, are shown in Table 1. The leukocyte counts were 3.58 ± 5.58 , 7.63 ± 5.48 , and 8.62 ± 7.21 for the NA, NorA, and AA groups, respectively, and the neutrophil percentages were 66.95 ± 17.22 , 72.21 ± 18.08 , and 76.42 ± 16.50 , respectively (both $P < 0.001$).

The ACSS performance was the lowest (77.86%, 25.08%, 46.63%, 57.39%, and 54.87% for specificity, sensitivity, positive predictive value, negative predictive value, and accuracy, respectively) on all parameters, whereas the diagnostic method using the SVM had the highest values (99.53%, 99.70%, 99.40%, 99.77%, and 99.61% for specificity, sensitivity, positive predictive value, negative predictive value, and accuracy, respectively) (Table 2). Although the MLNN method showed lower performance compared with that of the SVM, the performance of MLNN was significantly higher than that of the ACSS.

Table 2. Performance of diagnosis methods

	Specificity (%)	Sensitivity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
ACSS	77.86	25.08	46.63	57.39	54.87
MLNN	95.10	90.03	93.42	92.52	92.89
SVM	99.53	99.70	99.40	99.77	99.61

ACSS: Alvarado clinical scoring system, MLNN: Multilayer neural network, SVM: Support vector machine.

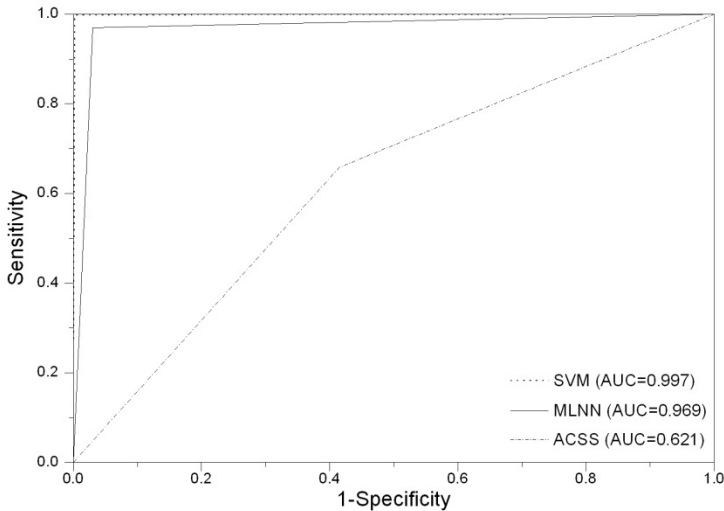


Fig. 2. A receiver operating characteristics (ROC) graph and the area under an ROC curve (AUC) for support vector machine (SVM), Multilayer neural network (MLNN), and Alvarado clinical scoring system (ACSS)

The ROC graphs for the three methods for diagnosing AA are shown in Fig. 2. The method accuracies in decreasing order were SVM (AUC, 0.997), MLNN (AUC, 0.969), and ACSS (AUC, 0.621). Taken together, these results confirm that the AI method had significantly better performance than that of the ACSS.

4 Discussion

Appendicitis is a common abdominal disease in the emergency department. Acute appendicitis, which is considered advanced appendicitis, can lead to death. Although various diagnostic methods have been suggested and have shown good performance, problems have recently come to the fore for the main diagnostic methods such as unstable performance of ACSS, un-usability of ultrasound, and poor safety of CT. We suggested a novel solution using AI methods such as SWM and MLNN.

We enrolled 760 patients with abdominal pain, and the total rate of female patients was significantly higher than that of males (38.82% vs. 61.18%, $P < 0.001$) within the NA (31.47% vs. 68.53%, $P < 0.001$), NorA (40.51% vs. 59.49%, $P < 0.05$), and AA groups (41.69% vs. 58.31%, $P < 0.05$). Hale et al. reported appendectomies in 4,950 patients that were collected over a 12-month period. They noticed that the number of normal appendicitis cases in female patients was significantly higher than that of male patients (19% vs. 9%) [11]. This is because ectopic pregnancy and mittelschmerz in women mimic appendicitis. In the present study, the number of females with appendicitis was also significantly higher than that of males. We thought that some of the women in the appendicitis group may be confused with dysmenorrhea, and many

female patients in this study actually were dysmenorrhic (57.14%, 58.87%, and 80.28% for NA, NorA, and AA groups, respectively).

The performance of the ACSS in this study was the lowest of the three methods. This is because of low value of leukocyte count, which plays an important role in the diagnosis of appendicitis, particularly in women and children. Although leukocyte counts were significantly different ($P < 0.001$) among the three groups, the mean leukocyte count ($8.62 \times 10^6/\text{mm}^3$) was lower than $10 \times 10^6/\text{mm}^3$, which is a threshold value to receive 2 points in the ACSS. Previous studies have reported that mean leukocyte counts in appendicitis groups are $> 10 \times 10^6/\text{mm}^3$ and that the ACSS performed well [1, 7, 12]. We cannot explain why our leukocyte counts were lower compared with those of previous studies. This phenomenon should be investigated in a future study.

However, the performance of the AI method was higher compared with that of the ACSS ($P < 0.001$). de Dombal et al. reported in 1972 that the performance of a computer-aid diagnostic system was significantly higher than that of clinicians [13]. Many researchers have used the AI method to diagnose disease and have shown good performance [8, 9, 10]. The weakness of the AI method is that it is highly dependent on the database (i.e., number of patients), but AI remains the best approach to solve nonlinear problems such as disease diagnosis. To overcome this weak point, the SVM is commonly used to solve nonlinear problems due to kernel function, which converts simple feature dimensions (or input data) into high dimensions [14, 15, 16, 17]. In this study, SVM had better performance on all measurements than that of the MLNN.

5 Conclusion

The AI model showed excellent performance to diagnose acute appendicitis without the need for an expert surgeon. This model may help reduce both negative appendectomies and a delayed diagnosis, particularly for junior surgeons.

Acknowledgements. This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry of Health & Welfare, Republic of Korea. (A102058).

References

1. Kim, E., Subhas, G., Mittal, V.K., Golladay, E.S.: C-reactive protein estimation does not improve accuracy in the diagnosis of acute appendicitis in pediatric patients. *Int. J. Surg.* 7(1), 74–77 (2009)
2. Fergusson, J., Hitos, K., Simpson, E.: Utility of white cell count and ultrasound in the diagnosis of acute appendicitis. *ANZ. J. Surg.* 72(11), 781–785 (2002)
3. Petroianu, A.: Diagnosis of acute appendicitis. *Int. J. Surg.* 10(3), 115–119 (2012)
4. Alvarado, A.: A practical score for the early diagnosis of acute appendicitis. *Ann. Emerg. Med.* 15(5), 557–564 (1986)

5. Pritchett, C., Levinsky, N., Ha, Y., Dembe, A., Steinberg, S.: Management of acute appendicitis: the impact of CT scanning on the bottom line. *J. Am. Coll. Surg.* 210(5), 699–705 (2010)
6. Yang, H., Wang, Y., Chung, P., Chen, W., Jeng, L., Chen, R.: Laboratory tests in patients with acute appendicitis. *ANZ. J. Surg.* 76(1-2), 71–74 (2006)
7. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer (2010)
8. Hu, X., Cammann, H., Meyer, H., Miller, K., Jung, K., Stephan, C.: Artificial neural networks and prostate cancer-tools for diagnosis and management. *Nat. Rev. Urol.* 10, 1–9 (2013)
9. Mat-Isa, N., Mashor, M., Othman, N.: An automated cervical pre-cancerous diagnostic system. *Artif. Intell. Med.* 42(1), 1–11 (2008)
10. Shi, H., Tsai, J., Chen, Y., Culbertson, R., Chang, H., Hou, M.: Predicting two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. *Breast Cancer Res. Treat.* 135(1), 221–229 (2012)
11. Hale, D., Molloy, M., Pearl, R., Schutt, D., Jaques, D.: Appendectomy: a contemporary appraisal. *Ann. Surg.* 225(3), 252–261 (1997)
12. Pouget-Baudry, Y., Mucci, S., Eyssartier, E., Guesdon-Portes, A., Lada, P., Casa, C., et al.: The use of the Alvarado score in the management of right lower quadrant abdominal pain in the adult. *J. Visc. Surg.* 147(2), e40–e44 (2010)
13. de Dombal, F., Leaper, D., Staniland, J., McCann, A., Horrocks, J.: Computer-aided diagnosis of acute abdominal pain. *Br. Med. J.* 2(5804), 9–13 (1972)
14. Chen, L., Xuan, J., Riggins, R., Clarke, R., Wang, Y.: Identifying cancer biomarkers by network-constrained support vector machines. *BMC Syst. Biol.* 5, 161 (2011)
15. Mourao-Miranda, J., Reinders, A., Rocha-Rego, V., Lappin, J., Rondina, J., Morgan, C., et al.: Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol. Med.* 42(5), 1037–1047 (2012)
16. Lancashire, L., Roberts, D., Dive, C., Renehan, A.: The development of composite circulating biomarker models for use in anticancer drug clinical development. *Int. Cancer* 128(8), 1843–1851 (2011)
17. Lv, G., Cheng, H., Zhai, H., Dong, L.: Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr. Power Syst. Res.* 75(1), 9–15 (2005)

On Extracting Important User Preferences

Sylvia Encheva

Stord/Haugesund University College, Bjørnsonsg. 45, 5528 Haugesund, Norway
sbe@hsh.no

Abstract. User preferences are very important in planning organization's future activities. Involvement of a decision support system can considerably shorten the time used for planing and for cost calculations by presenting what is the current status and what can be expected in relation to new users' preferences. The latter can be further combined with existing quality guidelines and standards for each particular organization.

Keywords: Many valued logics, lattices, decision making.

1 Introduction

Public and private organizations need to follow numerous quality guidelines and standards. One way to improve their performance is by enhancing awareness about users' preferences. This can be done by inviting users to fill in some forms reflecting on their preferences. While such inquiries may reveal sufficient information about current situations they will not provide good indications about preferences of new users or new products.

As a way to improve organizations' planing and performance we suggest use of a decision support system. This can considerably shorten the time used for planing and for cost calculations by presenting what is the current status and what can be expected in relation to new demands.

2 Background

Many valued logics have been applied in solving numerous theoretical and practical problems where uncertainty is involved, [8]. Both possibility theory and possibilistic logic often involve many-valued calculi, [3].

A three-valued logic, known as Kleene's logic is developed in [9] and has three truth values, truth, unknown and false, where unknown indicates a state of partial vagueness. These truth values represent the states of a world that does not change.

The semantic characterization of a four-valued logic for expressing practical deductive processes is presented in [1] and [2]. In most information systems the management of databases is not considered to include neither explicit nor hidden inconsistencies. In real life situation information often come from different contradicting sources. Thus different sources can provide inconsistent data while

deductive reasoning may result in hidden inconsistencies. The idea in Belnap’s approach is to develop a logic that is not that dependable of inconsistencies.

The Belnap’s logic has four truth values 'T, F, Both, None'. The meaning of these values can be described as follows:

- an atomic sentence is stated to be true only (T),
- an atomic sentence is stated to be false only (F),
- an atomic sentence is stated to be both true and false, for instance, by different sources, or in different points of time (Both), and
- an atomic sentences status is unknown. That is, neither true, nor false (None).

Five valued logic as in [7], [8] is presented in Fig.1. The five-valued logic introduced in [7] is based on the following truth values: unknown or undefined, possibly known but consistent, false, true, and inconsistent.

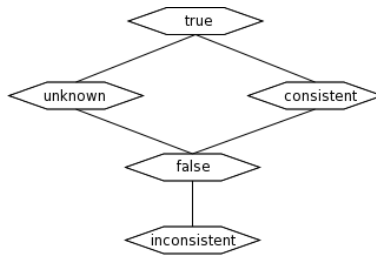


Fig. 1. The five truth values

Fuzzy multiple criteria decision-making methodology was used by [6] to develop a practical model for business purpose. Improved score functions measuring the degree of suitability of a set of alternatives, with respect to a set of criteria based on vague values is discussed in [10]. An algorithm for score functions is also introduced by taking into account the effect of an unknown degree (hesitancy degree) of the vague values on the degree of suitability to which each alternative satisfies the decision makers’s requirement. In [4] the authors develop a forecasting framework based on the fuzzy multi-criteria decision making (approach to help organizations build awareness of the critical influential factors on the success of knowledge management implementation, measure the success possibility of knowledge management projects, as well as identify the necessary actions prior to embarking on conducting knowledge management.

A lattice is a partially ordered set, closed under least upper and greatest lower bounds. The least upper bound of x and y is called the join of x and y , and is sometimes written as $x + y$; the greatest lower bound is called the meet and is sometimes written as $x \dot{y}$, [5, 11].

3 Preferences

In this section we are combining different user preferences and illustrating what could be expected with respect to demonstrating preferences in new situations. We consider

cases with both complete and incomplete information and draw conclusions based on theories from many valued logics.

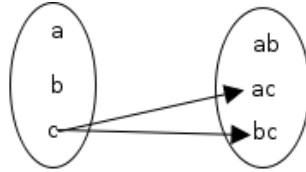


Fig. 2. High level satisfaction in product *c* only

Medium level satisfaction in product *c* and lack of information about products *a* and *b* implies expectation of low level satisfaction in new products having features from either *a* and *c* or from *b* and *c*, Fig. 2.

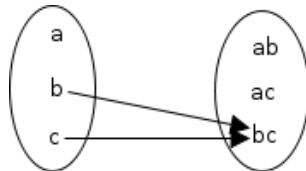


Fig. 3. High level satisfaction in products *b* and *c* only

Medium level satisfaction in products *b* and *c* and lack of information about about product *bc* implies expectation of low level satisfaction in *bc*, Fig. 3.

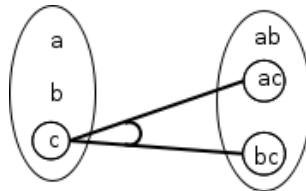


Fig. 4. High level satisfatio in products *ac* and *bc* only

High level satisfaction in products *ac* and *bc* implies expectation of high level satisfaction in product *c*, Fig. 4.

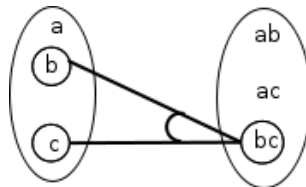


Fig. 5. High level satisfaction in products *b* and *c* only

High level satisfaction in products b and c and implies expectation of high level satisfaction in product bc , Fig. 5.

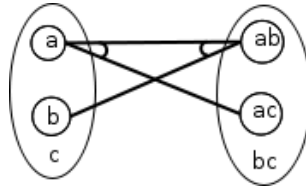


Fig. 6. High level satisfaction in products b and ac only

High level satisfaction in products b and ac implies expectation of medium level of satisfaction in products a and ab , Fig. 6.

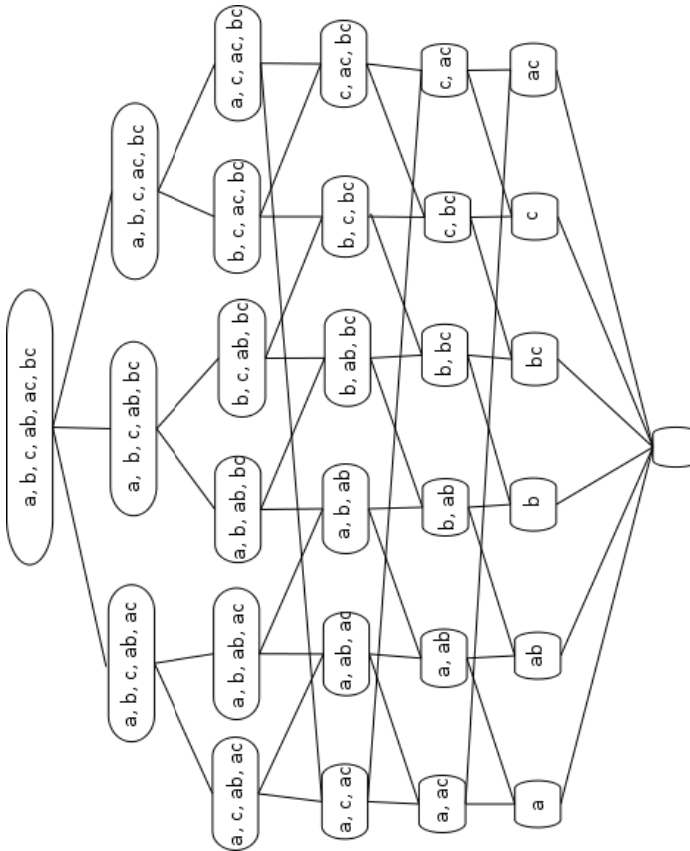


Fig. 7. A lattice illustrating all dependencies

All dependencies are incorporated in Fig. 7. Lattice theory and higher-order logic can be applied for developing a functional decision support system.

4 Conclusion

The aim of this work was to address the problem of drawing conclusions about users' preferences. In order to complete the task we have been relying on some available information and applying rules from the theory of many valued logics. In future work we plan to address problems related to development of user preferences' testing.

References

1. Belnap, N.J.: How a computer should think. In: Contemporary Aspects of Philosophy. Proceedings of the Oxford International Symposia, Oxford, GB, pp. 30–56 (1975)
2. Belnap, N.J.: A useful four-valued logic. In: Dunn, J.M., Epstein, G. (eds.) Modern Uses of Multiple-valued Logic, pp. 8–37. Reidel Publishing Co., Dordrecht (1977)
3. Bolc, L.: Many-Valued Logics 2: Automated Reasoning and Practical Applications. Springer (2003)
4. Chang, T.H., Wang, T.C.: Using the fuzzy multi-criteria decision making approach for measuring the possibility of successful knowledge management. *Information Sciences* 179, 355–370 (2009)
5. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)
6. Ding, J., Liang, G.: Using fuzzy MCDM to select partners of strategic alliances for liner shipping. *Inf. Sciences*, 197–225 (2005)
7. Ferreira, U.: A Five-valued Logic and a System. *Journal of Computer Science and Technology* 4(3), 134–140 (2004)
8. Fitting, M., Orłowska, E.: Beyond Two: Theory and Applications of Multiple-Valued Logic. *STUDFUZZ*, vol. 114. Springer, Heidelberg (2003)
9. Kleene, S.: Introduction to Metamathematics. D. Van Nostrand Co., Inc., New York (1952)
10. Ye, J.: Improved method of multicriteria fuzzy decision-making based on vague sets. *Computer-Aided Design* 39, 164–169 (2007)
11. Wille, R.: Concept lattices and conceptual knowledge systems. *Computers Math. Applications* 23(6-9), 493–515 (1992)

Object Retrieval Scheme Using Color Features in Surveillance System

Su-wan Park^{*}, JeongNyeo Kim, and Jong Wook Han

Cyber Security Research Department, Electronics and Telecommunication Research Institute,
Korea
{parksw10, jnkim, hanjw}@etri.re.kr

Abstract. In this paper, we have described the object retrieval scheme based on color for video surveillance that is influenced by the different light changes and overlapping/non-overlapping view cameras setting. The proposed video retrieval scheme separates object into top and bottom, and extracted dominant colors from each region. Each dominant color includes hue, saturation, value in HSV space and proportion of hue color. In addition, it uses the various threshold values and pre-defined weights based on the experiment and processes the similarity measurement to order the search results. Therefore, our retrieval scheme provides the delicateness and the robustness in varying surveillance environmental conditions. As well, it can be applied in real-time surveillance system.

Keywords: CCTV Surveillance system, Video Surveillance, Video Retrieval.

1 Introduction

Recently, the number of criminal is increasing very fast. As the number of installed surveillance cameras increase, the identification of objects in many views is an important topic. Especially, the case that illustrated recently with the identification of the bombers of the Boston marathon based on large sets of multimedia data provided by the public emphasizes the necessity of video object retrieval technology that can help to quickly go through hours of CCTV video and analyze the features on suspects. The early video searching engines rely on text-based annotations and descriptions of the video clips input by the clip owners. However, these approaches need extensive time and subjective notion to annotate and describe those video clips. Thus, automated video indexing and retrieval schemes have become one of the key issues recently. The video retrieval system can use the object features such as the color, size, object type and contour for video indexing, and it performs some similarity measurements between video shots or objects for retrieval scheme [1]. Furthermore, the recent works been performed to study the retrieval based on the object characteristics such as clothing to identify the suspects. Because color information has the strong temper, while size, speed and contour are changeable in the difference environment. However, these works still have the limitation by issue of color constancy including changing lighting conditions and multiple-cameras setting.

^{*} Corresponding author.

2 Previous Work

In video analysis, the moving object appeared in the video scenes is first detected by background subtraction [2]. The detected objects can be classified as either a person or vehicle based on shape features and motion features. The object is then tracked by the commonly used Kalman filter tracker [3]. In addition to the object tracking in a scene, the object tracking methods between IP cameras should be considered in surveillance system [4]. In particular, the object tracking between IP cameras with non-overlapping view is one of the biggest challenges lately [5]. The sort of object feature that can be utilized for tracking or retrieval in the surveillance system based on non-overlapping view is much more limited than the features of overlapping view, thus many studies mainly uses color information of object. In this paper, we will focus on the retrieval scheme only using color, even though there are many different retrieval approaches by environmental factors of surveillance system.

S. Calderara et.al.[6] proposed the architecture for people retrieval in multi-camera surveillance with the non-overlapping views, but it applies the video post-analysis including the learning phase to estimate the probability density function of its colors. A. J. Perrott et.al.[7] utilized the MPEG-7 standard to provide content-based retrieval of digital CCTV recordings in real-time as enforcing a standard Description Definition Language (DDL). J. Annesley et.al.[8] presented the results to evaluate the effectiveness of MPEG7 Color descriptors in visual surveillance retrieval problems. This dominant color descriptor clusters the data using the LUV color space and outputs the number of clusters (up to 9), spatial homogeneity, and each clusters with a color, variance and percentage value. Thus, this approach is expected to have the performance and computational requirements. Y. Tian et.al.[9] explored video parsing techniques that automatically extract index data from video, indexing which stores data in relational tables, retrieval which uses SQL queries to retrieve events of interest and the software architecture that integrates these technologies. However, it detects and tracks human faces mainly and object color in the vehicle searching still is determined by 6 colors of HSI space and the relative amount of black and white. J. S-C. Yuk et.al.[10] presented a surveillance video indexing and retrieval system based on object features similarity measurement, and it use real-time automatic indexing methodology and specifies the query using image or a sketch of the desired objects. However, it uses only the hue component of the HSV color space to be relatively insensitive to lighting conditions and saturation level of different cameras.

3 The Proposed Retrieval Scheme

3.1 System Architecture for Retrieval

The proposed retrieval scheme is aimed at searching a specific object using color information in the various surveillance circumstances by the different light changes and overlapping/non-overlapping view cameras setting.

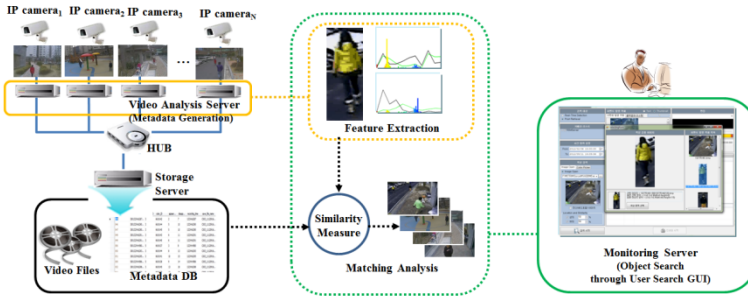


Fig. 1. Overview of the proposed retrieval scheme

Fig. 1 shows the architecture and modules for the proposed retrieval scheme. In this paper, we assume that the surveillance system consists of the IP cameras, video analysis servers, storage server and monitoring server. Our object retrieval scheme requires the metadata generation module and the object search module.

The metadata generation module serves the function indexing each object in video as collecting the features of each object. It first analyzes videos received from multiple surveillance cameras in real-time, and then generate object metadata using features of object. The object features can contain the properties such as object_ID, camera_ID, appearance_time, disappearance_time, video_file_name, top_color[n], bottom_color[n], location_coordinate and object_size. Since the proposed retrieval system focuses on color-based object search, especially, the color extraction point of time should be considered carefully. It is because the accuracy of extracted color and proportion affects the search results. If you use the average value of colors extracted from the frames that includes the consistently same object, it should be based on the accuracy of object tracking and object ID management technique. Thus, we use one key frame satisfying the prescribed object size from shots, and extract only once the features from the object in key frame. The color feature extraction method in detail will be addressed in the following section. Lastly the generated object metadata are stored in DB of storage server.

The object search module serves the function searching a specific object selected by user within many stored video files. User first inputs the search conditions such as the IP cameras, date, time and color through a search GUI application. The search conditions are translated into a query, and transferred to the object search module. The object search module first analyses the query, and then it processes object matching technique that finds objects having similarity between the query and the object metadata stored in DB. Here, we use the threshold values set by the query analysis result to fine the related metadata, and measure the similarity between query and metadata using the distance of colors and proportion. It will be addressed in the following section in detail.

3.2 Color Feature Extraction and Matching

In this section we propose the color feature extraction method for collecting color information consisting of each object and the object matching method for measuring the similarity between a specific user query and the stored object metadata in detail.

Color Feature Extraction

In this scheme, we use the dominant color descriptor that means the representative colors in an object region. The dominant colors are extracted from two separate regions since the object clothing is segmented into top and bottom items. The procedure of dominant color extraction method consists of the following steps:

- 1) RGB values consisting of the region of interest (ROI) are converted into HSV color space [9] that is represented by hue, in the range of 0 to 360, saturation and value (brightness), in the range of 0 to 100 respectively.

$$RGB: (R_i, G_i, B_i) \rightarrow HSV: (H_i, S_i, V_i), \quad i \in \{1, 2, \dots, K\},$$

where K is the number of pixels consisting of ROI.

- 2) HSV space is quantized into a small number of colors. We divided hue space into M(=18) colors to provide the detailed search capability. The hues consisting of ROI are accumulated in histogram of 18 colors, and the average of saturation and values corresponding to hue are calculated respectively. Each quantized hue point QH_j includes factors as follows:

$$QH_j = (N_{H_j}, A_{S_j}, A_{V_j}, P_{H_j}), \quad j \in \{1, 2, \dots, M\}$$

$$A_{S_j} = \frac{\sum_1^K S_i}{N_{H_j}} \quad \text{if } H_i \in QH_j, \quad A_{V_j} = \frac{\sum_1^K V_i}{N_{H_j}} \quad \text{if } H_i \in QH_j, \quad P_{H_j} = \frac{N_{H_j}}{\sum_1^M N_{H_j}}$$

where N_{H_j} is the number of pixels included in QH_j , A_{S_j} is the average of saturations corresponding to QH_j , A_{V_j} is the average of values corresponding to QH_j , and P_{H_j} is the percentage of QH_j in ROI.

- 3) If N_{H_j} is in top N, H_j are determined as one of the dominant colors of the ROI. The formula can be simplified and will be displayed as follows:

$$D^l = (H_D^l, S_D^l, V_D^l, P_D^l) \quad \text{if } N_{H_j} \text{ in Top N}, \quad l = (1, 2, \dots, N),$$

where x^l means the value x of top l -th.

- 4) For the ROI ranges of top and bottom, step 1) to step 3) are processed again.

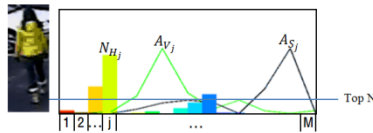


Fig. 2. Object color feature extraction method using the histogram

Object Matching

In this section, we address the method to judge whether the similarity exists between an input query and each object metadata in the database using the dominant colors. The retrieval results are then presented with the measured similarity value. The procedure can be split in two parts greatly. One is the analysis of input query, and the other is the metadata search in the database based on criteria that is given by query analysis.

1) Analysis of input query

- A. The dominant color D^l of query is checked whether achromatic color or chromatic color.

$$\text{Color type} \begin{cases} \text{Achromatic color,} & \text{if } S_D^l < s\delta_1 \text{ and } V_D^l < v\delta_1 \\ \text{Chromatic color,} & \text{otherwise} \end{cases}$$

- B. If D^l is achromatic color, D^l is again checked whether black, white, or gray by conditions of saturation and value according to threshold values ($v\theta_1 \sim v\theta_4, s\theta_1$) in Table 1. Then, the number of achromatic color increases by one and it is displayed as N_a . Otherwise, D^l is counted as the number of chromatic color and is displayed as N_c . Then, N_{a+c} means the sum of N and N_c .

Table 1. The threshold values to distinguish black, white, or gray

Range	Saturation	Value
Black	$0 \sim 100$	$0 \sim v\theta_1$
White	$0 \sim s\theta_1$	$v\theta_2 \sim 100$
Gray	$0 \sim s\theta_1$	$v\theta_3 \sim v\theta_4$

- C. Step A to step B is performed repeatedly until l is N which is the number of the dominant colors.
- D. Through step A to step C, it determines whether the input query is single color or multi-color. In addition, it judges whether the ROI includes only achromatic color or chromatic color, or combination of achromatic color and chromatic color.

2) Metadata search

Each metadata in database is searched based on the values specified by achromatic color or chromatic color. If the dominant colors are more than one such as two chromatic colors or combination of achromatic color and chromatic color, our system processes the each search depending on the color type, then it combines the search results through the inner join operation in SQL. In addition, the results for top and bottom can be combined again. These search results can be ordered by the similarity measure depending on the color type. The metadata search criterions for each color type and similarity measure operation are as follows:

- A. If the color type of query ROI is achromatic color, it utilizes the saturation and value in HSV space to distinguish black, white, or gray. The interval of threshold values ($v\theta_1 \sim v\theta_4, s\theta_1$) in Table 1 may be overlap to

tolerate the lighting effects, and the values can be adjusted according to the similarity rate inputted by user. The dominant colors included in metadata are searched by the set values.

- B. If the color type of query ROI is chromatic color, it mainly utilizes hue in HSV space, and uses the fixed threshold value of saturation and value to avoid searching the achromatic color. The hue range can be set differently with the similarity depending on the sensitivity of the search. Suppose two dominant colors for metadata and query:

$$M^k = (H_M^k, S_M^k, V_M^k, P_M^k), k = (1, 2, \dots, N_1)$$

$$Q^l = (H_Q^l, S_Q^l, V_Q^l, P_Q^l), l = (1, 2, \dots, N_2)$$

The set values for chromatic color can be summarized as follows, and these conditions are used in combination of all or partials:

Hue	Saturation	Value
$ H_M^k - H_Q^l < h\theta_1$	$S_M^k > s\theta_2$ and $ S_M^k - S_Q^l < s\theta_3$	$V_M^k > v\theta_5$ and $ V_M^k - V_Q^l < v\theta_6$

- C. The similarity measure for achromatic color and chromatic color can be computed as:

- i. $S_k = \sum_{k=1}^{N_1} (\alpha * \beta * \gamma * P_M^k)$, if $l = 1$, where α, β and γ are the weights given according to the similarity by $|H_M^k - H_Q^l|$, $|S_M^k - S_Q^l|$ and $|V_M^k - V_Q^l|$ respectively. The weights have the values between 0 and 1, and the greater the similarity is close to 1.
- ii. $P_k = \begin{cases} S_k/P_Q^l, & S_k \leq P_Q^l \\ P_Q^l/S_k, & S_k > P_Q^l \end{cases}$
- iii. $S_{lk} = \sum_{l=1}^{N_2} P_k$
- iv. $P_{lk} = \left(\frac{S_{lk}}{N_{a+c}} \right) * 100$, where N_{a+c} means the number of dominant colors of the query.

At this time, γ can be 0 if it ignores the light effect.

4 Experiment Results

The proposed videos retrieval system is evaluated through video contents obtained by 5 multiple cameras with the non-overlapping view installed around an apartment. We first indexed the 100 objects using color feature extraction method from video frames in real-time, and the indexed object features are stored in database as object metadata. Here, we extracted respectively 3 dominant colors from two separate regions of object appearance, giving top and bottom item, after each object is recognized and segmented. Then, we performed the object retrieval using the color descriptors of query and the metadata in DB. At this time, the several experimental parameters as discussed above are set basically as follows: $s\delta_1 = 10$, $v\delta_1 = 20$, $s\theta_1 = 20$, $v\theta_1 = v\theta_3 = 20$, $v\theta_2 = v\theta_4 = 80$, $h\theta_1 = 1.0$ (between 0.8 and 1.2), $s\theta_2 = 10$, $s\theta_3 = 30$ (between 20 and 80), $v\theta_5 = 15$, and $v\theta_6 = 50$ (between 50 and 100).

Hue 6 colors and Hue 18 colors, that are similar to the references [9] and [10], can recognize identically the same objects in different cameras, while the objects of similar color in different cameras are difficult to distinguish as shown in 3th and 4th rows of Fig 3. Using only hue values such as the reference [10], in addition, has a blurred distinction in achromatic color. If the saturation value is added for object retrieval, it can be utilized to distinguish the different objects of similar color in Fig.3. However, it still does not affect to distinguish achromatic colors as shown in (c, d) of 5th row.

In many variations by lighting conditions, the proposed scheme is able to retrieve the corresponding objects in both chromatic and achromatic colors, such as the red objects in (a, b) and the white objects in (c, d). In addition, our scheme can distinguish the objects with different color, such as the yellow-green and khaki in (e, f), and red and brown in (g, h), even though they have same or similar hue value. Therefore, our scheme using the segmentalized hue together with saturation and value in HSV space for the retrieval scheme provides the search rate of more than 80% and has better result than references mentioned.

Furthermore, our scheme provides the compactness of object data to index object in database as using the minimal color information. It also requires the low computational cost in the color extraction and similarity measure, compared with the reference [8], so that it can be performed in real-time.

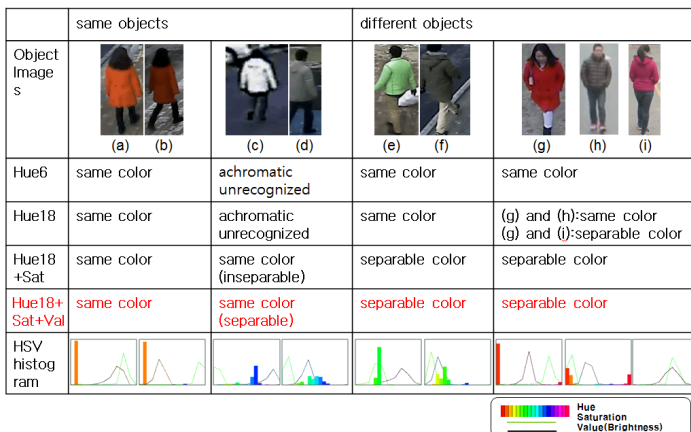


Fig. 3. Objects and their color histogram obtained from 5 multiple cameras, (a)~(e), and the comparison of retrieval schemes using hue, saturation or value

5 Conclusions

In this paper, we have described the object retrieval scheme based on color for video surveillance that is influenced by the different light changes and overlapping/non-overlapping view cameras setting. Since the video retrieval scheme searches moving objects wearing clothes which are similar to the query color, we separated object into top and bottom, and extracted dominant colors from each region. Here, each dominant

color includes hue, saturation, value in HSV space and proportion of hue color. For query analysis and the search of object metadata in DB, in addition, our scheme uses the various threshold values and weights previously set based on the experiment. Finally, we ordered the search results through the similarity measurement. Therefore, our scheme provides the robustness in surveillance environmental factors while it enables to search a little more delicate than the existing scheme. As well, it can be applied in real-time surveillance system and it can also be used for the legal evidence-video generation later as utilizing the search results.

Acknowledgment. This work was supported by the IT R&D program (10043959, Development of EAL 4 level military fusion security solution for protecting against unauthorized accesses and ensuring a trusted execution environment in mobile devices) of KEIT/MOTIE, Korea.

References

1. Patel, B.V., Meshram, B.B.: Content based video retrieval systems. *International Journal of UbiComp* 3(2), 13–30 (2012)
2. Cristani, M., Farenzena, M., Bloisi, D., Murino, V.: Background Subtraction for Automated Multisensor Surveillance: a Comprehensive Review. *EURASIP Journal on Advances in Signal Processing*, Article No. 43 (February 2010)
3. Hsia, K.H., Lien, S.F., Su, J.P.: Moving Target Tracking Based on CamShift Approach and Kalman Filter. *International Journal of Applied Mathematics & Information Sciences* 7(1), 193–200 (2013)
4. Hwang, T., Cho, S., Park, J., Choi, K.: Object Tracking for a Video Sequence from a Moving Vehicle: A Multi-modal Approach. *ETRI Journal* 28(3), 367–370 (2006)
5. Montcalm, T., Boufama, B.: Object Inter-camera Tracking with Non-overlapping Views: A New Dynamic Approach. In: *Proceedings of the 2010 Canadian Conference on Computer and Robot Vision*, pp. 354–361 (June 2010)
6. Calderara, S., Cucchiara, R., Prati, A.: Multimedia Surveillance: Content-based Retrieval with Multicamera People Tracking. In: *Proceedings of the ACM International Workshop on VSSN 2006*, pp. 95–100 (2006)
7. Perrott, A.J., Lindsay, A.T., Parkes, A.P.: Real-time multimedia tagging and content-based retrieval for CCTV surveillance systems. In: *Proceeding on SPIE*, vol. 4862 (July 2002)
8. Annesley, J., Orwell, J., Renno, J.P.: Evaluation of MPEG7 color descriptors for visual surveillance retrieval. In: *Proceedings of the International Conference on Computer Communications and Networks*, pp. 105–112 (2005)
9. Tian, Y., Hampapur, A., Brow, L., Feris, R., Lu, M., Senior, A.: Event Detection, Query, and Retrieval for Video Surveillance. *Artificial Intelligence for Maximizing Content Based Image Retrieval* (2009)
10. Yuk, J.S.-C., Wong, K.-Y.K., Chung, R.H.-Y., Chow, K.P., Chin, F.Y.-L., Tsang, K.S.-H.: Object-based surveillance video retrieval system with real-time indexing methodology. In: Kamel, M.S., Campilho, A. (eds.) *ICIAR 2007*. LNCS, vol. 4633, pp. 626–637. Springer, Heidelberg (2007)

Multi Criteria Decision Making Related to Services

Sylvia Encheva

Stord/Haugesund University College, Bjørnsonsg. 45, 5528 Haugesund, Norway
sbe@hsh.no

Abstract. A lot of data mining techniques are develop to handle large data sets. When applied on small data sets however they perform poorly. More often than not conclusions have to be drawn from relatively small data sets due to various reasons. Rough sets approximations can be applied in such situations since they do not need a critical amount of data in order to provide reliable results.

Keywords: Rough sets, evaluations, intelligent systems.

1 Introduction

Missing and contradictory data has been omitted nearly without hesitation from scientific investigations a few decades ago being regarded as a distraction. Obviously this implies partially correct conclusions since a lot of interesting dependencies can not be reviled. Use of Boolean logic in particular limits system's responses to true or false and cannot therefore recognize other occurrences like f. ex partially correct or incomplete information about services. Boolean logic appears to be quite sufficient for most everyday reasonings, but it is certainly unable to provide meaningful conclusions in presence of inconsistent and/or incomplete input, [4]. This problem can be resolved by applying methods from the theory of rough sets approximations.

In this work we are focussing on services' evaluations being subjects of cooperative decision making.

2 Background

A lattice is a partially ordered set, closed under least upper and greatest lower bounds. The least upper bound of x and y is called the join of x and y , and is sometimes written as $x + y$; the greatest lower bound is called the meet and is sometimes written as $x \dot{\bar{y}}$, [6].

A *context* is a triple (G, M, I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively [1], [6], and [15].

For $A \subseteq G$ and $B \subseteq M$, define

$$A' = \{m \in M \mid (\forall g \in A) \ gIm\}, \quad B' = \{g \in G \mid (\forall m \in B) \ gIm\}$$

where A' is the set of attributes common to all the objects in A and B' is the set of objects possessing the attributes in B .

A *concept* of the context (G, M, I) is defined to be a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The *extent* of the concept (A, B) is A while its *intent* is B . A subset A of G is the extent of some concept if and only if $A'' = A$ in which case the unique concept of the which A is an extent is (A, A') . The corresponding statement applies to those subsets $B \in M$ which is the intent of some concepts.

The set of all concepts of the context (G, M, I) is denoted by $\mathbf{B}(G, M, I)$. $\langle \mathbf{B}(G, M, I); \leq \rangle$ is a complete lattice and it is known as the *concept lattice* of the context (G, M, I) .

From classical stand point of view a concept is well defined by a pair of intention and extension. Existence of well defined boundaries is assumed and an extension is uniquely identified by a crisp set of objects. In real life situations one has to operate with concepts having grey/gradual boundaries, like f. ex. partially known concepts, [16], undefinable concepts, and approximate concepts, [9].

Rough Sets were originally introduced in [14]. The presented approach provides exact mathematical formulation of the concept of approximative (rough) equality of sets in a given approximation space. An *approximation space* is a pair $A = (U, R)$, where U is a set called universe, and $R \subset U \times U$ is an indiscernibility relation.

Equivalence classes of R are called *elementary sets* (atoms) in A . The equivalence class of R determined by an element $x \in U$ is denoted by $R(x)$. Equivalence classes of R are called *granules* generated by R .

Attributes reduction stands for removal of attributes that do not effect the primary system. Rough sets attribute analysis is usually applied in the process of establishing the relative importance of an attribute and consecutively remove it if it contains redundant information.

Data analysis with various applications is well presented in [1], [2], [7], [8], [13]. Multi-criteria methods for project evaluation are applied in [5], [11].

3 Summarized Assessments

Services are evaluated by experts where their summarized assessments are denoted by v - very high level of success, s - high level of success, m - moderate level of success, l - low level of success, and u - unknown level of success. A concept lattice relating services and criteria evaluations is presented in Fig. 1. A graphical representation of rough sets approximations can also be seen in Fig. 2.

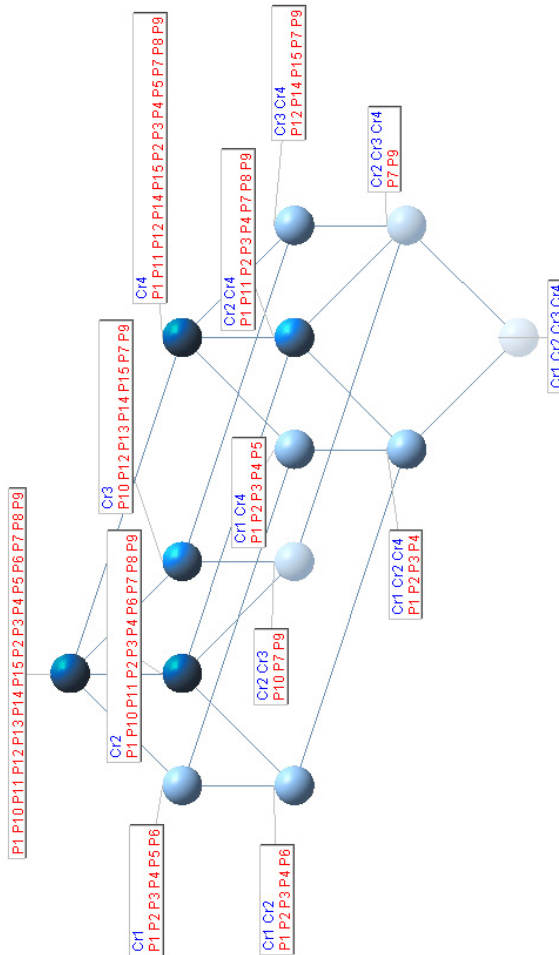


Fig. 1. A concept lattice

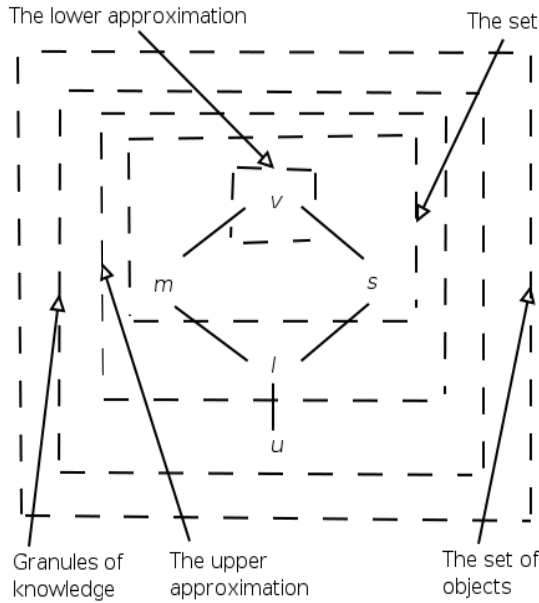


Fig. 2. Approximations

Below we present experts evaluations of services with respect to some criteria and outcomes.

The abbreviation P stands for a service and Cr for a criterion, where

Cr 1 - a nonempty cell indicates that two experts' evaluations found place within the lower approximation

Cr 2 - a nonempty cell indicates that two experts' evaluations found place within that part of the set that does not involve the lower approximation

Cr 3 - a nonempty cell indicates that two experts' evaluations found place within that part of the upper approximation that does not involve the lower approximation

Cr 4 - a nonempty cell indicates that two experts' evaluations found place outside of the upper approximation

- * - the first and the second evaluations are compared
- - the second and the third evaluations are compared
- x - the third and the forth evaluations are compared

Table 1. An illustrative decision table

	Cr 1	Cr 2	Cr 3	Cr 4
P 1	*	●		x
P 2	*	●		x
P 3	*	●		x
P 4	*	●		x
P 5	*	●		x
P 6	*	●		x
P 7		x	●	*
P 8		x	●	*
P 9		x	●	*
P 10		x	●	*
P 11		x	●	*
P 12	x		●	*
P 13	x		●	*
P 14	x		●	*
P 15	x		●	*

The indiscernable sets are $P_1 = \{ P 1, P 2, P 3, P 4, P 5, P 6, P 7 \}$, $P_2 = \{ P 8, P 9, P 10, P 11 \}$, $P_3 = \{ P 12, P 13, P 14, P 15 \}$. The set $\{ P 1, P 2, P 4, P 5, P 12, P 13, P 14, P 15 \}$ is a rough set because it can not be presented as an union of P_1 and P_3 . The upper and lower approximations of are $R^* = \{ P 1, P 2, P 3, P 4, P 5, P 6, P 7, P 8, P 9, P 10, P 11, P 12, P 13, P 14, P 15 \}$ and $R_* = \{ P 12, P 13, P 14, P 15 \}$.

4 Conclusion

Services’ assessment is on many occasions forced to extract information from imperfect, imprecise, and incomplete data. Therefore, precise reasoning rules are difficult and some times impossible to use. Applying rough sets approximations facilitates a balance between accuracy and precision.

References

1. Carpineto, C., Romano, G.: Concept Data Analysis: Theory and Applications. John Wiley and Sons, Ltd. (2004)
2. Garcia, M., Lloret, J., Sendra, S., Rodrigues, J.J.P.C.: Taking Cooperative Decisions in Group-Based Wireless Sensor Networks. In: Luo, Y. (ed.) CDVE 2011. LNCS, vol. 6874, pp. 61–65. Springer, Heidelberg (2011)

3. Gratzner, G.: *General Lattice Theory*. Academic Press, New York (1978)
4. Gradel, E., Otto, M., Rosen, E.: Undecidability results on two-variable logics. *Archive of Mathematical Logic* 38, 313–354 (1999)
5. Huylenbroeck, G., Martines, L.: The Average Value Ranking multi-criteria method for project evaluation in regional planning. *European Review of Agricultural Economics* 19(2), 237–252
6. Davey, B.A., Priestley, H.A.: *Introduction to lattices and order*. Cambridge University Press, Cambridge (2005)
7. Jiang, D., Tang, C., Zhang, A.: Cluster Analysis for Gene Expression Data: A Survey. *IEEE Trans. on Knowledge and Data Engineering* 16(1), 1370–1386 (2004)
8. Heyer, L.J., Kruglyak, S., Yooseph, S.: Exploring Expression Data: Identification and Analysis of Coexpressed Genes. *Genome Research* (1999)
9. Marek, V.W., Truszczynski, M.: Contributions to the theory of rough sets. *Fundamenta Informaticae* 39(4), 389–409 (1999)
10. Mayo, M., Mitrovic, A.: Optimising ITS behaviour with Bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education* 12, 124–153 (2001)
11. Monch, L., Lendermann, P., McGinnis, L.F., Schirrmann, A.: A survey of challenges in modelling and decision-making for discrete event logistics systems. *Computers in Industry* 62, 557–567 (2011)
12. Parsa, S., Parand, F.-A.: Cooperative decision making in a knowledge grid environment. *Future Generation Computer Systems* 23, 932–938 (2007)
13. Pfaltz, J.L.: Establishing Logical Rules from Empirical Data Intern. *Journal on Artificial Intelligence Tools* 17(5), 985–1001 (2008)
14. Pawlak, Z.: Rough Sets. *International Journal of Computer and Information Sciences* 11, 341–356 (1982)
15. Wille, R.: Concept Lattices and Conceptual Knowledge Systems. *Computers Math. Applications* 23(6-9), 493–515 (1992)
16. Yao, Y.Y.: Interval-set algebra for qualitative knowledge representation. In: *Proceedings of the Fifth International Conference on Computing and Information*, pp. 370–374 (1993)

Keyword Searchable Encryption with Access Control from a Certain Identity-Based Encryption

Koji Tomida¹, Masami Mohri², and Yoshiaki Shiraishi¹

¹ Nagoya Institute of Technology, Aichi 466-8555 Japan
tomida.koji@niztlab.com,
zenmei@nitech.ac.jp

² Gifu University, Gifu 501-1193 Japan
mmohri@gifu-u.ac.jp

Abstract. Data sharing on the cloud server is used because of the low management cost and its convenience. It is desirable for data to be stored on the cloud server in encrypted form for its confidentiality. To address the problem of searching on encrypted data, many searchable encryption schemes have been proposed. The searchable encryption enables the server to perform the keyword search on encrypted data without learning anything about the keyword and the original data. Some schemes have a function of access control over the encrypted data. But in these schemes the number of users providing the encrypted data to the server or performing the keyword search on encrypted data is limited. We propose a searchable encryption scheme with access control which does not limit the number of users providing and searching on the encrypted data.

Keywords: searchable encryption, access control, identity based encryption, bilinear map.

1 Introduction

Many searchable encryption schemes have been proposed since the first public key searchable encryption scheme [1] is proposed in 2004. In cloud service such that an entity providing data to the cloud server is different from an entity preserving data, data is desirable to be encrypted for its confidentiality. There are searchable encryption schemes with access control, which limits users who can search the ciphertext. A scheme in [2] is constructed based on broadcast encryption (BE) [2] and users providing the encrypted data are many and unspecified. The number of users who perform the search on the encrypted data are limited depending on its input parameter. A scheme in [3] is constructed based on identity based broadcast encryption (IDBE) [4] and the number of users who perform the search is not limited. This scheme assumes the situation where a data owner has many data to be stored in the server, and only the data owner can provide the encrypted data to the cloud server. We propose a searchable encryption scheme that many and unspecified users can provide the encrypted data and the number of users who perform the search is not limited. The proposed

scheme is constructed based on identity based encryption (IBE) [5]. We define a security game of the proposed scheme and prove its security.

2 Preliminary

2.1 Bilinear Map

Let \mathbb{G}_1 and \mathbb{G}_2 be two cyclic groups of order p for some large prime p . A bilinear map $e: \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ between these two groups satisfies the following properties:

1. **Bilinear:** $e(aP, bQ) = e(P, Q)^{ab}$ for all $P, Q \in \mathbb{G}_1$ and all $a, b \in \mathbb{Z}_p^*$.
2. **Non-degenerate:** If P is a generator of \mathbb{G}_1 then $e(P, P)$ is a generator of \mathbb{G}_2 .
3. **Computable:** There is an efficient algorithm to compute $e(P, Q)$ for any $P, Q \in \mathbb{G}_1$.

2.2 Bilinear Diffie-Hellman Problem (BDH)

The BDH problem is as follows: Fix a generator P of \mathbb{G}_1 . Given $P, aP, bP, cP \in \mathbb{G}_1$ as input, compute $e(P, P)^{abc} \in \mathbb{G}_2$. We say BDH is intractable if all polynomial time algorithms have a negligible advantage in solving BDH. In this paper BDH is assumed to be intractable.

3 Keyword Searchable Encryption with Access Control from a Certain Identity-Based Encryption

3.1 Model

Provider

The provider chooses a keyword W and users who are allowed to search. The provider encrypts W and gives it to the server.

Searcher

The searcher generates a trapdoor for a specific keyword corresponding to its own private key. The searcher sends it to the server to perform the keyword search and receives a response whether a ciphertext contains the specific keyword or not.

Server

The server receives a ciphertext from the provider and stores it. The server receives a trapdoor from the provider and searches ciphertext with trapdoor. The result is sent to the searcher whether ciphertext include the certain keyword or not without learning about the keyword.

PKG (Private Key Generator)

The PKG extracts a private key from a searcher's identity string ID, e-mail address for example, and give it to the searcher of identity ID.

3.2 Algorithm

The proposed scheme consists of five algorithms.

KeyGen(k)

Taking a security parameter k as input, KeyGen algorithm outputs system parameters $params$ and master secret key msk .

Extract($params, msk, ID$)

Taking system parameters $params$, master secret key msk and an identity string $ID \in \{0, 1\}^*$ of a searcher as input, Extract algorithm outputs a private key d_{ID} corresponding to ID .

Encrypt($params, ID, W$)

Taking system parameters $params$, any identity string ID , and a keyword W as input, Encrypt algorithm outputs a ciphertext C which is searchable encryption of W for the searcher of identity ID .

Trapdoor($params, d_{ID}, W$)

Taking system parameters $params$, the private key d_{ID} and a keyword W as input, Trapdoor algorithm outputs the trapdoor $T_{W, ID}$.

Test($params, C, T_{W, ID}$)

Taking the system parameters $params$, the ciphertext $C := \text{Encrypt}(params, ID', W')$ and the trapdoor $T_{W, ID}$ as input, Test algorithm outputs 1 if $W = W'$ and $ID = ID'$, otherwise outputs 0.

The flow of the proposed scheme is shown in Fig. 1.

3.3 Construction

We use a bilinear map $e: \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$, and three hash functions $H_1: \{0, 1\}^* \rightarrow \mathbb{G}_1^*$, $H_2: \mathbb{G}_2 \rightarrow \{0, 1\}^n$, $H_3: \{0, 1\}^* \rightarrow \mathbb{Z}_p^*$, where n is a natural number and p is a prime number. This construction is based on the identity based encryption scheme [5].

KeyGen(k)

The input security parameter k determines a prime, p , of two groups $\mathbb{G}_1, \mathbb{G}_2$ of order p , and a bilinear map $e: \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$. The KeyGen algorithm picks a random $s \in \mathbb{Z}_p^*$ and a generator $P \in \mathbb{G}_1$ and computes $P_{pub} := sP$. It chooses three cryptographic hash functions H_1, H_2, H_3 . It outputs the system parameters $params := \langle p, n, P, P_{pub}, H_1, H_2, H_3, e \rangle$ and the master key $msk := s$.

Extract($params, msk, ID$)

Given an identity $ID \in \{0, 1\}^*$, the Extract algorithm computes $Q_{ID} := H_1(ID) \in \mathbb{G}_1^*$. It outputs the private key $d_{ID} := sQ_{ID}$ where s is the master secret key.

Encrypt($params, ID, W$)

To encrypt a keyword $W \in \{0, 1\}^*$ the Encrypt algorithm first computes $Q_{ID} := H_1(ID) \in \mathbb{G}_1^*$, $H_3(W) \in \mathbb{Z}_p^*$ and chooses a random $r \in \mathbb{Z}_p^*$. It outputs the ciphertext $C := [rP, H_2(g_{ID}^r)]$, $g_{ID} := e(H_3(W)Q_{ID}, P_{pub}) \in \mathbb{G}_2$.

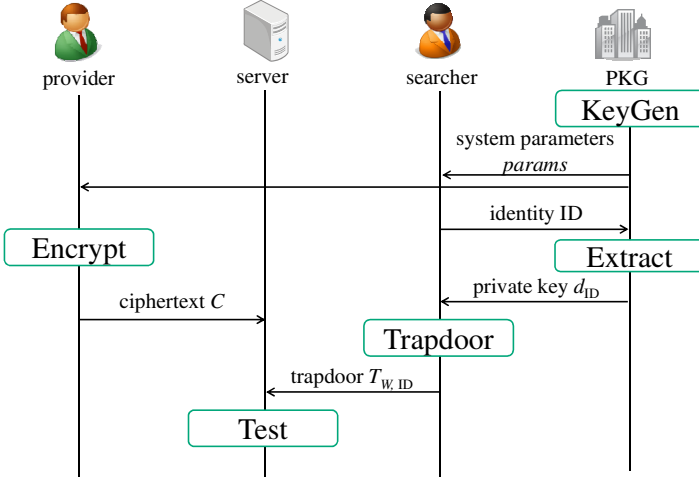


Fig. 1. Flow of the proposed scheme

Trapdoor($params, d_{ID}, W$)

Given a private key d_{ID} and a keyword W , the Trapdoor algorithm computes $T_{W, ID} := H_3(W)d_{ID}$ and outputs it.

Test($params, C, T_{W, ID}$)

Let $C := [C_1, C_2]$ be a ciphertext. The test algorithm tests if $H_2(e(T_{W, ID}, C_1)) = C_2$, if so outputs 1, otherwise 0.

3.4 Security Definition

We define security for the proposed scheme as a security game between a challenger and an attacker. The attacker chooses two different keywords W_0, W_1 and any identity ID^* of its choice and gives them to the challenger. The challenger picks a random $b \in \{0, 1\}$ and gives the attacker $C^* := \text{Encrypt}(params, ID^*, W_b)$ as challenge ciphertext. The attacker outputs $b' \in \{0, 1\}$. The attacker wins the game if $b = b'$. Throughout the game the attacker can obtain any trapdoor adaptively except for T_{W_0, ID^*} or T_{W_1, ID^*} . If the attacker’s advantage $\text{Adv}_A(k) = |\Pr[b = b'] - \frac{1}{2}|$ is negligible function then we say that the proposed scheme is keyword indistinguishable under adaptive keyword attacks.

4 Security Proof

We need to ensure that an $\text{Encrypt}(params, ID, W)$ must not reveal any information about the keyword W unless trapdoor $T_{W, ID}$ is available. The attacker should not be able to distinguish between the encryption of two different keywords W_0, W_1 of its choice for any identity ID of its choice even if the attacker can obtain any trapdoor

except for $T_{W_0, ID}$ or $T_{W_1, ID}$. We suppose the attacker can obtain the system parameters $params$ and the ciphertext C .

If the attacker can compute $H_2(e(H_3(W)Q_{ID}, P_{pub})^r)$ from $params$ and C without trapdoor, then the attacker can get to know the keyword included in the ciphertext. As a result, the attacker is able to distinguish the ciphertext.

This problem is equal to the problem whether or not the attacker can compute $e(P, P)^{\alpha\beta\gamma}$ from given $Q_{ID} := \alpha P, P_{pub} := \beta P$ and $C_1 := \gamma P$. This is BDH problem itself. Therefore under the BDH assumption the attacker is not able to distinguish between two encryptions.

The security of the proposed scheme can be strictly proved analogously to the security proof of the scheme in [1].

5 Comparison

We compare the proposed scheme with the schemes in [2] and [3] in view of their base scheme, the limit of the number of providers and searchers, and the ability to add a new searcher to the system as required. The result is shown in Table 1.

In these schemes searchers have their unique private keys. Depending on their private key, which is used to generate a trapdoor, their trapdoors are distinguished whether they are allowed to search for the ciphertext or not. In the scheme in [2], Setup algorithm (the first algorithm) takes parameter N , which denotes the number of users, and generates N private keys depending on input parameter N . Therefore the number of the searchers is limited to N and we are not able to add a new searcher as necessary. In the proposed scheme and the scheme in [3], private keys of searchers are generated from their own identity ID (any string is available as ID) and there is no limit of the number of searchers. In our scheme PKG generates private keys of searchers, on the other hand the provider generates private keys in the scheme in [3].

In the proposed scheme and scheme in [2], ciphertext of the keyword is generated from public system parameter and many and unspecified users can generate a ciphertext and provide it to the server. In the scheme in [3] ciphertext of the keyword is generated from not only system parameter but also master secret key. The master secret key should not to be shared by many users because of the system security

Table 1. Comparison of schemes

	scheme in [2]	scheme in [3]	proposed scheme
base scheme	broadcast encryption [2]	identity based broadcast encryption [4]	identity based encryption [5]
number of searchers	limited to the parameter	no limit	no limit
number of providers	many and unspecified	1	many and unspecified
addition of searcher	impossible	possible	possible

because any trapdoors and private keys can be generated by using the master secret key. So only the data owner who holds master secret key can generate the ciphertext.

6 Conclusion

This paper proposed the searchable encryption scheme with access control. This scheme can be applied to data sharing on the cloud server without leakage information in multi-user setting, where many and unspecified users can provide encrypted data to the cloud server and perform the keyword search on encrypted data if they are allowed to search for each ciphertext. The number of searchers is not limited and we can add a new searcher as required by generating a new private key. The proposed scheme is keyword indistinguishable under the assumption that BDH problem is intractable.

The first public key encryption scheme in [1] and the proposed scheme are also based on identity based encryption [5]. The scheme in [1] performs the search on encrypted data by testing keyword matching included in ciphertext and trapdoor without learning anything else about the keyword. Compared with the scheme in [1], the proposed scheme adds the identity information to both ciphertext and trapdoor and the search is performed by testing both keyword and identity matching included in ciphertext and trapdoor.

References

1. Boneh, D., Di Crescenzo, G., Ostrovsky, R., Persiano, G.: Public Key Encryption with keyword Search. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 506–522. Springer, Heidelberg (2004)
2. Attrapadung, N., Furukawa, J., Imai, H.: Forward-Secure and Searchable Broadcast Encryption with Short Ciphertexts and Private Keys. In: Lai, X., Chen, K. (eds.) ASIACRYPT 2006. LNCS, vol. 4284, pp. 161–177. Springer, Heidelberg (2006)
3. Katayama, T., Takagi, T.: Keyword Searchable Encryption allowing Access Control. In: SCIS 2008, 4E2-2 (2008)
4. Sakai, R., Furukawa, J.: Identity-Based Broadcast Encryption. IACR Cryptology ePrint Archive 2007: 217 (2007)
5. Boneh, D., Franklin, M.: Identity-based Encryption from the Weil Pairing. SIAM J. on Computing 32(3), 586–615 (2003)

Attribute-Based Encryption with Attribute Revocation and Grant Function Using Proxy Re-encryption and Attribute Key for Updating

Takeru Naruse¹, Masami Mohri², and Yoshiaki Shiraishi¹

¹ Nagoya Institute of Technology, Aichi 466-8555, Japan
naruse.takeru@niztlab.com
zenmei@nitech.ac.jp

² Gifu University, Gifu 501-1193, Japan
mmohri@gifu-u.ac.jp

Abstract. Ciphertext-Policy Attribute-Based Encryption (CP-ABE) is suitable for data access control on a cloud storage system. In CP-ABE, the data owner encrypts data under the access structure over attributes and a set of attributes assigned to users is embedded in user's secret key. A user is able to decrypt if his attributes satisfy the ciphertext's access structure. In CP-ABE, processes of user's attribute revocation and grant are concentrated on the authority and the data owner. In this paper, we propose a ciphertext-policy attribute-based encryption scheme delegating attribute revocation processes to Cloud Server by proxy re-encryption. The proposed scheme does not require generations of new secret key when granting attributes to a user and supports any Linear Secret Sharing Schemes (LSSS) access structure.

Keywords: cryptographic cloud storage, CP-ABE, attribute revocation and grant, proxy re-encryption.

1 Introduction

Sharing of data on a cloud storage has a risk of information leakage caused by service provider's abuse. In order to protect data, the data owner encrypts data shared on the cloud storage so that only authorized users can decrypt.

Ciphertext-Policy Attribute-Based Encryption (CP-ABE) [1][2] is suitable for data access control in the cloud storage system. The authority manages the attributes in the system. The data owner chooses an access structure and encrypts message under the access structure. The set of attributes assigned to users is embedded in his secret key. A user is able to decrypt a ciphertext if his attributes satisfy the ciphertext's access structure.

There are user's attribute revocation and grant in CP-ABE. In simple processes of user's attribute revocation, when his attributes are revoked, the data owner re-encrypts the shared data so that revoked user cannot decrypt. Then, the authority redistributes new secret keys so that other users can decrypt. In simple processes of user's attribute

grant, the authority generates a new secret key. These simple processes are concentrated on the data owner and the authority.

Some attribute revocable CP-ABE schemes have been proposed [3-5]. Yu et al. [3] proposed a scheme combining CP-ABE with proxy re-encryption. The authority can delegate re-encryption and secret key update to proxy servers. However, this scheme has a limitation in access policy because it can only express “AND” policy. Hur et al. [4] proposed a scheme using key encryption keys (KEKs). A service provider distributes KEKs to each user. The service provider re-encrypts a ciphertext by an attribute group key. Then, he encrypts attribute group key by using KEKs so that authorized user can decrypt. As the number of users system has increases, the number of KEKs also increases and management becomes complicated. Liangu et al. [5] proposed a scheme using user information (UI). UI is generated by Revocation Tree and Revocation List. An authorized user can decrypt ciphertexts by using secret key and UI. In this scheme, users whose attributes are revoked lose the access rights to all shared data by attribute revocation processes.

Moreover, in these schemes [3-5], the authority needs to generate a new key when granting attribute to users.

In this paper, we propose a CP-ABE scheme delegating attribute revocation processes to Cloud Server by proxy re-encryption and meets the following requirements.

- 1) Support any Linear Secret Sharing Schemes (LSSS) access structure.
- 2) Revoke the only specified attribute (attribute level user revocation).
- 3) Does not require the generation of new secret key when granting attribute to user.

2 Preliminaries

2.1 Bilinear Maps

Let G_1, G_2 be two cyclic groups of prime order p . Let P be a generator of G_1 . A bilinear map is a map $e : G_1 \times G_1 \rightarrow G_2$ with the following properties:

1. Bilinearity: for all $P, Q \in G_1$ and $a, b \in \mathbb{Z}_p$, we have $e(aP, bQ) \rightarrow e(P, Q)^{ab}$.
2. Non-degeneracy: $e(P, P) \neq 1$.
3. Computability: There is an efficient algorithm to compute $e(P, Q)$ for all $P, Q \in G_1$.

2.2 Linear Secret Sharing Scheme (LSSS)

Definition 1 (Linear Secret Sharing Schemes (LSSS) [2][6]). A secret-sharing scheme Π over a set of parties \mathcal{P} is called linear (over \mathbb{Z}_p) if

1. The shares for each party form a vector over \mathbb{Z}_p .
2. There exists a matrix an M with l rows and n columns called the share-generating matrix for Π . For all $i = 1, \dots, l$, the i 'th row of M we let the function ρ defined the party labeling row i as $\rho(i)$. When we consider the column vector

$v = (s, r_2, \dots, r_n)$, where $s \in Z_p$ is the secret to be shared, and $r_2, \dots, r_n \in Z_p$ are randomly chosen, then Mv is the vector of l shares of the secret s according to Π . The share $(Mv)_i$ belongs to party $\rho(i)$.

Suppose that Π is an LSSS for the access structure \mathbb{A} . Let $S \in \mathbb{A}$ be any authorized set, and let $I \subset \{1, 2, \dots, l\}$. Then, there exist constants $\{\omega_i \in Z_p\}_{i \in I}$ such that, if $\{\lambda_i\}$ are valid shares of any secret s according to Π , then $\sum_{i \in I} \omega_i \lambda_i = s$. Furthermore, these constants $\{\omega_i\}$ can be found in time polynomial in the size of the share-generating matrix M [6].

3 Our Scheme

3.1 Model

There are four entities in the proposed scheme as follows.

User: The user downloads the shared data from Cloud Server.

Data owner: The data owner encrypts the shared data then uploads to Cloud Server.

Authority: The authority manages attributes in the system and publishes the parameters used for encryption. It generates a secret key that user's attributes are embedded and PRE keys used for re-encryption and updating secret key. The authority is trusted party.

Cloud Server: Cloud Server stores shared data. It re-encrypts encrypted shared data and update secret key by using PRE keys received from the authority. Similar to previous schemes [3][4], we assume Cloud Server to be curious-but-honest. That is, it will honestly execute the tasks assigned by legitimate parties in the system. However, it would like to learn information of encrypted shared data as much as possible.

3.2 Overview

The proposed scheme is based on Waters's scheme of CP-ABE [2]. Water's scheme supports any LSSS access structure. We apply the idea of attribute revocation shown in [3] to the proposed scheme. In the proposed scheme, the attribute key is included in the ciphertext and secret key to delegate attribute revocation processes to Cloud Server. The attribute key is master key components corresponding to each attribute in the system. When user's attributes are revoked, the authority re-defines the attribute keys, and generates PRE keys for updating the attribute keys. Cloud Server re-encrypts ciphertext and updates secret key by updating attribute key by using PRE key. Each attribute is associated with version number for updating attribute key.

Cloud Server keeps user list UL , re-encryption key list RKL and the key for granting an attribute to secret key J . UL records user's ID , user's attribute information, secret key components, t_{ID} . t_{ID} is a random number that randomize each secret key to prevent users' collusion attack. t_{ID} should "bind" components of one user's key together so that they cannot be combined with another user's key components[2]. RKL records update history of attribute (version number) and PRE keys.

When granting attributes to users, Cloud Server generates user's secret key components correspond to granting attribute from t_{ID} and J , and sends secret key component to the user. The user joins secret key component to own secret key. Thus, it is possible to grant attributes to users without generation of new secret key by the authority.

3.3 Algorithm

Auth.Setup(U). The setup algorithm takes as input the number of system attributes U . It first chooses a group G_1 of prime order p , a generator $P \in G_1$. It then chooses random group elements $Q_1, \dots, Q_U \in G_1$ that are associated with the U attributes in the system. In addition, it chooses two random $\alpha, a \in \mathbb{Z}_p$, and random $Att_1, \dots, Att_U \in \mathbb{Z}_p$ as the attribute key.

The public parameters are

$$PK := \langle P, e(P, P)^\alpha, aP, Q_1, \dots, Q_U, T_1 = Att_1P, \dots, T_U \rangle.$$

The master key is $MK := \langle \alpha, Att_1, \dots, Att_U \rangle$.

The keys for granting an attribute are $J := \langle \{x, J_x = 1/Att_x\}_{1 \leq x \leq U} \rangle$.

DO.Enc($PK, (M, \rho), \mathcal{M}$). The Encryption algorithm takes as input the public parameters PK , an LSSS access structure (M, ρ) , and a message \mathcal{M} . The function ρ associates rows of M to attributes. Let M be an $l \times n$ matrix. It first chooses a random vector $\vec{v} = (s, y_2, \dots, y_n) \in \mathbb{Z}_p$. For $i = 1$ to l , it computes $\lambda_i := \vec{v} \cdot M_i$. It then chooses random $r_1, \dots, r_l \in \mathbb{Z}_p$ and outputs the ciphertext

$$CT := \langle C, C', (C_1, D_1), \dots, (C_l, D_l) \rangle =$$

$$\langle Ke(P, P)^{\alpha s}, sP, (\lambda_1(aP) - r_1Q_{\rho(1)}, r_1T_{\rho(1)}), \dots, (\lambda_l(aP) - r_lQ_{\rho(l)}, r_lT_{\rho(l)}) \rangle$$

with (M, ρ) .

Auth.Ext(MK, S). The key extraction algorithm takes as input the master key MK , and a set of attributes S . It first chooses a random $t_{ID} \in \mathbb{Z}_p$. It then outputs t_{ID} and the secret key

$$SK := \langle K, L, \forall x \in S K_x \rangle =$$

$$\langle \alpha P + t_{ID}(aP), t_{ID}P, \forall x \in S (t_{ID}/Att_x)Q_x \rangle.$$

U.Dec(SK, CT). The decryption algorithm takes as input a secret key SK for a set S and a ciphertext CT for access structure (M, ρ) . Suppose that S satisfies the access structure and let I be defined as $I = \{i : \rho(i) \in S\}$. Then, let $\{\omega_i \in \mathbb{Z}_p\}_{i \in I}$ be as set of constants such that if $\{\lambda_i\}$ are valid shares of the secret s according to M , then $\sum_{i \in I} \omega_i \lambda_i = s$.

The decryption algorithm first computes

$$\frac{e(C', K)}{\prod_{i \in I} (e(C_i, L) e(D_i, K_{\rho(i)}))^{\omega_i}} = \frac{e(P, P)^{\alpha s} e(P, P)^{\alpha s t_{ID}}}{\prod_{i \in I} (e(P, P)^{t \alpha \lambda_i \omega_i})} = e(P, P)^{\alpha s}.$$

It can then decrypt the message $\mathcal{M} = C/e(P, P)^{\alpha s}$.

Auth.ReKeyGen(MK, γ). The re-encryption key generation algorithm takes as input the master key MK and a set of attributes γ for update. For each $x \in \gamma$, it chooses random $Att'_x \in \mathbb{Z}_p$ as the new attribute key, and computes $T'_x := Att'_x P$, $rk_{x \rightarrow x'} := \frac{Att'_x}{Att_x}$. It then replaces each Att_x of the master key component with Att'_x , and each T_x of public parameter with T'_x . It outputs the redefined the master key MK' , the redefined public parameters PK' , and the PRE keys $rk := \{x, rk_x\}_{x \in \gamma}$.

C.ReEnc($y(= \rho(i)), D_i, RKL_y$). The re-encryption algorithm takes as input an attribute $y(= \rho(i))$ for update, the ciphertext component D_i and a PRE key list RKL_y . It first checks version of attribute y . If y has the latest version, it outputs \perp and exit. Let $Att_{y(n)}$ be defined as an attribute key of the latest version of attribute y . It computes $rk_{y \leftrightarrow y(n)} := rk_{y \leftrightarrow y'} \cdot rk_{y' \leftrightarrow y''} \cdots rk_{y(n-1) \leftrightarrow y(n)} = Att_{y(n)}/Att_y$. Then, it outputs the re-encrypted ciphertext component $D'_i := rk_{y \leftrightarrow y(n)} \cdot D_i = (Att_{y(n)}/Att_y) \cdot r_i Att_y P = r_i Att_{y(n)} P$.

C.ReKey(w, K_w, ID, RKL_w). The key regeneration algorithm takes as input an attribute w for update, the secret key component K_w and the PRE key list RKL_w . It first checks version of attribute w . If w has the latest version, it outputs \perp and exit. Let $Att_{w(n)}$ be defined as the attribute key for the latest version of attribute w . It computes $rk_{w \leftrightarrow w(n)} := rk_{w \leftrightarrow w'} \cdot rk_{w' \leftrightarrow w''} \cdots rk_{w(n-1) \leftrightarrow w(n)} = Att_{w(n)}/Att_w$. It then outputs the updated secret key component $K'_w := rk_{w \leftrightarrow w(n)}^{-1} \cdot K_w = (Att_w/Att_{w(n)}) \cdot (t_{ID}/Att_w) Q_w = (t_{ID}/Att_{w(n)}) Q_w$.

C.GrantAtt(v, J_v, t_{ID}, RKL_v). The attribute grant algorithm takes as input an attribute v , the key of granting an attribute J_v , t_{ID} and the PRE key list RKL_v . It first checks version of attribute v . Let $Att_{v(n)}$ be defined as the attribute key for the latest version of attribute v . It first computes $rk_{v \leftrightarrow v(n)} := rk_{v \leftrightarrow v'} \cdot rk_{v' \leftrightarrow v''} \cdots rk_{v(n-1) \leftrightarrow v(n)} = Att_{v(n)}/Att_v$. It then outputs secret key component for $K_v := t_{ID} \cdot rk_{v \leftrightarrow v(n)}^{-1} \cdot J_v = t_{ID} \cdot (Att_v/Att_{v(n)}) \cdot (1/Att_v) Q_v = (t_{ID}/Att_{v(n)}) Q_v$ and redefines the key of granting an attribute $J'_v := rk_{v \leftrightarrow v(n)}^{-1} \cdot J_v = (1/Att_{v(n)}) Q_v$.

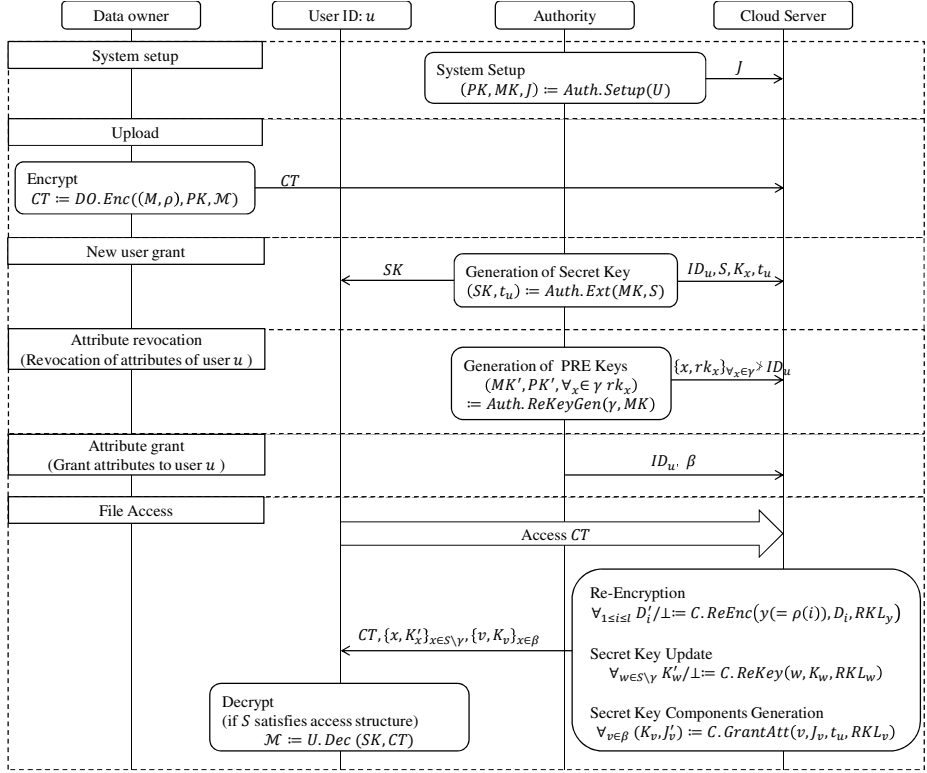


Fig. 1. Flow of the proposed scheme

We show the flow of our scheme in Fig 1. In Fig 1, γ denotes a set of user u 's attributes which are revoked and β denotes a set of attributes that granting to user u .

4 Security Proof

We prove that unauthorized users and Cloud Server cannot decrypt ciphertext CT that was encrypted by using the proposed scheme. Since we assume Cloud Server is honest, we do not consider active attacks from CloudServer by colluding with revoked users as in [3][4].

An unauthorized user cannot decrypt CT because his secret key does not contain components that corresponds to attributes necessary for decryption. In addition, each secret key is randomized with a freshly chosen exponent t_{ID} to prevent collusion attack. t_{ID} should "bind" the components of one user's key together so that they cannot be combined with another user's key components [2]. Therefore, unauthorized users cannot decrypt the ciphertext CT .

Cloud Server keeps secret key components K_x , t_{ID} and the keys for granting an attribute to secret key J . It can generate any K_x that corresponds attribute in the system, but CT cannot be decrypted only with K_x . Therefore, Cloud Server cannot decrypt the ciphertext CT .

Table 1. Comparison of schemes

	scheme[3]	scheme[4]	Scheme[5]	proposed scheme
Supporting Access Policy Type	'AND'	'AND','OR'	Any LSSS	Any LSSS
Attribute level user revocation	Possible	Possible	Impossible	Possible
Grant attributes to users	The authority generates a new secret key	The authority generates a new secret key	The authority generates a new secret key	Cloud Server adds attributes to user's secret key

5 Comparison and Conclusion

This paper proposed a ciphertext-policy attribute-based encryption scheme delegating attribute revocation processes to Cloud Server by proxy re-encryption. Cloud Server re-encrypts a ciphertext and updates a secret key by updating attribute key with PRE key for updating the attribute keys. We compared the proposed scheme with schemes of [3-5] and show the comparison result at Table 1.

The proposed scheme meets three requirements as follows;

First, the proposed scheme supports any LSSS access structure. Second, the authority can only revoke specified attribute by updating attribute key included in ciphertext corresponding to his attributes which are revoked. Finally, when granting attributes to a user, generation of a new secret key becomes unnecessary because Cloud Server generates secret key components corresponding to granting attributes.

References

1. Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-Policy Attribute-Based Encryption. In: IEEE Symposium on Security and Privacy, pp. 181–194 (2007)
2. Waters, B.: Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization. In: Catalano, D., Fazio, N., Gennaro, R., Nicolosi, A. (eds.) PKC 2011. LNCS, vol. 6571, pp. 53–70. Springer, Heidelberg (2011)
3. Yu, S., Wang, C., Ren, K., Lou, W.: Attribute Based Data Sharing with Attribute Revocation. In: 5th ACM Symposium on Information, Computer and Communications Security, pp. 261–270 (2010)
4. Hur, J., Nor, D.K.: Attribute-Based Access Control with Efficient Revocation in Data Outsourcing Systems. IEEE Transactions on Parallel and Distributed Systems 22, 1214–1221 (2011)
5. Liang, X., Lu, R., Lin, X., Shen, X.: Ciphertext Policy Attribute Based Encryption with Efficient Revocation. Technical report, Univ. of Waterloo (2011)
6. Beimel, A.: Secure Schemes for Secret Sharing and Key Distribution. PhD thesis, Israel Institute of Technology (1996)

Research on Stereo Image Authentication Watermarking with Self-recovery

Ting Luo, Gangyi Jiang^{*}, Mei Yu, Yigang Wang, Feng Shao, and Zongju Peng

Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
jianggangyi@126.com

Abstract. With rapid development of three dimensional video systems, the integrity of stereo image has become increasingly important. However, stereo image authentication watermarking methods are rarely reported, and only monocular watermarking methods are extensively studied. In order to ensure authentication and integrity of stereo image, different stereo image watermarking methods are presented in different ways. Firstly, existing monocular watermarking methods are extended directly to stereo image without correlations of stereo image. Secondly, monocular image watermarking methods are extended further as that disparity is used to recover tamper. Finally stereo vision based stereo image watermarking methods are proposed for embedding more recovery bits to improve tamper recovery. Experimental results show that stereo image watermarking method based on correlations of stereo image can improve watermarking performance.

Keywords: Three dimensional video system; Stereo image watermarking; Disparity; Stereo vision.

1 Introduction

With the development of three dimensional (3D) technologies, three dimensional Television (3DTV) are becoming main stream of the consuming entertainment than ever [1]. The flourish of computer has made distribution of 3D contents much easier and faster. Stereo image is a main representation of 3D image, and may be tampered by illegal users. Consequently the integrity of stereo image needs to be authenticated.

Authentication watermarking is a technology to embed particular information inside multimedia contents as a solution to keep integrity of stereo image in advance [2]. Many authentication watermarking methods focus on tamper location and recovery. Recovery bits generated for each image block are embedded into their unique mapping blocks [3]. If image blocks are tampered, corresponding recovery bits are extracted to recover tampers. However, a tampered block cannot be recovered if its unique mapping block is destroyed as well. In order to solve this problem, recovery bits of each block are embedded twice into two mapping blocks, respectively [4], to improve quality of recovery.

However, above authentication watermarking methods were designed for monocular image. Existing stereo image watermarking methods are mainly designed

^{*} Corresponding author.

for copyright protection. For example, Campisi extracted objects from depth map computed from left and right views, where watermark is embedded to resist JPEG compression [5]. Authentication stereo image watermarking with self-recovery is rarely reported. Stereo image is different from monocular image, which consists of left and right views. More importantly, two views of stereo image have high correlations, such as disparity and stereo vision. Correlations can be employed to design stereo image watermarking for improving performance.

In this paper, in order to authenticate the integrity of stereo image, firstly, two existing monocular image watermarking methods are directly extended to stereo image, where two views are considered as independent. Then, those two methods are further extended as that disparity between two views is used to recover tamper. Finally, stereo vision base stereo image watermarking method is presented. The comparison results of experiments are tested for five different stereo image watermarking methods, and it proves that stereo image watermarking methods using correlations between two views are superior to directly extended monocular image watermarking methods.

2 Stereo Image Watermarking Methods

In order to authenticate stereo image, five stereo image authentication watermarking methods are presented with self recovery. Let each view of stereo image is of $N_1 \times N_2$.

2.1 Four Extended Monocular Image Watermarking Methods

Firstly, two existing monocular image watermarking methods [3,4] are directly extended to stereo image, where two views are considered as two independent views. Thus first two stereo image watermarking methods just follow methods of [3] and [4] in the processes of watermark embedding, tamper detection. Only in the process of tamper recovery, the inpainting method [6] replaces 3×3 neighborhood method. Two stereo image watermarking methods are named as Lin's and Lee's.

Furthermore, monocular image watermarking methods are extended further using correlations of stereo image, especially in tamper recovery.

Since two views are captured for the same scene from two different cameras, contents are correlated with each other, that is, pixels in left view are matched with pixels in right view. Suppose two views are captured by two parallel cameras, that is, only horizontal disparity is taken into account. Matched pixels of stereo image are defined as.

$$I^L(i, j) \approx I^R(i - d(i, j), j) \quad (1)$$

where $I^L(x,y)$ and $I^R(x,y)$ are pixels in left and right views, respectively, $d(x,y)$ is the disparity for pixels, $1 \leq x \leq N_1$, and $1 \leq y \leq N_2$.

Process of tamper recovery in Lin's and Lee's methods is improved as that disparity is employed to recover tamper. However, disparity for pixels takes up 50% compared with stereo image, and it is not suitable to take extra bands for disparity transmission. Thus firstly disparity of pixels is computed [7], and then disparity is

computed for blocks of 4×4. If disparities of pixels are same in same block, which are assigned to disparity of blocks, otherwise, blocks in one view are not matched with any block in the other view. Disparity of blocks takes up 3.125%, and is suitable for real applications.

The tamper recovery process is modified that after tamper recovery from embedded recovery bits, some blocks are not recovered, and those tampered pixels are recovered by using disparities. If after recovery with disparity some pixels are still not recovered, the inpainting method is employed as well. Two further extended methods are named as Lin's+D and Lee's+D.

2.2 A Novel Stereo Image Watermarking Method

Besides disparity, more correlations of stereo image can be employed for improving performance of stereo image watermarking, such as stereo vision masking. Based on stereo vision masking model, watermark capacity can be increased much without perception of humane eyes. Thus, more recovery reference can be embedded for each block, and performance on tamper recovery can be better.

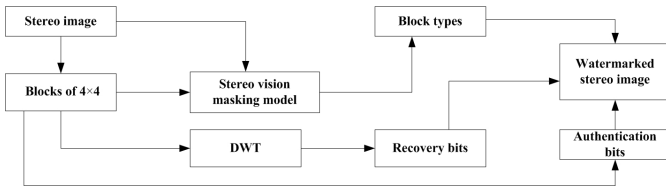


Fig. 1. Flowchart of watermark embedding

Based on stereo vision (SV) masking model, a stereo image watermarking is proposed and named as SV stereo image watermarking method. Watermark embedding process is illustrated in Fig. 1, and main steps are as follows.

Step 1. Stereo image is divided into blocks of 4×4. Just noticeable difference (JND) value of each pixel is computed based on stereo vision masking model, and average JND value for each block is computed as well. If average JND of blocks is equal or greater than 7, three LSBs of each pixel in the block are allocated for watermark embedding, otherwise, two LSBs are allocated. Recovery bits are generated by approximation coefficients of DWT instead of average intensity.

$$X_k(x, y) = Round(D_k(x, y) / Q) \tag{2}$$

where $D_k(x,y)$ is a approximation coefficient of DWT block (x,y) , $k \in \{1,2,3,4\}$, and $Q=17$. $X_k(x,y)$ are represented by 5 bits. Each block can be classified into four types according to Eq. (3).

$$\forall k \lfloor X_1(x, y) - X_k(x, y) \rfloor \leq \alpha, k \in \{2, 3, 4\} \tag{3}$$

where α is assigned to 1,3 or 7, and thus 11, 14 or 17 bits are used to represent those types of block. If Eq. (3) is not satisfied for all values of α , 20 bits are used for the last type.

Step 2. Authentication bits are computed by using parity check of pixels, which are combined with recovery bits in their predefined mapping blocks, the way similar as Lin's method [3]. Watermarked stereo image is achieved.

In the process of tamper detection, authentication bits are extracted as the reverse of watermark embedding, which are checked whether they are still same as authentication bits in the process of watermark embedding. If they are not same, blocks are identified as invalid. Two binary masking are for left and right views, respectively, and 1 is invalid and 0 is valid. Moreover, morphological erosion and dilation are operated on those two masking, tampered blocks are detected finally.

If any block is invalid, valid recovery bits are extracted from those mapping blocks as the reverse of recovery bits embedding. Some blocks are not recovered because corresponding mapping blocks are tampered as well, and are recovered by using the inpainting method finally.

3 Experimental Results and Discussions

In order to evaluate performance of proposed five stereo image watermarking methods, the first frame of "Laptop" and "Alt Moabit" with 640×480 pixels are taken as tested stereo images shown in Fig. 2.



Fig. 2. Tested stereo image

3.1 Single Region of Tamper

In this section, four experiments are tested on 'Laptop', where 'Laptop' is pasted by different objects at a single location with different ratios of tamper from around 3.27% to 34.11% of stereo image as shown in Figs. 3(a) to 3(d). Moreover, different texture regions are modified, such as smooth background is covered by 'pumpkin' as shown in Fig. 3(a), and texture background is covered by 'ball' as shown in Fig. 3(b).

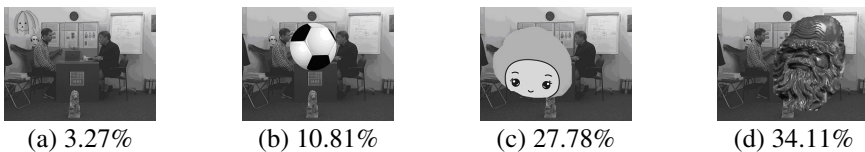


Fig. 3. Left view of tampered stereo image with different ratios

Five stereo image watermarking methods can detect tamper with tamper detection ratio of nearly 99%. SV method recovers tamper without visual perception as shown in Fig. 4.

Table 1 shows PSNR of tamper recovery relative to original stereo image with different methods. Lin's+D and Lee's+D perform better than their directly extended methods without disparity, respectively, and especially tamper ratios are higher. Only 3.27% of stereo image is tampered, methods with or without disparity recover tamper with same PSNR, because embedded recovery bits can recover tamper completely. Thus, disparity plays an important role in tamper recovery. PSNR of SV method is better than those of other four methods, and only is less than Lin's and Lin's+D when tamper ratio is around 3.27%. It denotes that proper use of stereo vision can be embedded more recovery bits to improve performance of stereo image watermarking method.



(b) Recovery of tamper for four different ratios of tamper

Fig. 4. Left view of tamper detection and recovery with SV method

Table 1. PSNR of tamper recovery for four different tamper ratios of 'Laptop' [dB]

	Lin's		Lin's+D		Lee's		Lee's+D		SV	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
3.27%	43.76	43.25	43.76	43.25	37.57	37.76	37.57	37.76	39.65	38.59
10.81%	35.82	35.83	36.01	36.12	34.40	34.62	35.33	35.78	37.75	36.94
27.78%	34.11	34.02	34.65	34.52	32.71	32.96	33.74	34.09	35.57	35.23
34.11%	34.69	33.94	35.63	35.36	29.78	29.56	30.10	29.34	36.16	35.19

3.2 Multiple Regions of Tamper

In order to show correlations of stereo image can improve performance of watermarking again, 'Alt Moabit' is tampered with multiple regions of tamper as shown in Fig. 5. Pixels being 255 are supposed to be cropped from images as shown in Figs. 5(b) and 5(c).

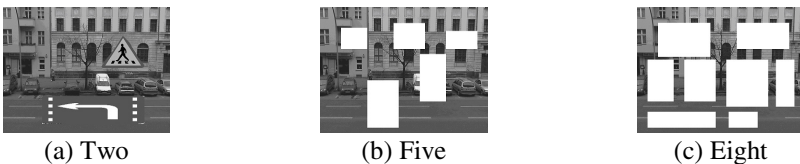


Fig. 5. Left view of tampered stereo image with different number of tampered regions

From Table 2, Lin's+D and Lee's+D are better than their originals. PSNRs of SV method for left view are better than those of other four methods. Although PSNR of

SV for right view is a little less than Lin's+D sometimes, the stereo visual quality is based on the view with better quality according to [10]. Thus, three experiments prove that methods with disparity are better than methods without disparity again. Moreover SV method is superior to other four methods as well. Correlations of stereo image improve stereo image watermarking methods much.

Table 2. PSNR of tamper recovery for tamper of 'Alt Moabit' [.dB]

	Lin's		Lin's+D		Lee's		Lee's+D		SV	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Two	36.83	37.13	37.21	37.82	26.62	26.34	27.01	26.89	38.07	36.21
Five	31.88	31.89	33.03	33.33	31.48	31.87	32.27	32.46	33.98	32.95
Eight	27.76	27.29	29.08	29.27	27.88	27.97	28.60	28.57	30.01	29.32

4 Conclusion

This paper has proposed five stereo image watermarking methods. Two methods are directly extended monocular watermarking methods, and other three methods use correlations of stereo image to improve performance in tamper recovery. Experimental results show stereo image watermarking methods with disparity or stereo vision performs better than other two methods.

Acknowledgments. This work was supported by Natural Science Foundation of China (61071120, 61171163, 61111140392, 61271270, 612712700), Natural Science Foundation of Ningbo (2012A610045), and Scientific Research Foundation of Ningbo University (XYL12001).

References

1. Son, J.Y., Son, W.H., Kim, W.H., et al.: Three-dimensional imaging for creating real-world-like environments. *Proceedings of the IEEE* 101, 190–205 (2013)
2. Haouzia, A., Noumeir, R.: Methods for image authentication: a survey. *Multimedia Tools and Application* 39, 1–46 (2008)
3. Lin, P., Hsieh, C., Huang, P.: A hierarchical digital watermarking method for image tamper and recovery. *Pattern Recognition* 38, 2519–2529 (2005)
4. Lee, T., Lin, S.: Dual watermark for image tamper detection and recovery. *Pattern Recognition* 41, 3497–3506 (2008)
5. Campisi, P.: Object-oriented stereo-image digital watermarking. *Journal of Electronic Imaging* 18, 043024 (2008)
6. Chan, T., Shen, J.: Mathematical models for local nontexture inpaintings. *Society for Industrial and Applied Mathematics* 62, 1019–1043 (2002)
7. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 1124–1137 (2004)

Fine-Grained Access Control in Object-Oriented Databases

Rahat Masood, Muhammad Awais Shibli, and Abdul Ghafoor

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology,
Sector H-12, Islamabad - 44000, Pakistan
{10msccsmmasood, awais.shibli, abdul.ghafoor}@seeecs.edu.pk

Abstract. Regardless of all the inherent features, Object-Oriented Databases (OODBs) are unable to come up with provable security. The emergence of diversified computing paradigms like Grid & Cloud Computing and Online Social Interaction Platforms has significantly driven the need of OODB technologies with reliable security features. This paper serves to improve the authorization issues within OODBs, by providing an annotation based Fine-Grained Access Control solution. The approach involves realization of Fine-Grained Access Control (FGAC) model for providing granular level access control in Object Oriented Databases.

Keywords: Object-Oriented Technologies, Implicit Authorization, Annotations, Finer Granularity, Least Privileges.

1 Introduction

Object-Oriented Databases (OODBs) are one of the unstructured databases that considerably fulfill almost all the requirements of today's technology. Despite of all the significant features, Object-Oriented databases invokes many provoking thoughts towards the security of data. The number of attacks including code injection, inference attacks, DOS attacks can be launched on these databases as their architecture is totally different from RDBMS. The security loopholes within these databases are not only because of their unique and discernible features but also because a very little attention has yet been given to their security considerations. Therefore, there is substantial need to review the security model and architecture of OODBs from scratch. Access control models to protect the databases have been around since the commercial deployment of these databases. Many new access control models have been proposed and integrated in OODBs however, their mechanisms are either not covering all object oriented features or are not designed to protect data at fine grained level. These shortcomings are either due to some assumptions that these access control models have made or because of the limited focus of these models towards security features. There is need to address the concerns regarding authorization in order to provide a comprehensive access control solution.

Fine-Grained Access Control (FGAC) which is an emerging access control technique for databases plays an important role in this regard. Therefore, the purpose of this paper is to present the concept of FGAC model into OODBs through java annotations. Based on the notion of meta-tagging, we aim to provide customizable FGAC annotations which can be attached to OODB platform for the realization of an authorization mechanism. Our proposed FGAC annotations provide a flexible way to define access control policies on instances, methods, classes and fields of OODBs. We also introduce the concept of customization within these FGAC annotations so that organizations can provide varying level of protection to their object-oriented applications.

This paper is structured as follows. Section 2 presents the relevant work done in the area of authorization of OODBs and FGAC techniques. Section 3 provides detailed description of our annotation based FGAC scheme. Discussion on the effectiveness of annotation based from security perspective is conducted in section 4. Section 5 gives future research directions and concluding remarks on the paper.

2 Related Work

Sohail et al. discusses two well-accepted and mostly deployed access control mechanisms, Discretionary Access Control (DAC) and Mandatory Access Control (MAC) [1]. However, these models have not addressed the security problems of information flow, multilevel object view, inference, aggregation and polyinstantiation. High level authorization specification and efficient implementation of security features in OODBs is presented in [2]. The technique is described in detail; however different access control metrics such as principle of least privileges, consistency, and policy conflict avoiding algorithms have not been verified by this technique. In [3], authors discuss the significance of finer-granularity in databases in terms of subject to object access control, inter-object and intra-object access control, granularity for access unit, class and structure inheritance.

Implicit authorization mechanism provides a finer level access control to OODBs and is based on subject, object and access hierarchies to provide authorization up to unit level of data [4]. Nevertheless, the authors have only proposed the idea, no implementation work has been found for the proof of concept. Surajit et al. [6] proposes a comprehensive design to support predicate authorization at column level. This extended model of SQL applies predicates to authorization grant model and based on the value of predicate, access to columns are provided.

Another technique in which preliminary access control is implemented with firewall functions and then fine-grained access control decisions are made based on user digital credentials and IP address is discussed in [7]. Parameterized views are proposed in [8] to secure databases by transferring identity of users to the database instead of application thus only displaying the relevant data to the user. Row level access control mechanism (a.k.a Oracle VPD) [9, 10] provided by Oracle, uses predicate filtering to secure database. Other techniques [11, 5] have also extended SQL language to define policy types in database. Apart from these techniques, various other FGAC approaches [12, 13] have also been discussed in research literature.

3 Annotation Based Fine-Grained Access Control System

Our work addresses the use of Java Annotations with the purpose to achieve Fine-Grained Access Control within Object-Oriented Databases. The proposed annotation based FGAC is successfully executed on all the OODB concepts (object fields, object methods, object classes and inheritance) and is proved to be applicable on all OODB features.

Being a versatile meta-data, annotations allow us to add additional code at the level of package declarations, type declarations, constructors, methods and fields. For the proposed work, we have introduced three “*FGAC annotations types*”, which in combination with each other; provide access control at granular level. In addition to custom annotations, we also use four pre-defined metadata annotations types for complete consolidation of fine-grained mechanism. The Pre-defined metadata annotations used are: Target, Retention, Documented and Inherited. For the effective realization of FGAC within OODB, three additional annotation types are created, having the distinctive properties of FGAC model. These annotated types are annotated with the code where the database is created.

- **Basic FGAC:** This annotation type of FGAC focuses on checking the attributes of user. Based on the attributes of the user, it is decided whether a user can insert, delete, update or view the data. Database developers just need to annotate this type with any of the object oriented item which they want to protect. This annotation must be used in conjunction with Target, Inherited and Retention annotations.
- **Actions FGAC:** This annotation type is used in combination with BasicFGAC and will be evaluated after BasicFGAC is checked. As indicated by the name, this FGAC annotation type specifies the action that could possibly be executed on any of the Object Orientation concept e.g. read only access to methods of the class, insertion access to parameters of the method having no delete privileges and similar action privileges on constructors, methods, class and fields.
- **Policy FGAC:** This class is extendable in a sense that developers of FGAC annotations can insert security attributes in this annotation type according to their own security requirements. This annotation may contain additional attributes of policies being applied on object data. For-instance, if an administrator wants to draw a policy that data can be viewed to HR and IT Managers but only from Monday to Friday then developers can edit this annotation to reflect this policy. All of three annotations are created in a package named AnnotationFGAC and this package will be imported by other classes to use the FGAC annotations.

Generally an authorization system has three important modules, namely, *Policy Administration Module*, *Policy Enforcement Module* and *Policy Evaluation Module*.

Policy Administration Module, in common, is used to create and maintain policies. For our proposed system, FGAC policies are created by database security administrators of the organization and are pushed to Policy Repository for storage at server

side. **Policy Enforcement Module** is generally used to enforce policies at the arrival of user request. In our annotation based system, policy enforcement is an entity having the capability of interpreting the request from user. **FGAC Policy Engine** (also interpreted as Policy Decision Module) is the third important module of our annotation based system, which act as an intermediate entity between user application and database. This module evaluates the validity of user request upon its arrival and has embedded functions to check for the values of attributes, obligations and conditions associated with a query. A complete FGAC annotation model is illustrated in Fig. 1.

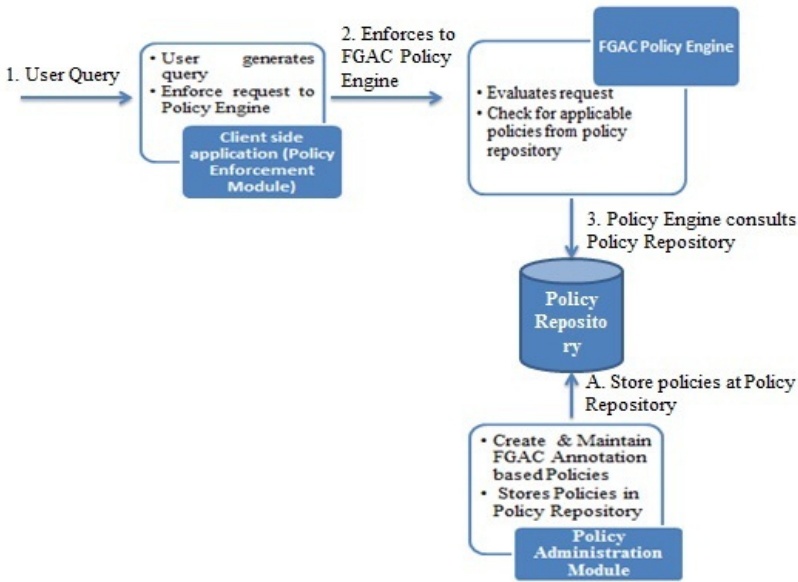


Fig. 1. Annotation Based FGAC Authorization System

Firstly, an administrator creates the FGAC annotations through Policy Administration Module and stores it in policy repository to be used by FGAC evaluation engine. Now, whenever a request is generated by user, Policy Enforcement Module interprets the request and sends it to FGAC Policy Engine for further evaluation. FGAC Policy Engine checks the policies stored in policy repository and based on the decision return by FGAC Policy Engine, request is either sent to database for data retrieval or is rejected on the basis of incorrect attribute values.

4 Discussion

We have formally verified the system with respect to features and functionalities that a reliable FGAC system must hold. Table 1 shows our findings.

Table 1. Features of FGAC model in correspondence with Annotation based FGAC system

FGAC Features & Functionalities		Corresponding Annotation based FGAC feature
Degree of least Privileges	Creating and update Policy Functions and Application Contexts	@BasicFGAC {username, IP Address, Session key, Role Domain} @PolicyFGAC {time & date, extendable context information}
	Association of object security level with user privileges	@BasicFGAC {username, IP Address, Session key, Role Domain} @Target {parameters, constructors, fields, methods, classes} @Inherited @ActionsFGAC {Insert, Editable, Viewable, Delete}
Separation of Duties (SODs)	Categorization of types of controls and processes	@Target {parameters, constructors, fields, methods, classes} @Inherited @PolicyFGAC {time & date, extendable context information}
	Assigning responsibilities to roles and individuals / Specification of roles and responsibilities in a segregated manner	@BasicFGAC {username, IP Address, Session key, Role Domain} @ActionsFGAC {Insert, Editable, Viewable, Delete} @PolicyFGAC {time & date, extendable context information}
	Prevention from role and responsibilities conflicts	Not Addressed Yet
User & Object Privilege Management	Granting and revoking privileges from user / Creation and deletion of users and objects	@BasicFGAC {username, IP Address, Session key, Role Domain} @Target {parameters, constructors, fields, methods, classes}
	Assigning and revoking granularity level to objects / Edit user privileges and granularity level of objects	@Target {parameters, constructors, fields, methods, classes} @BasicFGAC {username, IP Address, Session key, Role Domain} @ActionsFGAC {Insert, Editable, Viewable, Delete}
Policy Conflict Prevention	Implementation of FGAC policy conflict algorithms	Not Addressed Yet

5 Conclusion

In this paper, our aim was to investigate the various characteristics of fine-grained techniques to get them reliably implemented within OODBs. We conclude that effective implementation of this annotation based access control technique will provide provable security among object databases. Further evaluation is needed to successfully deploy annotation based technique in real environments. In addition, we aim at enhancing the FGAC annotations to incorporate features like delegation, revocation, polymorphism, abstraction and access based on multiple attributes.

References

1. Imran, S., Hyder, I.: Security Issues in Databases. In: Second International Conference on Future Information Technology and Management Engineering, China (2009)
2. Xianjun, N.: A Logic Specification and Implementation Approach for Object Oriented Database Security. In: Workshop on Knowledge Discovery and Data Mining, USA (2008)
3. Ambhore, P., Meshram, B.B., Waghmare, V.B.: An Implementation of Object Oriented Database Security. In: Fifth International Conference on Software Engineering Research, Management and Applications, Washington, USA (2007)
4. Hu, H., Xiang, H.: Implicit Authorization Mechanism of Object-Oriented Database. World Academy of Science, Engineering and Technology (2010)
5. Policy Based Access Control PBAC: New Security Features in Sybase Adaptive Server Enterprise. In: Sybase (2003)
6. Chaudhuri, S., Dutta, T., Sudarshan, S.: Fine Grained Authorization through Predicated Grants. In: Third International Conference on Data Engineering. IEEE (2007)
7. Pan, L.: A Unified Network Security and Fine-Grained Database Access Control Model. In: Second International Symposium on Electronic Commerce and Security, China (2009)
8. Roichman, A., Gudes, E.: Fine-grained Access Control to Web Databases. In: Proceedings of the 12th ACM Symposium on Access Control Models and Technologies (SACMAT), France (June 20-22, 2007)
9. Hossz, G., Indraratne, H.: Fine-Grained Data Access for Net-working Applications. In: IGI Global (2009)
10. Nanda, A.: Fine Grained Access Control. In: Proligence. New York Oracle User Group, New York (December 11, 2003)
11. Zhu, H., Lü, K.J.: Fine-Grained Access Control for Database Management Systems. In: Cooper, R., Kennedy, J. (eds.) BNCOD 2007. LNCS, vol. 4587, pp. 215–223. Springer, Heidelberg (2007)
12. Agarwal, R., Bird, P., Grandison, T., Kiernan, J., Logan, S., Rajaibi, W.: Extending Relational Database Systems to Automatically Enforce Privacy Policies. In: Proceedings of the 21st International Conference on Data Engineering (ICDE 2005), pp. 1013–1022. IEEE Computer Society, Washington DC (2005)
13. Franzoni, S., Mazzoleni, P., Valtolina, S., Bertino, E.: Towards Fine-Grained Access Control Model and Mechanisms for Semantic Databases. In: Fifth IEEE International Conference on Web Services (ICWS). IEEE Computer Society, Marriott Salt Lake City Downtown (2007)

Doubly Encrypted Identity-Based Encryption for File Transfer Service

Makoto Sato¹, Masami Mohri², Hiroshi Doi³, and Yoshiaki Shiraishi¹

¹ Nagoya Institute of Technology, Aichi 466-8555 Japan
sato.makoto@niztlab.com,
zenmei@nitech.ac.jp

² Gifu University, Gifu 501-1193 Japan
mmohri@gifu-u.ac.jp

³ Institute of Information Security, Kanagawa 221-0835 Japan
doi@iisec.ac.jp

Abstract. File transfer service demands that what users have to do is as little as possible and that no one can see the contents of a file except for a sender and a receiver of the file. In identity-based cryptography (IBC), one can use receiver's identity (ID) as a public key. There is no need to maintain public key certificates and to communicate preliminarily to get public keys. However, in common identity-based encryption (IBE), the decryption right is concentrated on the Private Key Generator (PKG) which generates every user's private key. Therefore, the PKG is asked for complete trust which is difficult to find in many realistic scenarios. In this paper, we propose an encryption scheme which encrypts a message doubly. By using our scheme, the decryption right is distributed to three servers, and the only receiver can decrypt ciphertext.

Keywords: identity-based encryption, distribution of the decryption right, file transfer service, double encryption.

1 Introduction

In order to transmit a file containing confidential information safely and easily, following two ways may be thought of to realize it with existing techniques. First, encrypting file by using applications supporting public key infrastructure (PKI) may be mentioned [1-2]. However, a receiver has to be issued a public key certificate which certifies the validity of the key by the certificate authority (CA), the trusted authority. And a sender has to communicate with a server maintaining receiver's certificate to get a public key. At the points of these introduction and running costs, the method above has room to be improved. Second, the transmission by large file transfer service may be hit upon. Some services provide secure channels but not encrypt files. So there is a possibility that the contents of a file are seen by service providers. Though there are services encrypt files, such services need both sender and receiver's user registration and authentication. Removing not only sender and receiver's burden but also service provider's one is desirable.

In identity-based cryptography (IBC), one can use receiver's identity (ID) like E-mail address, name and address, and so on as a public key. So, there is no need to maintain public key certificates and to communicate preliminarily to get public keys. This means a sender can encrypt files without troublesome preparations. However, in common identity-based encryption (IBE), the decryption right is concentrated on the Private Key Generator (PKG) which generates every user's private key. Therefore, the PKG is asked for complete trust which is difficult to find in many realistic scenarios [3].

As distribution of the PKG's decryption right, the application of threshold cryptography [4] to the PKG is widely known [3], [5]. When one uses threshold cryptography, generates parts of the master key called share from PKG's master secret key and send them to multiple PKGs. Each PKG can not generate whole private key from its own share. However, thinking of practical use, the application of threshold cryptography to the PKG makes a system complex, and expanding the size of the service gets harder. And there are IBEs to which the possibility of applying threshold cryptography is not sure. In addition, that not only PKGs but also receivers should participate in generating a secret key is a burden for users.

In this paper, we propose an encryption scheme that only a receiver can decrypt ciphertexts and what users have to do is little, aiming at secure and easy file transmission. We organize the rest of this paper as follows. We define terms, explain bilinear maps, Boneh and Franklin's BasicIdent scheme [5] and Computational Diffie-Hellman Assumption in G_I in Section 2. Next, we show our scheme's model and algorithm in Section 3, then define attack models against our scheme and state our scheme's security briefly in Section 4. Section 5 gives conclusion.

2 Preliminary

2.1 Definition of Terms

In this section, we give some definitions of terms used in this work.

Random component / Message component: Random component means the component of a ciphertext not including any information of a plaintext. To the contrary, Message component means the component including information of a plaintext.

Partial decryption / Decryption: Since ciphertexts generated by our scheme is doubly encrypted, we selectively use two terms, Partial decryption and Decryption. Partial decryption means the first time decryption done by two servers named RCD and MCD (we introduce them in section 3.1). Decryption means the second time decryption done by receivers.

2.2 Bilinear Maps

Let G_1, G_2 be two cyclic groups of order q . An admissible bilinear map $\hat{e} : G_1 \times G_1 \rightarrow G_2$ satisfies the following properties in arbitrary $P, Q \in G_1, a, b \in \mathbb{Z}_q^*$:

Bilinear: $\hat{e}(aP, bP) = \hat{e}(P, P)^{ab}$.

Non-degenerate: If P is a generator of G_1 , $\hat{e}(P, P)$ is a generator of G_2 .

Computable: There are efficient algorithms to compute $\hat{e}(P, Q)$.

2.3 Boneh and Franklin's BasicIdent Scheme

Boneh and Franklin's BasicIdent scheme was proposed in 2001, and is standardized in RFC5091 [6]. The scheme was proved to be secure against passive attack in the random oracle model. The scheme is composed of the four algorithms: **Setup**, **Extract**, **Encrypt**, **Decrypt**.

Setup: Given a security parameter $k \in \mathbb{Z}^+$, the algorithm works as follows:

Step1: Output two groups G_1, G_2 of prime order q , and an admissible bilinear map $\hat{e} : G_1 \times G_1 \rightarrow G_2$, and choose a random generator $P \in G_1$.

Step2: Pick a random $s \in \mathbb{Z}_q^*$ and compute $P_{pub} = sP$.

Step3: Choose cryptographic hash functions $H_1 : \{0,1\}^* \rightarrow G_1^*$, $H_2 : G_2 \rightarrow \{0,1\}^n$. G^* denotes the set $G \setminus \{O\}$, where O is the identity element in the group G . The message space is $\mathcal{M} = \{0,1\}^n$ and the ciphertext space is $\mathcal{C} = G_1^* \times \{0,1\}^n$. The system parameters are $params = \langle q, G_1, G_2, \hat{e}, n, P, P_{pub}, H_1, H_2 \rangle$. The master key is s .

Extract: For a given string $ID \in \{0,1\}^*$ the algorithm does:

Step1: Compute $Q_{ID} = H_1(ID) \in G_1^*$.

Step2: Set $d_{ID} = sQ_{ID}$ as the private key corresponding to the ID.

Encrypt: To encrypt $M \in \mathcal{M}$ with the public key ID, do the following:

Step1: Compute $Q_{ID} = H_1(ID) \in G_1^*$.

Step2: Pick a random $r \in \mathbb{Z}_q^*$, and set the ciphertext to be $C = \langle rP, M \oplus H_2(\hat{e}(Q_{ID}, P_{pub})^r) \rangle$

Decrypt: Let $C = \langle U, V \rangle \in \mathcal{C}$ be a ciphertext encrypted with the public key ID. To decrypt C with the private key d_{ID} compute: $V \oplus H_2(\hat{e}(d_{ID}, U)) = M$

2.4 Computational Diffie-Hellman Assumption in G_1

Let G_1 be a group of prime order q . Computational Diffie-Hellman problem (CDHP) in G_1 is to find abP from given random $\langle P, aP, bP \rangle$, where $a, b \in \mathbb{Z}_q^*$, $P \in G_1$. That the computation is believed to be hard is the Computational Diffie-Hellman Assumption.

3 Our Scheme

3.1 Model

We show the model of our scheme at Fig. 1, and define each entity as follows.

Private Key Generator (PKG): The PKG generates a private key corresponding to receiver's public key, ID, with its master secret key, and sends it to the receiver over a secure channel.

Random Component Depository (RCD): The RCD partially decrypts the random component received from the sender with its private key, and sends it to the receiver over a secure channel.

Message Component Depository (MCD): The MCD partially decrypts the message component received from the sender with its private key, and sends it to the receiver over a secure channel.

Sender: The sender generates ciphertexts with PKG’s public parameters, receiver’s ID, MCD and RCD’s public keys, then sends the random component to the RCD, and the message component to the MCD over public channels.

Receiver: The receiver is authenticated by the PKG, MCD, RCD for proving he/she is a valid receiver. After authentication, he/she receives a random component from the RCD, a message component from the MCD, a private key from the PKG, and decrypts a ciphertext.

3.2 Channels

In this paper, we discuss under the assumption that an adversary can get every input entered into each server. This enables senders to do a batch processing. On the other hand, as we consider the PKG’s collusion with an adversary, set channels as follows.

1) Channels between a sender and three servers (PKG, MCD, RCD) are not encrypted.

2) Each Channel between three servers and a receiver is an encrypted secure one.

By setting this way, two advantages are given. First, senders can send ciphertexts under non-secure environment that one cannot use secure channels. Second, it is not necessary that senders and receivers are both online when sending message, since senders do not communicate with receivers directly.

3.3 Algorithm

We show the process flow of our scheme in Fig. 2. In our scheme, we use 4 cryptographic hash functions:

$$H_1 : \{0,1\}^* \rightarrow G_1^*, H_2 : G_2 \rightarrow \{0,1\}^n, H_R, H_M : G_1 \rightarrow \{0,1\}^n.$$

The security analysis will view H_1, H_2, H_R, H_M as random oracles.

Then, we present our scheme by describing the 10 algorithms:

PKG.Setup: Given a security parameter $k \in Z^+$, the algorithm works as follows:

Step1: Output two groups G_1, G_2 of prime order q , and an admissible bilinear map $\hat{e} : G_1 \times G_1 \rightarrow G_2$, and choose a random generator $P \in G_1$.

Step2: Pick a random $s \in Z_q^*$ and compute $P_{pub} = sP$.

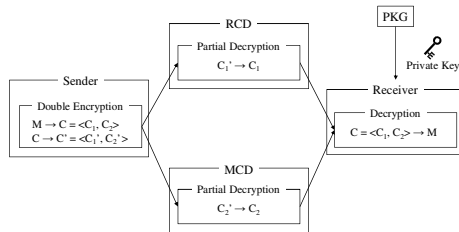


Fig. 1. Model

Step3: Choose cryptographic hash functions H_1, H_2, H_R, H_M . The message space is $\mathcal{M} = \{0,1\}^n$ and the ciphertext space is $\mathcal{C} = G_1^* \times \{0,1\}^n$. The system parameters are $params = \langle q, G_1, G_2, \hat{e}, n, P, P_{pub}, H_1, H_2, H_R, H_M \rangle$. The master-key is s .

PKG.Extract: Given a string $ID \in \{0,1\}^*$, master key s , the algorithm works as follows: Compute $Q_{ID} = H_1(ID) \in G_1^*$ and set $d_{ID} = sQ_{ID}$ as the private key corresponding to the ID.

RCD.KeyGen: Given $b \in Z_q^*$, the algorithm set $(RCD.PK, RCD.SK) = (bP, b)$. $RCD.PK$ and $RCD.SK$ are RCD's public key and RCD's secret key, respectively.

MCD.KeyGen: Given $u \in Z_q^*$, the algorithm set $(MCD.PK, MCD.SK) = (uP, u)$. $MCD.PK$ and $MCD.SK$ are MCD's public key and MCD's secret key, respectively.

Encryption: To encrypt $M \in \mathcal{M}$ with the public key ID, do the following:

Step1: Compute $Q_{ID} = H_1(ID) \in G_1^*$ and pick a random $r \in Z_q^*$.

Step2: Set the ciphertext to be $C = \langle C_R, C_M \rangle = \langle rP, M \oplus H_2(\hat{e}(Q_{ID}, P_{pub})^r) \rangle$

RCD.Encryption: Given a random component C_R , the algorithm works as follows:

Step1: Pick a random $a_1 \in Z_q^*$, and set $C_{R1}' = a_1P$.

Step2: Set $C_{R2}' = C_R \oplus H_R(a_1bP)$, and $C_R' = \langle C_{R1}', C_{R2}' \rangle = \langle a_1P, C_R \oplus H_R(a_1bP) \rangle$.

MCD.Encryption: Given a message component C_M , the algorithm works as follows:

Step1: Pick a random $a_2 \in Z_q^*$, and set $C_{M1}' = a_2P$.

Step2: Set $C_{M2}' = C_M \oplus H_M(a_2uP)$, and $C_M' = \langle C_{M1}', C_{M2}' \rangle = \langle a_2P, C_M \oplus H_M(a_2uP) \rangle$.

RCD.Decryption: Given a message component C_R' and $RCD.SK = b$, set $C_R = C_{R2}' \oplus H_R(bC_{R1}')$

MCD.Decryption: Given a message component C_M' and $MCD.SK = u$, set $C_M = C_{M2}' \oplus H_M(uC_{M1}')$

Decryption: To decrypt C with the private key $d_{ID} \in G_1^*$, compute: $M = C_M \oplus H_2(\hat{e}(d_{ID}, C_R))$

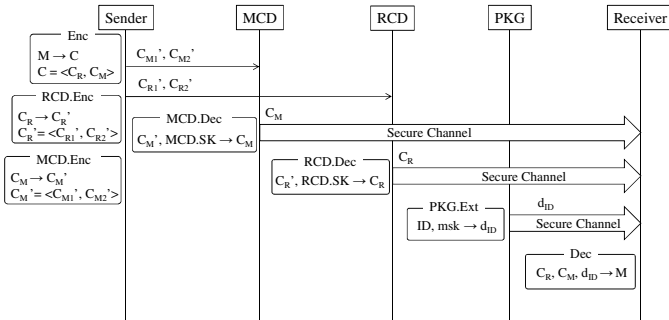


Fig. 2. Process Flow

4 Security

We define security requirements of our scheme as “Suppose there is at least one server which is reliable, our scheme is secure against passive attack”. Note that “reliable” means all servers (RCD, MCD, PKG) do not execute any processing except for ones which is defined as algorithm, and do not leak any secret. In passive attack, adversaries could get ciphertexts corresponding to plaintexts.

According to security requirements, we take account of following three attack models. 1) Both RCD and MCD collude with an adversary. 2) Both MCD and PKG collude with an adversary. 3) Both RCD and PKG collude with an adversary. Obviously, if all servers collude with an adversary, the proposed scheme is not secure.

Next we show how to prove security of the proposed scheme. In attack model 1), the security of proposed scheme is proved in the following proof sequence: proposed scheme \rightarrow BasicIdent. In attack model 2) and 3), proof sequence: proposed scheme \rightarrow CDHP in G_j . We will give a strict proof another time for want of space.

5 Conclusion and Future Direction

The common problem of identity-based cryptography (IBC) is that the Private Key Generator (PKG) which generates all user’s private key can decrypt arbitrary ciphertexts. To the reason, if IBC is used to configure file transfer service which encrypts files, PKG can see the contents of the file. In this work, we proposed an identity-based encryption scheme that only receivers can decrypt ciphertexts. Since the decryption right is distributed to three servers, our scheme is secure even if two out of three servers’ secrets leak.

One future direction is to improve the security of our scheme by applying Fujisaki and Okamoto’s conversion method [7] so as to be secure against active attack.

References

1. The Transport Layer Security (TLS) Protocol Version 1.2, <http://tools.ietf.org/html/rfc5246>
2. PKCS #7: Cryptographic Message Syntax Version 1.5, <http://tools.ietf.org/html/rfc2315>
3. Kate, A., Goldberg, I.: Distributed Private-Key Generators for Identity-Based Cryptography. In: Garay, J.A., De Prisco, R. (eds.) SCN 2010. LNCS, vol. 6280, pp. 436–453. Springer, Heidelberg (2010)
4. An Introduction to Threshold Cryptography, <ftp://ftp.rsasecurity.com/pub/cryptobytes/crypto2n3.pdf>
5. Boneh, D., Franklin, M.: Identity-Based Encryption from the Weil Pairing. *SIAM Journal on Computing* 32(3), 586–615 (2003)
6. Identity-Based Cryptography Standard (IBCS) #1: Supersingular Curve Implementations of the BF and BB1 Cryptosystems, <http://tools.ietf.org/html/rfc5091>
7. Fujisaki, E., Okamoto, T.: Secure Integration of Asymmetric and Symmetric Encryption Schemes. *Journal of Cryptology* 26, 80–101 (2013)

Toward a Honeypot Solution for Proactive Security in Vehicular Ad Hoc Networks

Dhavy Gantsou^{1,*} and Patrick Sondi²

^{1,2}Univ Lille Nord de France, F-59000 Lille

¹LAMIH-DIM UMR 8201, 59313 Valenciennes, France
dhavy.gantsou@univ-valenciennes.fr

²IFSTTAR-LEOST, F-59650 Villeneuve d'Ascq, France
patrick.sondi@ifsttar.fr

Abstract. Vehicular Ad Hoc Network (VANET) is an application of the general concept of Mobile Ad hoc Network (MANET) in the transportation domain. Playing a key-role in Intelligent Transportation Systems (ITS), VANETs are vulnerable to threats that may jeopardize the in-vehicle security, thus subsequently causing accidents. Moreover, on looking at the connected car paradigm that enables interaction between both in-vehicle services as well as devices, and services with outside ones, VANETs will have to inherit security risks of conventional information technologies (IT) systems. This makes it necessary to resort to approaches providing supports and services that are capable to tackle the resulting wide range of security risks. Many works have been dedicated to VANETs security. Mostly the focus has been on data security, relying on cryptographic centric tools in order to ensure privacy, anonymity, and data security. Contrary to this, we propose an approach that aims to proactively address issues beyond data security that could hinder VANETs to become a reality if not properly faced. Although inspired by honeypots used in wired networks, it differs because of the need for tools and methods that should be sufficiently efficient and flexible to ensure context-awareness of security measures.

Keywords: Cyber security, Honeypot, VANET, Intelligent Transport System.

1 Introduction

Many solutions exist for provisioning security in VANETs. However, they often consist in ensuring anonymity [1], [2] and secure communication through group signature [3], [4]. Some solutions resort to Public Key Infrastructure (PKI) for user authentication [5], [6], while others consist in performing message authentication [7], [8]. Nevertheless, almost all existing solutions are based on cryptographic centric solutions.

Connecting vehicles to external resources including cloud services, networks, and user devices requires the provision of better security and privacy. In fact, in order to

* Corresponding author.

provide information from various sources to the driver while improving the road traffic conditions, safety related vehicular real-time control systems and infotainment systems have to communicate. This gateway between vehicles and external resources may lead to security attacks that could have a greater impact than on conventional information technology systems. Therefore, prevention and early detection of attacks are essential. It is important to know both how the attackers proceed, and how the vehicle reacts in case of such attacks. To achieve this goal, another approach that relies on honeypot [9] and initially proposed for wired networks should be investigated. A honeypot is usually a computer system with no production tasks in the network. It is deployed to be probed, attacked, and compromised. This paper presents the results of investigations we made, in order to highlight the adequacy and issues surrounding the use of honeypot in VANETs, although they were initially dedicated to wired networks.

The remainder of the paper is organized as follows: section 2 presents related work on VANETs security and honeypot solutions for VANETs. Section 3 introduces the honeypot concept, and section 4 describes the VANETs honeypot design and the operational scenarios. In section 5, we outline future work and conclude the paper.

2 Related Work

A categorization of the possible attacks that could jeopardize the operations of the ITS shows three classes of scenarios. Aims of these attacks are: transfer of corrupt data, unauthorized access to data, and deny of service. In order to address these security threats, we need to protect not only the data exchanged by the vehicles, but also the entire architecture due to its function in the safety of the overall intelligent transportation system. Moreover, the techniques used by the attackers are more and more elaborated and efficient. There is a lack of a security mechanism that could allow learning on the attackers methods without surcharging the operational vehicular ad hoc network infrastructure.

The approach based on honeypot proposed in this paper aims to address these questions. Only few works have investigated the concept of honeypot. Most of them have been interested only in wired networks [9], [10]. Two other studies propose some scenarios on wireless deployment of honeypot [11], [12], but they do not consider high mobility, thus impeding a straightforward transposition of their work to VANETs. However, the authors of [13] presented some interesting scenarios concerning the use of honeypot over in-vehicle networks. Such an approach does not include all the aspects of VANETs where in-vehicle, inter-vehicle and sometimes vehicle-to-infrastructure communications must be taken into account.

3 The Honeypot Concept

In general, a honeypot is a computing resource [9], whose sole task is to be probed, attacked, compromised, used or accessed in any other unauthorized way. The resource could be essentially of any type: a service, an application, a system or set of systems

or simply just a piece of information/data. The key assumption is that any entity connecting to or attempting to use this resource in any way is by definition suspicious.

Generally speaking, honeypots are often classified by their level of interaction: low-interaction honeypots and high-interaction honeypots. In this work we will use this classification which allows a general description of almost all honeypot tools in only two categories.

3.1 Low-Interaction Honeypots

This kind of honeypot is often designed to give attackers the illusion of an interaction with the real system. Through this interaction, the honeypot will collect useful information ranging from high level data on attackers operational modes to various statistics data on the attacks. There is no need to implement the entire service specification for such a honeypot, but only the functionalities that are essential to accomplish this task in a way that will give an attacker or a worm the illusion that they interact with the real system.

The main advantage of low-interaction honeypots is the simplicity of their implementation and maintenance. An attacker interacting with the honeypot cannot corrupt the real system since he only communicates with a simulated system. Several tools implementing the concept of low-interaction honeypot are available.

3.2 High-Interaction Honeypots

Contrary to the previous kind of honeypot which only simulates some functionalities of the system, a high-interaction honeypot is a classical and complete computer component or system. It may be a server, an access point, a switch or a router. Several honeypots of different types (different hardware/operating systems/ frameworks) can be combined to form a honeynet. The attackers interacting with the honeypot evolve in a partitioned and monitored environment, and they cannot alter the production system. The point separating the honeynet with the remainder of the network is called a honeywall. The main advantage of high-interaction honeypot is that the attackers interact with a real system and manipulate real services, which allows collecting detailed information about their actions, their methods and their motivations. It is also possible to recover the data and the tools that they introduced in the system. However, there are two disadvantages related to high-interaction honeypots: intensive monitoring and maintenance operations on the honeypot, and a risk of compromising the honeypot itself due to its high interactions with a real operating system.

4 Application of Honeypot Concept to Vanets

In order to apply the honeypot solutions to VANETs security, it is mandatory to adapt their design and implementation to the specific characteristics of VANETs such as high mobility, lack of fixed infrastructures in certain areas, and specific threats.

Therefore, a specific solution must be defined for in-vehicle honeypot, and another one for equipping the Road Side Units, in the areas providing ITS infrastructure.

4.1 In-vehicle Honeypot

Regarding the factors that may affect performance and security in the specific context of VANETs, we can notice that:

- due to mobility of the vehicle, the attacker has a limited time-window to perform the attack. Therefore trying to attack such a system will lead to time wasting since the vehicle escapes from his range.
- The honeypots dedicated to classical wireless networks usually simulate hundreds of virtual access points (AP) in order to reduce the probability that the attackers find the one connecting to the production system. Due to the lack of infrastructure, this technique cannot be used in VANETs. Indeed, since there are only few vehicles, the attackers will detect the presence of a honeypot if there are many SSIDs. This leads to the honeypot architecture shown in Fig. 1.

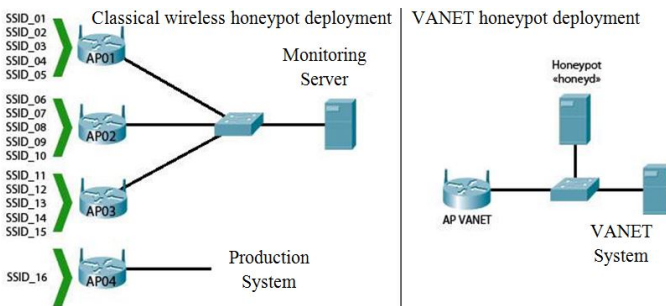


Fig. 1. Classical wireless honeypot architecture vs VANET honeypot

For testings, we chose honeyd, a honeypot simulating virtual machines with unused IP addresses, for its flexibility and low impact on resource consumption.

1) Settings

The following settings are performed on the embedded system inside the vehicle before deploying the honeypot:

- the operating system is a Linux distribution with a 32 bits kernel, running since 1 hour.
- the services activated are olsrd (TCP/UDP port 698) as the ad hoc routing protocol, SSH (TCP port 22), Telnet (TCP port 23), FTP (TCP port 20/21) and httpd (TCP port 80). The honeypot uses the MAC address `iC:7F:E5:FC:5o:9D`.
- The system is configured to route the packets to honeyd.

- Then, the honeypot is created by configuring the file `honeyd.conf`, and by activating the scripts that will simulate de functioning of the services activated. Honeyd does not provide IPv6 support, so we used IPv4 addressing.

2) Collecting Data and Generating Attacks Signatures

Honeyd collects the data related to the attacks through three services, namely:

- Service Level Logging which writes in the log files all HTTP requests/responses, OS version of the attacker's system and the web navigator that he used;
- Packet Level Logging which writes in the log files the intrusion attempts, the protocol used to perform them (TCP, UDP, ICMP), the source, the destination, the ports, and the OS version of the attacker's system;
- Daemon Level Logging manages the logs that are usually written in the system.

3) Testing

We used Zenmap and OpenVas in order, respectively, to see the services activated and to check the failures intentionally introduced to attract attackers. Other failures could be added by introducing PHP code in the web scripts.

4.2 Road-Side Unit Honeypot

The RSUs are positioned along the roads and mainly in towns. Usually they serve as gateways, thus making them a popular target of attackers. Moreover, such units are generally fixed, connected to a permanent power source, and can face the attempts of the same attacker during a long period. One can notice that usually the RSUs have not the same constraints than the vehicles in ITS.

In this context, the strategy of using many fake virtual APs RSUs can be applied. For these reasons, high-interaction honeypot suits better for protecting RSUs.

The honeypot implemented for the testings can be accessed through the Access Points of the network, through the RSU and through the Internet. The honeywall was implemented using ROO honeywall which is based on Sebek and Argos, and which runs over CentOS. It served for filtering and collecting the data related to the attacks.

5 Conclusion

The deployment of a honeypot solution in VANETs faces several situations which are the lack of a centralized infrastructure managing the network, the critical role of security in the system designed for drivers' safety, the importance of using IPv6 to provide better support for mobility, and the variety and complexity of the interactions. We have shown an example of honeypot deployment in VANETs. Though honeyd does not currently provide IPv6 support, it seems to be the appropriate honeypot tool on which to rely in order to implement VANETs compliant honeypots. However, the low-interaction of the different services which are only simulated, and the failures in

the implementations of both the RFC and the TCP/IP stack make honeyd easy to be detected by attackers. These experiments have allowed us to emphasize the issues related to the deployment of high-interaction honeypot in VANETs. Moreover, the honeywall that are generally used generates many data that must be properly collected and analyzed. In our future work, we will investigate the implementation of IPv6 in VANETs honeypot. Moreover, we seek to improve the adequacy of the honeypot solution with the particular context of vehicular ad hoc network.

References

1. Fischer, L., Aijaz, A., Eckert, C., Vogt, D.: Secure revocable anonymous authenticated inter-vehicle communication (SRAAC). In: 4th Workshop on Embedded Security in Cars, ESCAR 2006 (2006)
2. Fonseca, E., Festag, A., Baldessari, R., Aguiar, R.: Support of anonymity in vanets - putting pseudonymity into practice. In: Wireless Communications and Networking Conference, IEEE WCNC 2007, pp. 3400–3405 (2007)
3. Sun, X., Lin, X., Ho, P.-H.: Secure vehicular communications based on group signature and ID-based signature scheme. In: Proceedings of International Conference on Communications (ICC 2007), Scotland (2007)
4. Chaurasia, B.K., Verma, S., Bhasker, S.M.: Message broadcast in VANETs using group signature. In: Proceedings of the IEEE WCSN 2009, pp. 91–96 (2008)
5. Wasef, A., Jiang, Y., Shen, X.: DCS: an efficient distributed-certificate-service scheme for vehicular networks. *IEEE Transactions on Vehicular Technology* 59, 533–549 (2010)
6. Laberteaux, K.P., Haas, J.J., Hu, Y.C.: Security certificate revocation list distribution for VANET. In: Proceedings of the Fifth ACM Workshop on Vehicular Networks (2008)
7. Wen, H., Ho, P.H., Gong, G.: A novel framework for message authentication in vehicular communication network. In: Proceedings of the IEEE GLOBECOM 2009, pp. 1–6 (2009)
8. Zhang, C., Lu, R., Lin, X., Ho, P.H., Shen, X.: An efficient identity-based batch verification scheme for vehicular sensor networks. In: Proceedings of the IEEE INFOCOM 2008, pp. 89–824 (2008)
9. Spitzner, L.: *Honeypots: Tracking Hackers*. Addison-Wesley (2000)
10. Diebold, P., Hess, A., Schäfer, G.: A Honeypot Architecture for Detecting and Analyzing Unknown Network Attacks. In: Proc. 10th Kommunikation in Verteilten System 2005 (KiVS 2005), Kaiserslauten, Germany (2005)
11. Siles, R.: HoneySpot: The Wireless Honeypot. The Spanish HoneyNet Project (SHP) (December 2007)
12. Oudot, L.: *Wireless Honeypot Countermeasures*. Symantec.com (2010)
13. Verendel, V., Nilsson, D.K., Larson, U.E., Jonsson, E.: An Approach to using Honeypots in In-Vehicle Networks (2008)

DWL Tool for Creating a Customized Web-Based System Generator

Ling-Hua Chang^{1,*}, Sanjiv Behl², Tung-Ho Shieh³, and Chin-Chih Ou⁴

¹ Kun Shan University, Department of Information Management
No. 949, Da Wan Rd., Tainan City, Taiwan, R.O.C.
changlh@mail.ksu.edu.tw

² Thomas Edison State College,
101 W. State St., Trenton, NJ 08608-1176
sanbehl@yahoo.com

³ Kun Shan University, Department of Electro-optical Engineering
No. 949, Da Wan Rd., Tainan City, Taiwan, R.O.C.
thshieh@mail.ksu.edu.tw

⁴ Kun Shan University, Department of Electronic
No. 949, Da Wan Rd., Tainan City, Taiwan, R.O.C.
ou@mail.ksu.edu.tw

Abstract. We developed a new customized software tool for automatically generating a complete html program based on the values or parameters inputted by the user. We call it DWL (Dreamweaver Like or DWL for short) since it is similar to Adobe Dreamweaver. We illustrate how it can be used by building a web-based system for a company. We not only offer them this web-based system but also give them this DWL tool which can be used to change the settings to modify it as their requirements change in the future.

Keywords: Web-based Generator, E-commerce Generator.

1 Introduction

We introduce the DWL tool for aiding in the development of web-based systems. We use Barry Boehm's spiral model [1] to implement this tool because the software development process model combines elements of both design and prototyping-in-stages. Thus, it combines the advantages of both the top-down and the bottom-up approaches. We develop a prototype of a web-based system with the help of DWL. This prototype can be used by the system analysts and designers to get an idea of what the end product would be like. It can be tested to see if it meets the system requirements or if it needs to be refined further. If it needs to be refined further or modified, then the system development process would go back to the analysis or design phase, and so on, until a satisfactory prototype is developed.

* Corresponding author.

2 Related Work

Dreamweaver [2] is the most popular tool in the market today, because of the ease of editing and designing a web site. It also generates the html code for the graphical user interface. Thus it is a user-friendly tool. However, since it is not free, we developed DWL for our use.

Dreamweaver only generates html code and not a web server program in say, JSP or PHP, for example. However, there are many web server generators available on the market today like phpMyAdminv [3] and FireStorm/DAO [4]. Alonso [5] shows how to build a generator using Java, XML, and Oracle tools that can produce code for JSP, PL/SQL, and PSP. This domain engineering research project demonstrates that it is possible to successfully implement software product lines in industrial environments using existing tools. Guillen [6] describes a tool (that they call GARP) that automatically generates Web reports from a database scheme. This goal is achieved by the creation of a set of JSP files that contains all the information required by the reports. These files are generated using the XML (eXtensible Markup Language) and XSL (eXtensible Sheet Language) languages. The developed tool is applied for the generation of a Web report, which demonstrates the main features of the tool and the advantages of using XML and XSL for the generation of the required set of JSP files. We plan to use XML and XSL for the JSP generator we would be developing in the near future.

3 DWL System Structure

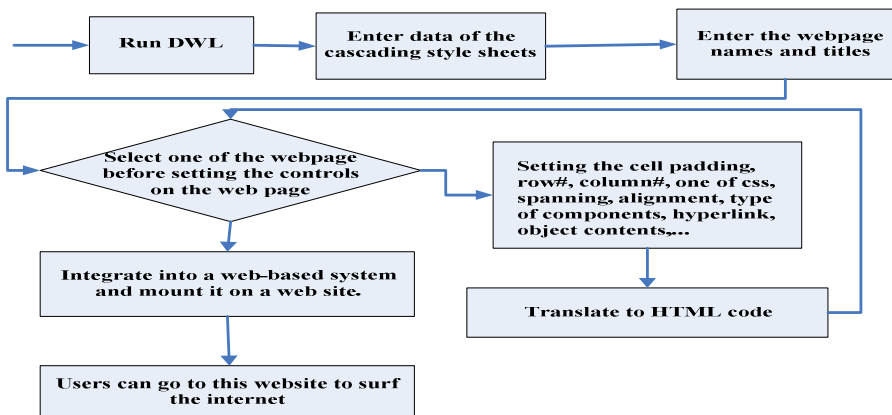


Fig. 1. The process of using DWL to translate a webpage system

Fig. 1 shows how to implement a customized web-based system generator. We need to go through the system analysis and design phase before using DWL. Fig. 1 shows a flow chart of the entire process.

Users use the interface screens to enter the data to create webpage content like text, pictures, hyperlinks, etc.

CSS is a style sheet language used for describing the presentation semantics (the look and formatting) of a document written in html. CSS allows the separation of presentation from structure. CSS can define color, font, text alignment, size, borders, spacing, layout and many other typographic characteristics, and can do so independently for on-screen and printed views. Fig. 2 shows a snapshot of the window that can be used to define the style sheets that would be applied to different parts of a web page. The Back button will take you to the previous window. Delete button will delete the CSS style that is highlighted or selected. New button can be used to specify the name of a new CSS style that will appear in the window. You can select the style sheet whose characteristics you want to specify. Clicking on the Next button will take you to a window shown in Fig. 3 where you can specify the characteristics of the Font, Background, Text, Box, Border etc. After setting the characteristics for each CSS element, you can click on the OK button to go back to Fig. 2. Then you can press the create button to translate a CSS file including all the snapshot styles.

Fig. 3 shows that the CSS *topmenu* is applied to manage five button images; viz. Home, About Us, Products, Concept and Contact (see oval mark in Fig. 4). The left margin is set to 50 pixels as seen in the rectangular mark of Fig.3, which means these five button images will be moved to right 50 pixels. Therefore CSS *topmenu* can be used to arrange images in the correct position.

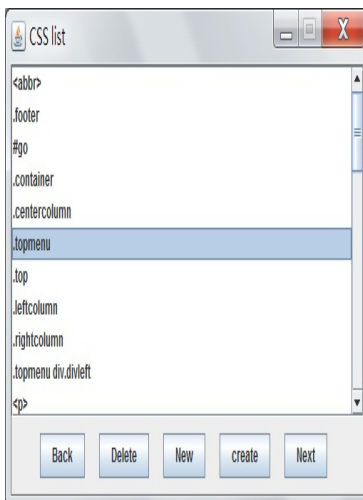


Fig. 2. Window for creating/deleting style sheets

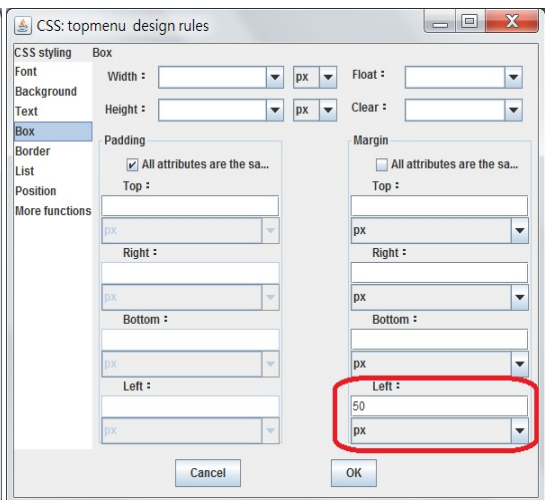


Fig. 3. Specifying the characteristics of each element of style sheet *topmenu*

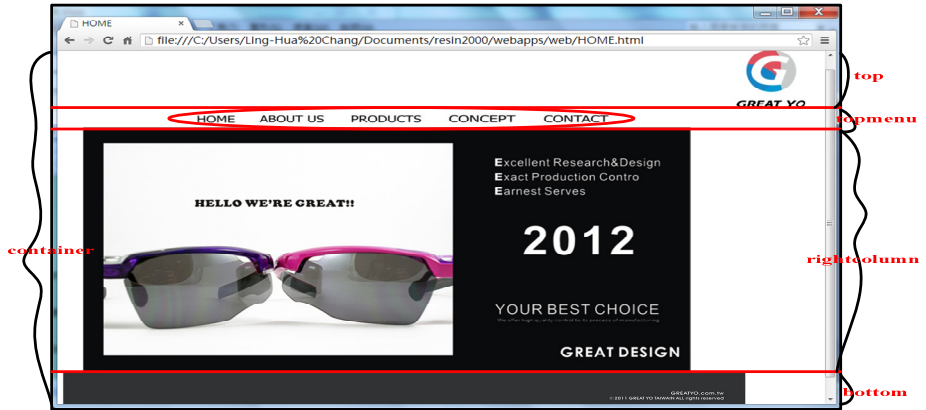


Fig. 4. Divisions of the HOME page of E-Great_Yo

We have used DWL to design a web-based system for a company called Great Yo. The E-commerce web system (named E-Great_Yo for short) has five functions viz. Home, About Us, Products, Concept and Contact. Fig. 4 shows the home page. It shows the company’s logo at the top-right corner and has five buttons for navigating to each of the other web pages. It also displays the company name and logo.

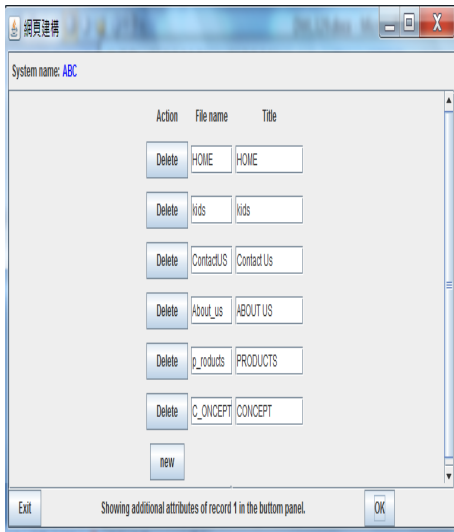


Fig. 5. Setting names of each web page

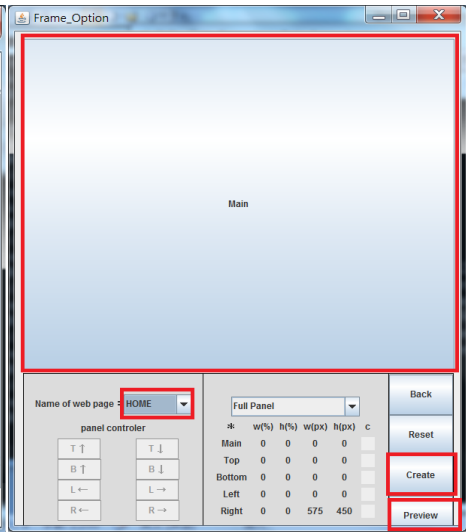


Fig. 6. Selecting a web page

Fig. 5 shows a screen shot of the window that can be used to set the names and titles of each web page. Clicking on the OK button takes you to the window shown in Fig. 6 where you can select the web page whose layout you want to set – the figure shows the HOME page being selected. Clicking on the Main button (the first

rectangular mark in Fig. 6) takes you to Fig. 7 where you can set the layout of HOME page and its components.

Fig. 4 shows how the html document for HOME page is divided into four sections. Each section in turn might be divided further into subsections and so on. We use CSS to manage components in these sections including size, location, color, padding, margin etc. We name the components the same as the name of the CSS style being applied to them. Therefore the complete section is named *container* and its subsections are named *top*, *topmenu*, *rightcolumn* and *bottom* (see the first and second rectangular mark in Fig. 7). The first rectangular mark shows that the complete section of HOME page is set to *container* at field *CSS* and field *Type of Component* is set to *Div* which means this *container* section is divided into several subsections as seen from the second rectangular mark. The second rectangular mark shows that these subsection names are set as *top*, *topmenu*, *rightcolumn* and *bottom*. The CSS fields are set to *top* (for managing the logo), *topmenu* (for managing the five buttons), *rightcolumn* (for the image introducing the company), and *bottom* (for all rights reserved image).

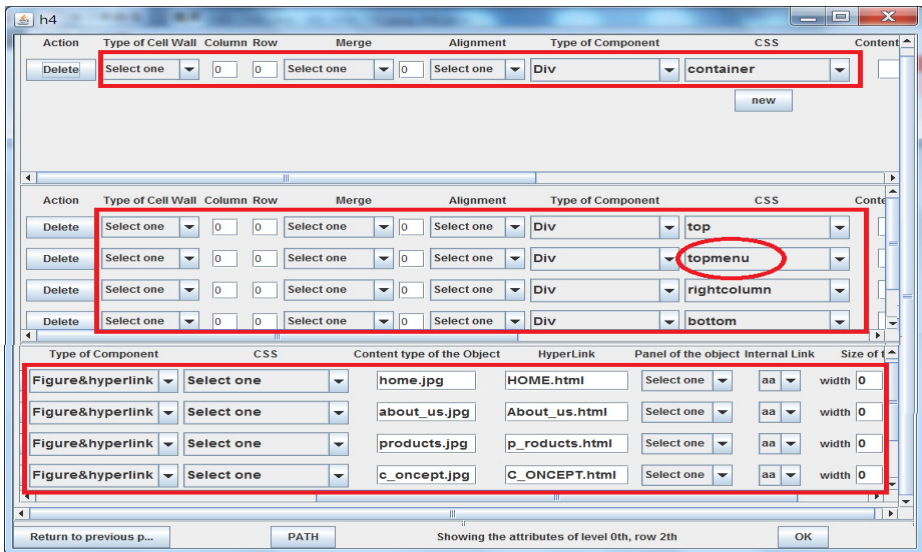


Fig. 7. Setting the components of the HOME page

The topmenu section is divided into five subsections, one for each button image. The *Type of Component* field is set to *Figure & hyperlink* for each button image. The *Content type of the Object* field shows each file name for the five button images and *HyperLink* field shows where you would navigate to when a button is pressed.

4 Experimental Results and Analysis

We compare DWL with Dreamweaver. Dreamweaver has a very good built layout engine and which currently DWL cannot keep up with. That is why they say that it

only takes 10 minutes to complete a webpage. We add some functions into our tool to reduce the time to build a webpage. Note that each webpage in an E-commerce web system has something in common. Take E-Great_Yo as an example. Each webpage in E-Great_Yo has the company's logo at the top-right corner and has five buttons and all rights reserved image at the bottom. So once you build the first page, we can use the common part for building the subsequent pages.

We developed an efficient DWL generator : DWL generates templates which can be used to modify each web page quickly by adding or updating lines of text, figures, hyperlinks, etc. The users of the web-based system can use DWL tool to make changes to it.

5 Conclusions and Future Work

DWL is a convenient tool for generating a web-based system. We illustrate its use by generating a system for a company as described in this paper. It enables customers to better understand the company and its products, leading to increased sales and revenue. In the future, we hope to use this tool to develop customized web-based systems for other small and medium sized businesses.

We have already established the usefulness of DWL. It can save time in writing, debugging and testing a program. It can also reduce the cost of producing software written in html. However, DWL is not as complete yet as we would like it to be. We would like to add a JSP generator to it. Hopefully in the new future our tool can automatically generate a web-based system completely without writing any code.

References

1. Whitten, J.L., Bentley, L.D., Dittman, K.C.: System analysis design methods. McGraw-Hill (2004)
2. Dreamweaver,
<http://www.adobe.com/tw/products/dreamweaver.html?promoid=BPCOM>
3. phpMyAdmin, <http://www.phpmyadmin.net/>
4. FireStorm/DAOandJSPCodeGenerator,
<http://www.codefutures.com/jsp-code-generator/>
5. Alonso, O.: Generating text search applications for databases. IEEE Software, 98–105 (2003)
6. Guillen, M.: GARP: a tool for creating dynamic Web reports using XSL and XML technologies. In: Proceedings of the Fourth Mexican International Conference on Computer Science, pp. 54–59 (September 2003)

The Task-Oriented Circulation Planning

Tzu-I Yang¹, Chorng-Shiuh Koong², and Chien-Chao Tseng¹

¹ National Chiao Tung University, Department of Computer Science, Hsinchu, Taiwan

² National Taichung University of Education,

Department of Computer and Information Science, Taichung, Taiwan

{tiyang, cctsensg}@cs.nctu.edu.tw, csko@mail.ntcu.edu.tw

Abstract. In the field of architecture, circulation refers to the way people move through and interact with a building. From the consumers' aspects, they may pay more attention on the relative position of different commodities and their demands. These behaviors can be deemed as the task-oriented works, which are less discussed comparing to the path planning algorithm. The task-oriented circulation planning may involve the dynamic allocation, demands, and resources, which may be more complex than the exists path problems. In this study, a task-oriented planning of circulation planning is proposed, which is modified and based on the Minimal Spanning Tree algorithm. Properties include circulation, turning points and exit points are involved in planning the optimal solutions based on different kinds of tasks. We also implemented and demonstrated a circulation planning on department of northern Taiwan, and try to provide some suggested circulations by using consumers' shopping lists as input. The proposed application can also be used to optimize the deployment of different commodities.

1 Introduction

Circulation, in the field of architecture and interior design, refers to the way people move through and interact with a building. In public buildings, circulation is of high importance; for example, in buildings such as museums, it is key to have a floor plan that allows continuous movement while minimizing the necessity to retrace one's steps, allowing a visitor to see each work in a sequential, natural fashion. Structures such as elevators, escalators, and staircases are often referred to as circulation elements, as they are positioned and designed to optimize the flow of people through a building. For department and shopping center, on the other hand, the circulation may be affected by the special effort, the demands and the limitation of resources. The circulation of different consumers may vary dynamically according to their tasks. For example, a consumer may have a list of necessities to purchase can be recognized as the task-oriented activities. The circulation of this kind is identified as the task-oriented circulation.

Compare to the exist path planning methodologies, the task-oriented circulation may be more complex than the exist optimal solutions. For example, the travelling salesman problem (TSP) asks the shortest possible route that visits each city exactly once and returns to the origin city with a given list. Although, the tasks can be

deemed as the city in TSP, there exists more than one counter may offer exactly the same merchandise which means it can be solved through TSP algorithms.

Path planning algorithms, for example, Ant colony optimization (ACO) [1, 2] is a probabilistic technique for solving computational problems which can be reduced to finding better paths through graphs. However, the algorithm is a typical probability algorithm, the algorithm parameters are usually given by experiments, that makes the performance of the algorithm relate to individual experience, difficult achieve optimization and not applicable to circulation planning.

Global positioning system (GPS) algorithm, for example, Dijkstra's algorithm [3] is often used in routing as a subroutine in other graph algorithms, or in GPS Technology. There are some researches [4, 5], which based on Dijkstra's algorithm with better efficiency can also help solve the path planning. However, the limitation may be similar to the TSP that if there exist duplicated nodes, counter that offer the same merchandise, may cause the algorithm fails to find the solutions.

Raghavendra et al. [6] proposed File Spanning Tree (FST) which is based on the Spanning tree (MST) In the mathematical field of graph theory, with the definition that connects the root node to some other nodes such that its vertices hold all the required files for executing that program. FST given the idea of nodes that may have more than one resource. However, a single graph can have many different spanning trees. A minimal file spanning tree (MFST) is an FST such that there exists no other FST which is a subset of it.

In this study, we present a Minimum Task Spanning Tree (MTST) algorithm for circulation planning based on the MFST. The counter in our model can be seen as the distributed notes in MFST with one or more resources that can be visited by the consumer. A modified algorithm is applied to the plan that includes circulations, turning points, the entry point, and the exit point.

2 Problem Abstraction

Space allocation for the department store may vary because of different types of commodity, furnishings, counter arrangements and the aesthetics. Different consumers may also have their own shopping list. Hence, the abstraction of the space allocation of the

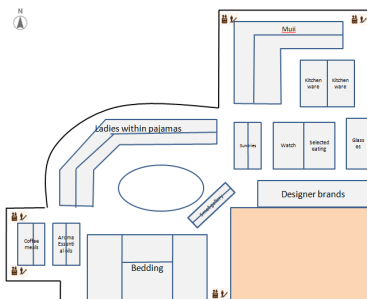


Fig. 1. The layout of the proposed example

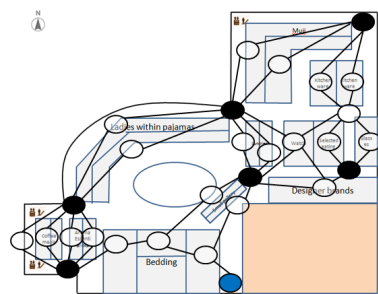


Fig. 2. The abstraction of the proposed layout

department is required. Then it can be transformed into a graph according to the relative position of the different counters (Figure 2). The MTST is used to find the minimal task spanning tree in the abstract layout of proposed example.

3 System Design

We implemented the modified algorithm with graphic user interface (GUI). It is developed by Java, and MySQL as the database (Figure 3). The system includes a manager’s and user’s GUI. The manager’s GUI allows the manager to deploy the correspondent counters, and the user’s GUI is used to simulate the circulation. The manager’s GUI consists of two main modules: inventory and counter deployment. You can set up the corresponding counters to specific positions and the path between nodes by mouse dragging. The distances of links and paths can be set manually or calculate automatically (Figure 4).

The user’s GUI which include the task editing and the shopping list choosing interface. With the MTST algorithm, the relative position on shopping list can be retrieved from the database and simulated the circulation planning, which can provide suggestions for the user. The recommend purchase route will be presented due to the shopping list in different sorts (Figure 4). The System simulates all possible paths to complete the task so that the consumer can complete their goal, obtaining all of their purchases as quickly and smoothly as possible.

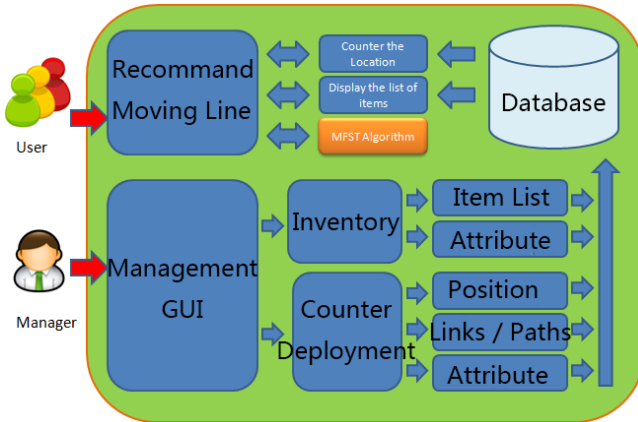


Fig. 3. The system architecture

4 Experiment

In the space allocation and circulation planning of the general department store, the sixth floor of Taipei City Takashimaya department store is taken as an example (Figure 1). The experiment is described as follows:

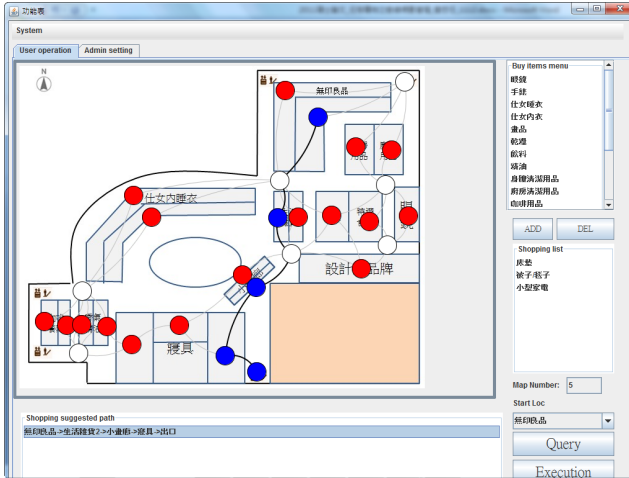


Fig. 4. The demonstration of user interface

1. The assumptions of shopping list include: mattresses, quilts, small appliances, and positions given consumers in the store.
2. Input the purchase list into the program (Figure 4).
3. The system transforms the position of the counter abstraction converted into graphs (Figure 5a).
4. The system abstracts the graphics with resources (Figure 5b).
5. By applying the MTST, we obtain the suggested circulation planning. Assuming that the consumer entry point at the MUJI counters.
6. Obtained to complete the task, the key drivers of consumers line
 - a) MUJI → Watch → Small gallery → Bedding
 - b) MUJI → Sundries → Small gallery → Bedding
 - c) MUJI → Sundries 2 → Small gallery → Bedding
 - d) MUJI → Ladies pajamas → Fragrance of essential oils → Bedding 2

Hence, we have shown the graphs with the solutions for task-oriented users, suggesting optimal moving lines to complete the consumers' task.

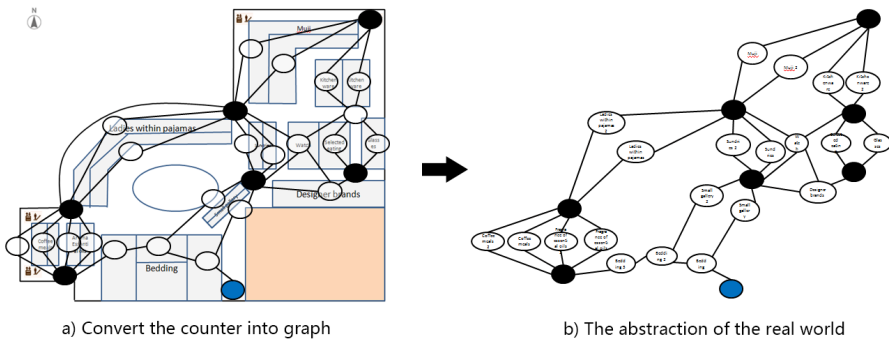


Fig. 5. An example of department store

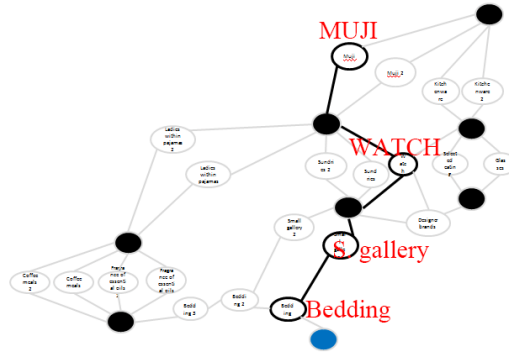


Fig. 6. The example of result

5 Conclusion

The related researches of path planning involve many different methodologies and algorithms. However, task-oriented circulation planning is still rarely seen. In this study, we provide a simulation tool by using MTST algorithm based on the MFST, which can help the consumer to complete task-oriented shopping efficiently. For the proprietors, on the other hand, may help them to draw up new circulations for different kinds of consumers and maximize the profit. In the future research, we will work on adding the dynamics, such as the flows of crows and the obstacles into the graph for reality.

References

1. Colomi, A., Dorigo, M., Maniezzo, V.: Distributed optimization by ant colonies. In: Proceedings of the First European Conference on Artificial Life, pp. 134–142 (1991)
2. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. *IEEE Computational Intelligence Magazine* 1, 28–39 (2006)
3. Dijkstra, E.W.: A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269–271 (1959)
4. Berliner, H.: The B* tree search algorithm: A best-first proof procedure. *Artificial Intelligence* 12, 23–40 (1979)
5. Hart, P.E., Nilsson, N.J., Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics* 4, 100–107 (1968)
6. Raghavendra, C.S., Prasanna Kumar, V.K., Hariri, S.: Reliability analysis in distributed systems. *IEEE Transactions on Computers* 37, 352–358 (1988)

A Framework for Selecting the Optimal Technique Suitable for Application in a Data Mining Task

Haruna Chiroma¹, Sameem Abdul-Kareem¹, and Adamau Abubakar²

¹Department of Artificial Intelligence, University of Malaya, Kuala Lumpur, Malaysia
freedonchi@yahoo.com, sameem@siswa.um.edu.my

²Department of Computer Science, Faculty of Information and Communication Technology,
International Islamic University Malaysia, Gombak, Kuala Lumpur, Malaysia
100adamu@gmail.com

Abstract. This paper presents a conceptual framework for selection of data mining technique based on the 8 selection criteria's: optimization capability, computation complexity, flexibility, interpretability, scalability, ease of problem encoding, autonomy, and accessibility. The framework is suitable for choosing appropriate technique for application in a particular task of data mining. The paper has set the stage for further research work.

Keywords: Data mining task, Neural networks, Evolutionary algorithms, Data visualization.

1 Introduction

The process of uncovering knowledge from very large database is referred to as data mining [1]. Data mining has become a popular methodology of mining knowledge from database due to its flexibility on different databases and accurate results [2]. Its objective is to allow writers of database application to build data mining models such as but not limited to: decision tree, classifier, regression model and segmentation from large data repository. These models can be applied in accomplishing several tasks including predictive and analytic as well as sharing these models with other applications [3]. It is also a discipline that has attracted attention from researchers and it can provide competition superiority to organization by exploring their data warehouse [4]. Techniques of data mining technology are applied in many problem domains [5].

Several techniques of data mining, useful for application in developing models does exist, such as but not limited to: neural networks, evolutionary algorithms, statistical inference. According to [6] data mining itself is never a solution but [7] pointed out that data mining assist decision makers in arriving at intelligent and well-timed decision.

No data mining tool is appropriate for all aspect of data mining task. Choosing a specific technique depends on requirements of the modeler [3] and characteristic of the data [8]. Systematic procedure for choosing the optimum technique for application in data mining task does not exist.

We propose a framework for modelers to select appropriate data mining technique based on their requirements, features of dataset and task classification. To sum up our contribution of proposing framework for selecting suitable data mining technique, we provide a systematic framework for choosing appropriate technique suitable for application in a data mining with laborious trial and error or conjecture selection methods of modelers being reduced.

The remaining sections are: Section 2 presents data mining techniques. Section 3 discusses classification of data mining task. Section 4 describes selection criteria's of data mining techniques. Section 5 provides the graphical representation of the propose systematic framework for selecting data mining technique before concluding and unveiling further research direction in section 6.

2 Data Mining Techniques

Since 1960, data mining techniques has created a branch of applied artificial intelligence [9]. In [10] data mining is defined as the discovery of valuable knowledge from huge databases through mathematical, statistical, artificial intelligence and machine learning methodologies. These methodologies constitute major constituents of data mining techniques in which they have being developing through many years [11]. Artificial intelligence tools are applied in data mining in order to deviate from limitations attributed to traditional statistical tools [12].

Table 1. Data mining techniques

Dimension	Technique	Reference
ANN	Support vector Machine	[5]
	Radial Basis Function network	[13-14]
	Self-organizing maps	[12]
	Fuzzy recurrent neural network	[15]
	Probabilistic neural networks	[16]
	Functional link neural network	[17]
	Modular neural network	[18]
	Bayesian networks	[19]
	Generalized regression neural networks	[20]
	Elman neural network	[21]
	Gene regulatory network	[22]
	Group method of data handling network	[23]
	Fuzzy neural network	[15]
	Time delay neural network	[24]

Table 1. (continued)

EA	Genetic algorithms	[25]
	Genetic programming	[26]
	particle swarm optimization	[27]
RI	Decision tree	[28 – 29]
	K - Nearest Neighbor	[5]
	If-then-else	[30]
	Fuzzy logic	[31]
	Expert systems	[32]
	Decision table, M5 model tree, M5' Rules	[33]
	SI	Linear Regression
SI	Hidden Markov chain model	[34]
	Regression model	[35]
	Set theory	[36]
	Naive Bayes methods	[37]
	Discriminant analysis	[38]
	Principal component analysis	[39]
DV	3D graph, Hygraphs, SeeNet	[40]
	customer map	[41]
	Bar chart, Grid form model, Solar plexus, parallel coordinates, 3D class preserving projection, Smart 2D placement	[42]
	Correlation image	[43]

Data visualization techniques are also applied in data mining to identify relationship within dataset and predict future results [44]. Data mining techniques takes a dataset and established a model of the data. Then, the model is deployed to describe patterns and relationships that exist in the data. Activities of data mining are categories into three namely, discovery, predictive modeling and forensic analysis [11]. Some widely use data mining techniques are presented in Table 1 and their application in data mining is attracting attention [4].

3 Data Mining Task Classification

Classes of data mining task are: classification and prediction, association rule mining, clustering analysis and sequential patterns and time series depending on the problem to be solve by the data mining [11, 45 – 46]. Each of these data mining application

classes is reinforced by algorithms methods in order to extract relevant relationships that can be found in the dataset [10]. The methods depends on the class of problem that they can solve [35].

4 Selecting Data Mining Techniques

Data mining is considered with other relevant issues not just as a secluded task. Issues to be considered in data mining are [34]: First, application domain. For instance, stock market prediction and credit card fraud may require separate data mining techniques. Secondly, features of the data. For example, time series data may require data mining technique with the characteristics of time series sequence. Thirdly, models in the domain. For example in finance there is Kareken – Wallace model through which data mining may take advantage of the model. As reveal in [34], every data mining technique is attributed to its own constraints. Suitability of a technique for application in a particular task depends on some expectations that make it a favorite over others.

5 Propose Framework for Choosing Techniques for Application in Data Mining Task

We present a pictorial representation of the proposed framework for selecting data mining technique suitable for application in a data mining task. Based on previous literature of data mining research [34] and Table 1, this paper proposes a framework for choosing data mining techniques for application in a particular data mining task (Fig. 1). The criteria’s in the framework influence the choice of data mining technique

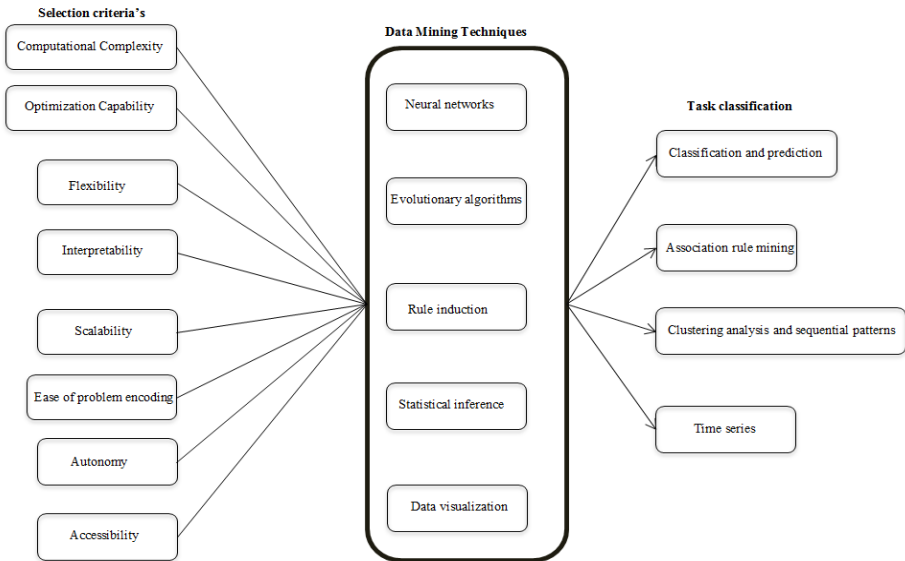


Fig. 1. Framework for selecting data mining techniques

for application in a certain task. The framework considers optimization capability, computation complexity, flexibility, interpretability, scalability, ease of problem encoding, autonomy, and accessibility for selecting a technique. The data mining task in Fig. 1 is classified as discussed in section III. Five points scale (very low = 1, low = 2, medium = 3, high = 4 and very high = 5) are used for comparing data mining techniques based on the 8 criteria's in the propose conceptual framework. For example, each of the 8 criteria's is allocated points according to characteristics of the technique. Technique with highest number of points after summation is selected as appropriate for application in a data mining task such as prediction or classification.

6 Conclusions and Further Research

This paper provides a systematic framework for choosing appropriate technique suitable for application in a particular task of data mining with laborious trial and error or conjecture selection methods of modelers being reduced. Our expectation is that the propose framework will become an increasingly powerful procedure for the data mining community in the selection of appropriate technique.

Our future research direction will be to investigate to what extend through which these criteria's influence the choice of data mining techniques. The research is to be conducted by developing a questionnaire based on the proposed conceptual framework. Conduct a survey by distributing questionnaires in IEEE or ACM data mining conferences to elicit the opinion of experts. Lastly, collected data will be analyze and arrive at conclusions about choosing data mining techniques based on the perception of experts. We hope to further the research presented in this section, in the near future.

References

- [1] Sung, S.Y., Wang, K., Chua, B.L.: Data mining in a large database environment. National University of Singapore, IEEE (1996)
- [2] Shahbaz, M., Athar, S.M., Shaheen, M., Khan, A.: Data mining methodology in perspective of manufacturing databases. *J. Am. Sci.* (2010)
- [3] Aggarwal, N., Kumar, A., Khatter, H., Aggarwal, V.: Analysis the effect of data mining techniques on database. *Advances in Eng. Software* 47, 164–169 (2012)
- [4] Bose, I., Mahapatra, R.K.: Business data mining – a machine learning perspective. *Inf. Management* 39(3), 211–225 (2001)
- [5] Abbasi, Z.K., Soleimanian, F.G.: Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation. *Procedia Technol.* 1, 65–71 (2012)
- [6] Zengin, K., Esgi, N., Erginer, E., Emin, M.A.: A sample study on applying data mining research techniques in educational science: developing a more meaning of data. *Procedia Social and Behavioral Sci.* 15, 4028–4032 (2011)
- [7] Wang, Y.H., Tseng, M.H., Liao, H.C.: Data mining for adaptive learning sequence in English language instruction. *Expert. Syst. Appl.* 36, 7681–7686 (2009)
- [8] Carrier, C.G., Povel, O.: Characterising data mining software. *Intelligent Data Analysis* 7, 181–192 (2003)

- [9] Liao, S., Chu, P., Hsiao, P.: Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Syst. Appl.* 39, 11303–11311 (2012)
- [10] Turban, E., Aronson, J.E., Liang, T.P., Sharda, R.: *Decision support and business intelligence systems*, p. 305. Pearson Education (2007)
- [11] Rygielski, C., Wang, J., Yen, D.C.: Data mining techniques for customer relationship management. *Technol. Society* 24, 483–502 (2002)
- [12] Wang, S.: Application of self-organising maps for data mining with incomplete data sets. *Neural Comput. Appl.* 12, 42–48 (2003)
- [13] Pahariya, J.S., Ravi, V., Carr, M., Vasu, M.: Computational Intelligence Hybrids Applied to Software Cost Estimation. *Int. J. Comput. Inf. Syst. Industrial Management Appl. (IJCSIM)* 2, 104–112 (2010)
- [14] Chen, S.C., Huang, M.Y.: Constructing credit auditing and control & management model with data mining technique. *Expert Syst. Appl.* 38, 5359–5365 (2011)
- [15] Aliev, R.A., Aliev, R.R., Guirimov, B., Uyar, K.: Dynamic data mining technique for rules extraction in a process of battery charging. *Appl. Soft Comput. J.* 8(3), 1252–1258 (2008)
- [16] Mantzaris, D., Anastassopoulos, G., Adamopoulos, A.: Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. *Neural Networks* 24, 831–835 (2011)
- [17] Dehuri, S., Cho, S.: A hybrid genetic based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets. *Neural Comput. Appl.* 19, 317–328 (2010)
- [18] Mitra, P., Mitra, S., Pal, S.K.: Evolutionary Modular MLP with Rough Sets and ID3 Algorithm for Staging of Cervical Cancer. *Neural Comput. Applic.* 10, 67–76 (2001)
- [19] Rivas, T., Paz, M., Martín, J.M., Matías, J.F., García, J.F., Taboada, J.: Explaining and predicting workplace accidents using data-mining techniques. *Reliability Eng. Syst. Safety* 96(7), 739–747 (2011)
- [20] Cravener, T.L., Roush, W.B.: Prediction of amino acid profiles in feed ingredients: genetic algorithms calibration of artificial neural networks. *Ann. Feed Sci. Tech.* 90, 131–141 (2001)
- [21] Ding, S., Zhang, Y., Chen, J., Jia, W.: Research on using genetic algorithms to optimize Elman neural networks. *Neural Comput. Appl.* (2012), doi:10.1007/s00521-012-0896-3
- [22] Ma, P.C.H., Chan, K.C.C.: An effective data mining technique for reconstructing gene regulatory networks from time series expression data. *J. Bioinf. Comput. Bio.* 5(3), 651–668 (2007)
- [23] Oh, S., Pedrycz, W.: Multi-layer self-organizing polynomial neural networks and their development with the use of genetic algorithms. *J. Franklin Institute* 343, 125–136 (2006)
- [24] Kim, H.-J., Shin, K.-S., Park, K.: Time Delay Neural Networks and Genetic Algorithms for Detecting Temporal Patterns in Stock Markets. In: Wang, L., Chen, K., S. Ong, Y. (eds.) *ICNC 2005. LNCS*, vol. 3610, pp. 1247–1255. Springer, Heidelberg (2005)
- [25] Welch, J., Reeves, T.E., Welch, S.T.: Using a genetic algorithm-based classifier system for modeling auditor decision behavior in a fraud setting. *Int. J. Intell. Syst. Accounting, Finance & Management* 7(3), 173–186 (1998)
- [26] Kaboudan, M.A.: Compumetric forecasting of crude oil prices. In: *Proc. IEEE Congress Evolutionary Comput.*, pp. 283–287 (2001)
- [27] Assareh, E., Behrang, M.A., Assari, M.R., Ghanbarzadeh, A.: Application of particle swarm optimization and genetic algorithms techniques on demand estimation of oil in Iran. *Energy* 35, 5223–5229 (2010)

- [28] Andreou, A.S., Papatheocharous, E.: Software Cost Estimation using Fuzzy Decision Trees. In: 23rd IEEE/ACM International Conference on Automated Software Engineering, pp. 371–374 (2008)
- [29] Hwan, J.S., Hoon, C.P., Hyun, S.J.: Applying text and data mining techniques to forecasting the trend of petitions filed to e-People. *Expert Syst. Appl.* 37, 7255–7268 (2010)
- [30] Leung, R.W., Lau, H.C.W., Kwong, C.K.: On a responsive replenishment system: A fuzzy logic approach. *Expert Syst. Appl.* 20, 20–32 (2003)
- [31] Deshmukh, A., Romine, J., Siegel, P.H.: Measurement and combination of red flags to assess the risk of management fraud: a fuzzy set approach. *Managerial Finance* 23(6), 35–48 (1997)
- [32] Vranes, S., Stanojevic, M., Stevanovic, V., Lucin, M.: INVEX: investment advisory expert system. *Expert Syst. Appl.* 13(2), 105–119 (1996)
- [33] Ugür, E.K., Selbas, R., Sencan, A.: Data mining techniques for thermophysical properties of refrigerants. *Energy Conversion and Management* 50, 399–412 (2009)
- [34] Zhang, D., Zhou, L.: Discovering Golden Nuggets: Data Mining in Financial Application. *IEEE Trans. Syst. Man Cybernetics—Part C: Appl. Reviews* 34(4) (November 2004)
- [35] Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., Sun, X.: The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Syst.* 50, 559–569 (2011)
- [36] Changchien, S.W., Lu, T.C.: Mining association rules procedure to support on-line recommendation by customers and products fragmentation. *Expert Syst. Appl.* 20, 325–335 (2001)
- [37] Bermúdez, L., Pérez, J.M., Ayuso, M., Gómez, E., Vázquez, F.J.: A Bayesian Dichotomous, Model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics* 42(2), 779–786 (2008)
- [38] Yeh, I., Lien, C.: The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Syst. Appl.* 36(2), 2473–2480 (2008)
- [39] Brockett, P.L., Derrig, R.A., Golden, L.L.: Fraud classification using principal component analysis of RIDITS. *J. Risk and Insurance* 69(3), 341–371 (2002)
- [40] Shaw, M.J., Subramaniam, C., Tan, G.W., Welge, M.E.: Knowledge management and data mining for marketing. *Decision Support Syst.* 31, 127–137 (2001)
- [41] Woo, J.Y., Bae, S.M., Park, S.C.: Visualization method for customer targeting using customer map. *Expert Syst. Appl.* 28, 763–772 (2005)
- [42] Kopanakis, I., Theodoulidis, B.: Visual data mining modeling techniques for the visualization of mining outcomes. *J. Visual Languages Comput.* 14, 543–589 (2003)
- [43] Dwinnell, W.: Data visualization tips for data mining: pattern recognition provides data insight. *PC AI Mag.* 16(1), 51–57 (2002)
- [44] Weaver, D.C.: Applying data mining techniques to library design, lead generation and lead optimization. *Curr. Opin. Chem. Biol.* 8, 264–270 (2004)
- [45] Larose, D.T.: *Discovering knowledge in data: An introduction to data mining.* John Wiley & Sons, Hoboken (2005)
- [46] Han, J., Kamber, M.: *Data mining: Concepts and techniques.* Morgan Kaufmann, San Francisco (2001)

Discovery of Closed Consensus Temporal Patterns by Group Decision Making

Tony Cheng-Kui Huang

Department of Business Administration, National Chung Cheng University, 168,
University Rd., Min-Hsiung, Chia-Yi, Taiwan, R.O.C.
bmahck@ccu.edu.tw

Abstract. The aggregation of individuals' preferences into a consensus ranking is a decision support problem which has been widely used in various applications, such as decision support systems, voting systems, and recommendation systems. Especially when applying recommendation systems in business, customers ask for more suggestions about purchasing products or services because the tremendous amount of information available can be overwhelming. Therefore, we have to gather more preferences from recommenders and aggregate them to gain consensus. For an example of the preference ranking, $C > A \geq D \geq B$ indicates that C is favorable to A, and A (D) is somewhat favorable but not fully favorable to D (B), where $>$ and \geq are comparators, and A, B, C, and D are items. This shows the ranking relationship between items. However, no studies, to the best of our knowledge, have ever developed a recommendation system to suggest a temporal relationship between items. That is, "item A could occur during the duration of item B" or "item C could occur before item D". This type of recommendation can be applied to the reading order of books, course plans in colleges, or the order of taking medicine for patients. In this study, we propose a novel recommendation model to discover closed consensus temporal patterns, where closed means that the patterns are only the maximum consensus sequences.

Keywords: group decision making, data mining, recommendation systems, consensus temporal pattern, closed pattern.

1 Introduction

In group decision making, aggregating individuals' preferences and obtaining consequences has become a significant and interesting issue, called the group ranking problem. Generally, this type of decision cannot be made by any optimal approaches, requiring members' opinions to subjectively determine a compromising result, which sometimes could be neither efficient nor effective in theory. Members, however, are requested to negotiate with each other to reach a consensus and believe that the decision is (1) fair and open and (2) the best result for us at this time [14].

The group ranking problem was already been investigated for more than two centuries [4] and has been applied to many fields, such as recommendation systems

[5][10][12], machine learning [6], sport tournaments [11], and decision support systems [3][7]. In recommendation systems, we can consider more preference recommendations for items, such as books, CDs, and other products, through the group rating. For example, in the recommendation system of Amazon.com, five users have rated four books, a, b, c, and d, based on their preferences as (1) $c > a \geq d \geq b$, (2) $c \geq a \geq d > b$, (3) $a > c > d > b$, (4) $c > a > d > b$, and (5) $c > a > d \geq b$, where $>$ denotes the former is favorable to the latter, and \geq denotes the former is somewhat favorable but not fully favorable to the latter. As a result, we generally discern that the reading order of the five books recommended by users is c first, and then is a, d, and finally b. As in the example, the goal is to aggregate each user's ranking and produce a result ultimately. In this case, however, user (3) has a different opinion about $a > c$ than the other users. Therefore, the task of integrating these preferences into a consensus solution is one of the major steps in this problem [15].

To extract the aggregated information from individuals' preferences, group ranking models are classified into two types, full and partial rankings. The former requires participants to appraise whole alternatives, whereas the latter allows for cases in which they can only compare with a subset of alternatives. The former representation proposed by Kemeny and Snell [8] was developed for complete and weak orderings. Bogart [2] extended the idea to include partial orderings. Both types have their advantages and are applied to different environments. For example, if we would like to acquire the complete orderings, the preferences for all alternatives must be ranked by participants. Moreover, conflict resolution takes place to obtain the final ranking of all alternatives if the preferences for some alternatives have different rankings. In the partial ranking, yet, participants can naturally give their preferences for a subset of alternatives if they are not required to offer the rankings for the whole alternatives.

Traditionally, the group ranking problem deals with preference among items, meaning which items are preferable to other items. However, the interest in the temporal relationships between items is of importance in real-life circumstances as well. We can give the temporal preferences of items; for example, "item a could occur during the duration of item b" or "item a could occur before item c". The recommendation in temporal preferences can be availed in many cases. To the best of our knowledge, however, no studies have addressed this problem.

To establish the temporal recommendation model, we first have to discuss whether we should consider the complete set of items in rankings or not. For example, a complete set needs to provide the relationships between a and b, b and c, c and d, and a and d and among a, b, and c, b, c, and d, and a, b, c, and d. However, with only the relationships between a, b, c, and d, we can obtain the all above relationships. In light of the description, we follow the idea of mining frequent closed itemsets [16] to uncover closed consensus temporal orderings (or called patterns); that is, a pattern I is closed if no superset of I exists with the same support.

Since no previous research has addressed the topic of recommendation systems for recommending temporal preferences with the closed idea, we first propose an algorithm to find this kind of pattern, closed consensus temporal patterns, in the data mining field. The algorithm is developed by extending the well-known generalized sequential pattern (GSP) algorithm, which uses a stage-by-stage process for

generating frequent patterns [13]. To correspond with the consensus requirement in decision making, however, we modify some steps of the GSP algorithm and develop some special functions for our proposed patterns. Moreover, we tackle a conflict case if the temporal relationship between items suggested by users reaches a specified conflict threshold.

2 The Proposed Algorithm

In this section, we propose the closed consensus temporal mining (CCTM) algorithm to find closed consensus temporal patterns. The CCTM algorithm is developed by modifying the well-known GSP algorithm [13]. In the CCTM algorithm, there are two phases which are repeatedly executed to generate patterns. The first phase generates candidate temporal sequences (candidate sequences hereafter) of length k , denoted by C_k , from the frequent consensus temporal sequences (frequent sequences hereafter) of length $k-1$, denoted by L_{k-1} . In each cycle, one more item and its time relation will be added to each candidate sequence, based on the frequent sequences in the preceding cycle. After finding all candidate sequences, the second phase scans the database once to determine the support of each candidate sequence, and the result consists of all frequent patterns of length k .

The process of finding the candidate sequences for the different lengths of k (the first phase) is described as the following.

1. For $k=1$: The set of candidate sequences of length 1, C_1 , will be generated by enumerating all distinct items of the database.
2. For $k=2$: Traditionally, C_2 is obtained by directly joining L_1 with L_1 ; that is, $C_2=L_1 \times L_1$, where \times denotes the joining operation. However, since the first and second items in C_2 , e.g., i_x and i_y , have two time relations, “&” and “ \rightarrow ”, pairs for the two comparators must be generated. Let us consider an explanatory example. Suppose that $\{i_x\}$ and $\{i_y\}$ belong to L_1 and $\otimes = \{\&, \rightarrow\}$. Then C_2 has a total of two candidate sequences, $\{i_x \& i_y\}$ and $\{i_x \rightarrow i_y\}$. In a word, C_2 can be generated by $L_1 \times \otimes \times L_1$.
3. For $k>2$: Let $s_\alpha = \{a_1 \otimes'_1 a_2 \otimes'_2 \dots a_{k-1} \otimes'_{k-1} a_k\}$ be a k frequent sequence in L_k . According to the *downward-closure property*, the $(k-1)$ -subsequences of s_α , $s_{\alpha_1} = \{a_1 \otimes'_1 a_2 \otimes'_2 \dots a_{k-2} \otimes'_{k-2} a_{k-1}\}$ and $s_{\alpha_2} = \{a_2 \otimes'_2 a_3 \otimes'_3 \dots \otimes'_{k-2} a_{k-1} \otimes'_{k-1} a_k\}$, must be frequent [13], because the support of s_α cannot be larger than the supports of s_{α_1} and s_{α_2} . Therefore, if sequences $\{a_1 \otimes'_1 a_2 \otimes'_2 \dots a_{k-2} \otimes'_{k-2} a_{k-1}\}$ and $\{a_2 \otimes'_2 a_3 \otimes'_3 \dots \otimes'_{k-2} a_{k-1} \otimes'_{k-1} a_k\}$ exist in L_{k-1} , then $\{a_1 \otimes'_1 a_2 \otimes'_2 \dots a_{k-1} \otimes'_{k-1} a_k\}$ must exist in C_k . All the sequences in C_k can be generated by joining the sequences in L_{k-1} this way.

Next, we will discuss how to execute the second phase, i.e., to determine the supports of all sequences in C_k . To this end, a tree structure, called a candidate tree, is adopted as a basis. Basically, it is similar to the prefix tree adopted in previous studies [1][9]. The major difference lies in that the traditional approach connects each tree

branch with an item name, whereas, in the new approach, two components are attached – an item name and a time relation.

In this phase, however, there is an exception procedure in L_2 . If L_2 has two sequences, called s_x and s_y , whose two items, i_1 and i_2 , are identical but whose time relations are not, we then calculate their ratio. The ratio is calculated by $s_x.count/s_y.count$ if $s_x.count \leq s_y.count$; otherwise $s_y.count/s_x.count$, where $s_x.count$ and $s_y.count$ are the supports of s_x and s_y . Accordingly, if the ratio is equal to or larger than a conflict threshold θ , the two sequences, $\{i_x \& i_y\}$ and $\{i_x \rightarrow i_y\}$, are removed from L_2 , and $\{i_x \sim i_y\}$ is added into L_2 . In addition, since the support of $\{i_x \sim i_y\}$ is calculated by adding the supports of $\{i_x \& i_y\}$ and $\{i_x \rightarrow i_y\}$, we adopt a diminishing coefficient ρ to multiply the support of $\{i_x \sim i_y\}$ to decrease the support of $\{i_x \sim i_y\}$. When $k > 2$, we adopt this approach as well. In the following experimental study, we will adjust the two arguments, θ and ρ , to observe the changes in the number of patterns.

3 Conclusions

Most existing approaches focus on giving recommendations of ranking items based on group decisions. Unfortunately, no studies have ever addressed the topic of recommendation systems in temporal preferences between items with a closed idea. This work presents a novel mining approach to find closed consensus temporal patterns to broaden the spectrum for the application of recommendation systems.

Closed consensus temporal pattern mining represents a new and promising research area in recommendation systems and data mining. The model can be extended by considering more temporal relationships, such as meet, during, and overlap, fuzzy temporal patterns, and other kinds of time-related knowledge. Finally, more real-life collected datasets in other types of recommended items can be used to test the proposed model's effectiveness.

Acknowledgement. The author appreciates Krannert School of Management, Purdue University, providing the research resources to support the revision of this paper during his visiting period. This research is supported by the National Science Council of the Republic of China under the grant NSC 101-2918-I-194-008.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings International Conference Very Large Data Bases, pp. 487–499 (1994)
2. Bogart, K.: Preference structures II: distances between asymmetric relations. *SIAM Journal of Applied Mathematics* 29(2), 254–265 (1975)
3. Cook, W.D., Golany, B., Kress, M., Penn, M., Raviv, T.: Optimal allocation of proposals to reviewers to facilitate effective ranking. *Management Science* 51(4), 655–661 (2005)
4. Cook, W.D.: Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research* 172(2), 369–385 (2006)

5. Chen, Y.L., Cheng, L.C.: A novel collaborative filtering approach for recommending ranked items. *Expert Systems with Applications* 34(4), 2396–2405 (2008)
6. Fagin, R., Kumar, R., Sivakumar, D.: Efficient similarity search and classification via rank aggregation. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 301–312. ACM, San Diego (2003)
7. Fernandez, E., Olemdo, R.: An agent model based on ideas of concordance and discordance for group ranking problems. *Decision Support Systems* 39(3), 429–443 (2005)
8. Kemeny, J.G., Snell, L.J.: Preference ranking: An axiomatic approach. In: *Proceedings of Mathematical Models in the Social Science*, pp. 9–23 (1962)
9. Liu, G., Lu, H., Xu, Y., Yu, J.X.: Ascending frequency ordered prefix-tree: efficient mining of frequent patterns. In: *Proceedings of the Eighth International Conference on Database Systems for Advanced Applications*, pp. 65–72 (2003)
10. Li, Q., Myaeng, S.H., Kim, B.M.: A probabilistic music recommender considering user opinions and audio features. *Information Processing and Management* 43(2), 473–487 (2007)
11. Moon, J.: *Topics in Tournaments*. Holt, Reinhart (1968)
12. Pon, R.K., Cárdenas, A.F., Buttler, D.J., Critchlow, T.J.: Measuring the interestingness of articles in a limited user environment. *Information Processing and Management* 47(1), 97–116 (2011)
13. Srikant, R., Agrawal, R.: Mining sequential patterns: Generalizations and performance improvements. In: Apers, P.M.G., Bouzeghoub, M., Gardarin, G. (eds.) *EDBT 1996*. LNCS, vol. 1057, pp. 3–17. Springer, Heidelberg (1996)
14. Smith, R.E.: *Human resources administration: a school-based perspective*, 4th edn. Eye On Education, Inc., Laechmont (2009)
15. Wang, Y.M., Yang, J.B., Xu, D.L.: A preference aggregation method through the estimation of utility intervals. *Computers and Operations Research* 32(8), 2027–2049 (2005)
16. Yan, X., Han, J., Afshar, R.: CloSpan: Mining closed sequential patterns in large databases. In: *Proceedings of the 2003 SIAM International Conference on Data Mining (SDM)*, San Francisco, pp. 166–177 (2003)

A Collaborative Filtering Recommendation Algorithm Based on Tag Clustering

Rujuan Liu* and Zhendong Niu

The School of Computer Science and Technology, Beijing Institute of Technology
Beijing 100081, China
susuliuxiao@gmail.com

Abstract. Social tagging system is applied widely in Web 2.0 nowadays, which is designed to express the user's interest and willingness more accurately. And tag clustering is an important research topic in personalized recommendation of social tagging systems. This paper presents a personalized recommendation method based on tag clustering. In this method, tag clustering is realized by calculating the tag similarity, and recommendation is made based on tag clustering results. Experiments using CiteULike data sets show, proposed method can optimize ranking of objective resources, and help users to discover new resources easier.

Keywords: Social Networks, Tag Clustering, Personalized Recommendation.

1 Introduction

At present, most web sites possess social tagging system, such as delicious, Last, FM, flickr, CiteULike, etc. In these systems, tag is an important source to reflect user data as a link between users and resources. User can label new resources they are interested in using tags, and access resources according to tags other people with similar interests labeled.

However, traditional collaborative filtering recommendation algorithm hasn't considered tag information in recommendation process, so rich personalized information contained in tag can't be acquired, which has unable to adapt the requirement of personalized recommendation in social tagging system. In recent years, the recommendation system based on the tag data gets extensive attention of the academic area and much research has been devoted to address the key issues of tag recommendation algorithm.

A. Hotho et al. proposed FolkRank algorithm based on graph, which mainly uses links information of tag, the resources and users[1] [2], but this approach does not considered personalized connotation contained in user tag, unable to recommend personalization tags for different user. R. Nakamoto et al. present a new model of

* Corresponding author.

collaborative recommendation based on tag, in which users' tag information is extracted as user vector and resources are recommended by computing the similarity of different users[3][4]. S. Zhao et al. take a similar approach to carry out collaborative recommendation by using WordNet to calculate users' similarity [5]. P. Symeonidis et al. propose tag recommendation method based on tensor decomposition in [6]. R. Krestel et al. present a tag recommendation algorithm based on Latent Dirichlet Allocation (LDA) method [7], in which latent topics are extracted from resource according to a relatively stable set of tag. In addition, literature [8] studies how to make personalized movie recommendations to users by using tag information. Jesse Vig et al. study how to make recommendation explanation by using tag [9].

In order to solve the above-mentioned problem, this paper adopted a clustering method based on tag similarity calculation and experiments show that this method can improve the performance of personalized recommendation system. Compared with previous recommendations method, recommendation quality and performance is proved to be enhanced.

Section 2 introduces tag similarity computation method. Section 3 presents personalized recommendation algorithm based on tag clustering. Section 4 shows experiment evaluation result and verify performance of proposed recommendation algorithm. Section 5 concludes the paper and discusses future work.

2 Computation for Tag Similarity

A good tag clustering method is important for personalized commendation, which can make the clustering algorithm fully reflect both user preferences for a particular topic and related degree of resources. Core problem of clustering algorithm is how to compute similarity between tags, which makes tag clustering describe personalized features of tags more accurately. This section gives the formal description of social network and Web resource space vector model. And similarity is calculated based on both user and the resource link.

Social network is a three-dimensional structure made up of users, resources and tags. Its formal description is as follows:

Definition 2.1: A social tagging system is a 4-tuple $D = (U, T, R, A)$ where:

U is a set of users, i.e., $U = \{u_1, u_2, \dots, u_l\}$;

T is a set of social tags, i.e., $T = \{t_1, t_2, \dots, t_n\}$;

R is a set of Web resources, i.e., $R = \{r_1, r_2, \dots, r_m\}$;

A is a set of connections, i.e., $A = \{(u, t, r) | u \in U, t \in T, r \in R\}$, which express links among web users, Web resources and social tags.

The correlation between tag t and resource r reflects marking frequency user mark resource r using tag t . Greater weight means that higher correlation between tag and resource and higher frequency users label the resource using the tag. Weights can be calculated by TF * IDF formula shown as following formula:

$$w(t, r) = \frac{b(t, r) \times \lg(m/n_k + 0.01)}{\sqrt{\sum [b(t, r) \times \lg(m/n_k + 0.01)]^2}} \quad (1)$$

In formula 1, $b(t, r)$ means marking times social tag t for resources r , m is the total number of Web resources, n_k means number of Web resources labeled by social tagging t , and denominator is the normalization factor. Relationship coefficient of tag t_i and t_j marking same resources r can be calculated shown as following formula:

$$Re(t_i, t_j, r) = 1 - \frac{|w(t_i, r) - w(t_j, r)|}{w(t_i, r) + w(t_j, r)} \tag{2}$$

In formula 2, $w(t_i, r)$ and $w(t_j, r)$, represent respectively relationship coefficient of t_i, t_j with resource r . Based on the above analysis, the similarity degree between tags is denoted by $Sim_r(t_i, t_j)$ in formula 3, which can be obtained by comparing their weights in the resource r . $R(t)$ means a set of resources which are marked the tag t .

$$Sim_r(t_i, t_j) = \frac{\sum_{r \in R(t_i) \cap R(t_j)} Re(t_i, t_j, r)}{MAX(|R(t_i) \cap R(t_j)|)} \tag{3}$$

Table 1. “recommender_system” related tags and corresponding similarity in CiteULike

<i>Social tag</i>	<i>similarity</i>	<i>Social tag</i>	<i>similarity</i>
<i>recommender_systems</i>	<i>0.561</i>	<i>cf</i>	<i>0.326</i>
<i>recommender</i>	<i>0.452</i>	<i>recsys</i>	<i>0.318</i>
<i>recommender_systems</i>	<i>0.385</i>	<i>Recommend</i>	<i>0.303</i>
<i>music_recommender</i>	<i>0.352</i>	<i>collaborative_filter</i>	<i>0.295</i>
	<i>0.332</i>	<i>multidimensional</i>	<i>0.276</i>

Algorithm performance evaluation is conducted based on data sets released by CiteULike in this paper. Table 1 shows ‘recommender_system’ related tag set and corresponding similarity using above mentioned formula based on CiteULike data set.

3 CF Recommendation Based on Tag Clustering

At present, different kinds of clustering method have been adopted in many researches. In this paper, a kind of clustering algorithms based on idea of k - means algorithm is used by calculating tag similarity. In this method, k - means algorithm has been improved to increase accuracy of tag clustering.

3.1 Basic Tag Clustering Algorithm

At first, similarity degree between two basic tags is calculated, and tags are classified to an original basic category when their similarity is greater than a certain threshold. Tag clustering algorithm is shown as algorithm 1, which can be run offline.

Algorithm 1: Basic Tag Clustering Algorithm

Input: Unprocessed tag set $T = \{t_1, t_2, \dots, t_i, \dots, t_m, t_n \in T, (i=1, \dots, n)\}$;

Output: Tag clustering set $Clust = \{c_1, c_2, \dots, c_k\}$.

Step 1 Initialization. Put tag pairs (t_i, t_j) marking same resources and their similarity degree into SimRe;

Step 2 Merging clusters with the largest similarity degree in SimRe until the highest similarity difference of tag pairs in cluster is less than the limit;

Steps 3 Take all tags in same cluster as clustering center, and recalculate similarity degree of tag pair in cluster until the clustering center no longer changes;

3.2 CF Recommendation Based on Tag Clustering

The recommendation process is divided into two phases. In first stage, collaborative filtering algorithm can make the preliminary recommendations to provide resource collection for users. In second stage, a new ranking of resource collection is implemented, considering both user model and tag clustering, to generate a personalized recommendation as a result. The whole process is as shown in algorithm 2.

Algorithm 2: Recommendation Algorithm based on tag clustering

Input: User Profile; Unprocessed resource set $R = \{r_1, r_2, \dots, r_j, \dots, r_m, r_j \in R, (j=1, \dots, m)\}$;

Output: Resource collection R'' recommended for user.

Step 1. Calculate correlation degree of tag and resource for each tag, and rank for resources to get basic resource recommendation list R' ;

$$tnr(t, r) = |\{a = (u, t, r) | t \in T, r \in R\}| \quad (4)$$

$$\cos(t, r) = \frac{tnr(t, r)}{\sqrt{\sum_{t \in T} tnr(t, r)^2}} \quad (5)$$

Step 2. Calculate correlation degrees between resources and users taking tag clustering as a bridge. For each $c_i \in Clust$, user interest can be expressed by marking degree of tag subclass $c_i (c_i \in Clust)$ as formula 6 shows;

$$i_d(u, c_i) = \frac{|\{a = (u, t, r) | t \in c_i, r \in R'\}|}{|\{a = (u, t, r) | t \in T, r \in R'\}|} \quad (6)$$

$w_d(r, c_i)$ indicates correlation degrees between resource $r (r \in R')$ and tag subclass $c_i (c_i \in Clust)$ as formula 3.4 shows:

$$w_d(r, c_i) = \frac{|\{a = (u, t, r) | u \in U, t \in c_i\}|}{|\{a = (u, t, r) | u \in U, t \in T\}|} \quad (7)$$

The correlation degree between resources and users can be expressed by $a_d(u, r)$, as formula 7 shows:

$$a_d(u, r) = \sum_{c_i \in Clust} i_d(u, c_i) * w_d(r, c_i) \quad (8)$$

Step 3. Conduct linear weighted average of the result of Step 1 and Step 2, calculate user interest value for resource, and get final list of personalized recommendation, for $r \in R'$, as formula 9 shows:

$$Rank(u, t, r) = \alpha * cos(t, r) + (1 - \alpha)a_d(u, r) \quad (8)$$

Final list of recommendation resources for user u is represented as:

$$R'' = \{ r \mid r \in R', t \in T, u \in U, Rank(u, t, r) \geq \varepsilon \} \quad (9)$$

The minimum threshold ε can be determined by a variety of methods. In this paper, $Rank(u, t, r)$ is calculated and weighted average value of non-zero is calculated to determine ε .

4 Experimental Results

In this paper, data set released by CiteULike is applied for algorithm performance evaluation. After data preprocessing, data contains 3276 users, 30667 papers and 11377 tags. Five layer cross validation (5 - fold cross validation) method is applied. Experiment is proceeded in Java JDK 1.6.0 environment, using MySQL database to store user, tag and literature information, and the correlation between social tagging and resources (here is the paper data) is computed under MyEclipse 6.5. Clustering analysis experiment for tag data is under waikato intelligent analysis environment WEKA3.7.0. Through many times of clustering analysis, tag data sparseness can be solved to a certain extent.

After running algorithm 2, imp [10] method is applied to evaluate the recommendation results at first.

$$imp = \frac{1}{r_p} - \frac{1}{r_b} \quad (10)$$

In formula 10, r_b is the ranking of recommended resources for general CF recommendation algorithm, and r_p is the ranking of recommended resources of recommendation algorithm based on tag clustering. Figure 1 shows experimental results by imp evaluation method. We can see that recommendation effectiveness is best respectively when α value is set to 0.7 in recommendation algorithm based on tag clustering and 0.2 in general CF recommendation algorithm. So if α is set too high or too low, the performance of the algorithm will be impacted. At the same time, it can be demonstrated that commendation algorithm based on tag clustering can improve the ranking of the target resource actually.

In addition, hit ratio is applied to evaluate the recommendation results. Hit ratio refers to proper recommendation probability for each user in test set by the system. Hit ratio is defined as the following formula:

$$hit_per = \frac{hit}{|test_set|} \quad (11)$$

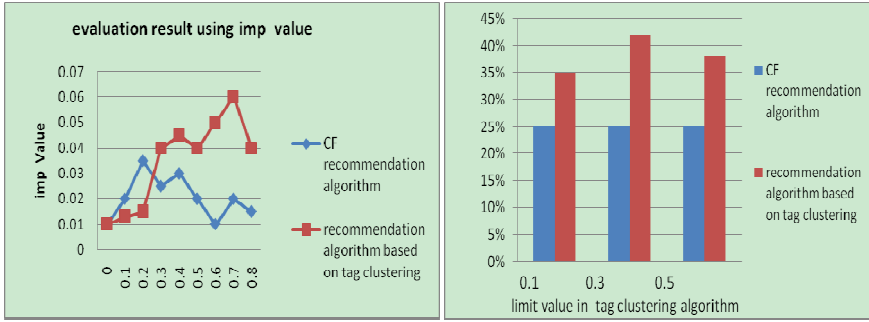


Fig. 1. Evaluation results using Imp value Fig. 2. Evaluation results using hit ratio

Figure 2 is the experimental results by comparing hit ratio. From the diagram we can see that hit ratio of literature in recommendation algorithm based on tag clustering is higher than general CF recommendation algorithm, and when the threshold limit is set to 0.3, hit ratio is the highest in the first method.

5 Summary and Future Work

This paper implements a personalized recommendation algorithm for social tagging system based on tag clustering. Compared with traditional collaborative filtering algorithm, this method has obvious improvement in the quality of recommendation, which has been proved by experiments. On the other hand, this research is still in the early stage, which only recommends resources. Based on current research, future work will continue to optimize the tag clustering algorithm and improve performance of recommended algorithm to meet the need of personalized recommendation service.

Acknowledgement. This work is supported by the National Basic Research Program of China (973 Program No 2012CB7207002), the National Natural Science Foundation of China (No. 61250010), the Program for Beijing Municipal Commission of Education (grant No.1320037010601) and the 111 Project of Beijing Institute of Technology.

References

1. Jaschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag Recommendations in Folksonomies. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 506–514. Springer, Heidelberg (2007)
2. Hotho, A., Jaschke, R., Schmidt-Thieme, L., Stumme, G.: FolkRank: A Ranking Algorithm for Folksonomies. In: Proc. FGIR (2006)

3. Nakamoto, R., Nakajima, S., Miyazaki, J., Uemura, S.: Tag-based Contextual Collaborative Filtering. In: 18th IEICE Data Engineering Workshop (2007)
4. Wang, J., de Vries, A.P., Reinders, M.J.T.: Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion. In: Sigir 2006 (2006)
5. Zhao, S., Du, N., Nauerz, A., Zhang, X., Yuan, Q., Fu, R.: Improved Recommendation based on Collaborative Tagging Behaviors. In: IUI 2008, Maspalomas, Gran Canaria, Spain, January 13-16, pp. 13–16 (2008)
6. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Tag recommendations based on tensor dimensionality reduction. In: RecSys 2008, Lausanne, Switzerland, October 23-25 (2008)
7. Krestel, R., Fankhauser, P., Nejdl, W.: Latent Dirichlet Allocation for Tag Recommendation. In: RecSys 2009, New York, USA, October 23-25 (2009)
8. Sen, S., Vig, J., Riedl, J.: Tagommenders: Connecting Users to Items through Tags. In: WWW 2009, Madrid, Spain, April 20-24 (2009)
9. Vig, J., Sen, S., Riedl, J.: Tagsplanations: Explaining Recommendations Using Tags. In: IUI 2009, Sanibel Island, Florida, USA, February 8-11 (2009)
10. Voorhees, E.M.: The TREC-8 Question Answering Track Report. In: Proceedings of TREC8, pp. 77–82 (1999)

BSTree: An Incremental Indexing Structure for Similarity Search and Real Time Monitoring of Data Streams

Abdelwaheb Ferchichi and Mohamed Salah Gouider

Higher Institute of Management, Department of Computer Science, Tunis, Tunisia
ferchichiabdelwaheb@yahoo.fr,
ms.gouider@isg.rnu.tn

Abstract. In this work, a new indexing technique of data streams called BSTree is proposed. This technique uses the method of data discretization, SAX [4], to reduce online the dimensionality of data streams. It draws on Btree to build the index and finally uses an LRV (least recently visited) pruning technique to rid the index structure from data whose last visit time exceeds a threshold value and thus minimizes response time for similarity search queries.

Keywords: incremental indexing, mining data streams, similarity search, dimensionality reduction, symbolic representation.

1 Introduction

In this paper, a new data stream indexing structure called BSTree is proposed. It is inspired by the recent works of Camerra et al. [2], as well as those of Rudolf Bayer on the Btree. The proposed technique uses the symbolic discretization method, SAX [4], to reduce the dimensionality of the data streams, the sliding windows to perform an online processing of incoming streams. Several changes were also made to the Btree structure to adopt the characteristics of the data streams.

2 Contribution

2.1 BSTree : Balanced Stream Tree

The new approach is based on:

a) SAX for Summarization and Discretization of the Data Stream

SAX representation is detailed in [4], in this sub-section, we would merely introduce the motivations of our choice, which can be summarized as follows:

- The ease of implementation of SAX, unlike other techniques such as DFT and DWT.
- The existence of a large number of algorithms and data structures that allow efficient handling of symbolic representations.

- The symbolic nature of the SAX representation allows the use of the lexicographical order for sorting the data in the BSTree structure.
- The distance measurement MinDist defined by the authors of SAX between two symbolic strings is very close (\leq) to the distance between the original data streams.

b) A Windowing System to Extract the Features of the Data Stream

To browse the continuous data stream for extracting subsets of elements that will be discretized with "SAX" before being inserted into the BSTree index structure, we use a Sliding Window. Whenever w elements are observed and included in the Sliding Window, a new symbol "SAX" is generated and inserted into the BSTree structure.

c) The Indexing Structure BSTree

A BSTree of order m has the following properties:

- The root either is a leaf node or has at least two non-empty sub-trees and at most m non-empty sub-trees.
- Each internal node has at least $\lceil m/2 \rceil$ non-empty sub-trees and at most m non-empty sub-trees ($\lceil m/2 \rceil$ is the smallest integer $> m/2$).
- The number of MBRs in each non-leaf node is one less than the number of non-empty sub-trees of this node.
- Each MBR has a predefined number c of distinct symbols (no duplicates) which are sorted in lexicographical order.

2.2 Index Building

The algorithm Build_Index, in table 1, consists of an insertion procedure to construct the index structure in a single pass and incrementally, and a procedure for removal or more specifically for pruning, to maintain the size of the BSTree structure and respect the constraint on the size of the memory.

The pruning procedure uses the LRV technique (Least Recently Visited). The insertion process continues until the size of the BSTree reaches a maximum height, defined by the user. The BSTree is then pruned by deleting branches whose last visit time exceeds a threshold value $tmpTh$ specified in advance and whose variation may affect the quality of the BSTree index structure after the pruning phase and also construction time of the index.

The Build_Index procedure begins with an iterative call to another procedure, BSTree_Insert, used to insert new SAX symbols created until reaching the maximum height of the tree defined by the user. Then we move to a pruning phase of the BSTree index structure based on a threshold value $tmpTh$ of the last visit time per node (call to the procedure LRV-Pruning). Finally we go back to the construction phase by recursively calling the Build_Index procedure which ensures an incremental construction of the BSTree index structure.

Table 1. Algorithm Build_Index

```

Algorithm. Build_Index (BSTree T,Stream S, c,w, m,
htree, Curr_htree, tmpTh )
// S : The data stream to be indexed
// T : The BSTree index structure
// c : capacity of the MBR
// w : size of the sliding window
/* m : Order of BSTree, it is the maximum number of ele-
ments (MBRs) per node */
// htree : The maximum height of the BSTree structure
/* Curr_htree : height of BSTree when calling to
Build_Index algorithm
/* tmpTh : The threshold used in the pruning procedure
"LRV-Pruning" */
Begin
  While Curr_htree < htree do
    Curr_htree =BSTree_Insert(T,S,c,w,m) ;
  end While
  LRV-Pruning(T, tmpTh);
  Curr_htree = getHauteur(T) ;
  // Recursive call to Build_Index procedure
  Build_Index (T,S,c,w,m,htree,Curr_htree ,tmpTh)
end

```

(a) The Insertion Procedure BSTree_Insert

In the `BSTree_Insert` procedure, a sliding window is used to retrieve the values of the data stream. The SAX procedure uses this window to build the SAX symbol. If the MBR covering this new symbol exists in the index structure, then the SAX symbol will be inserted into the MBR using the procedure `MBR_insert`. Otherwise, the MBR will be searched in a file that contains all possible combinations of the alphabet according to the size of the MBR and the created symbol is then inserted into the MBR before the latter is inserted into the BSTree index structure using the procedure `Index_insert`.

(i) The Procedure MBR_Insert

The `MBR_insert` procedure operates the ascending lexicographic order of symbols in an MBR. The procedure browses the MBR until finding the first element lexicographically greater than the newly built symbol. The new symbol is then inserted into this position by performing a translation of the elements of the MBR that are higher than this symbol.

(ii) The Procedure Index_Insert

The procedure `Index_insert` is the same as the standard insertion procedure used in B-tree, except that the comparison between elements when traversing the exploits the lexicographic order of the symbols.

(iii) The procedure SAX

The SAX procedure is detailed in [4].

(b) The Pruning Procedure LRV-Pruning

In the index structure BSTree, a timestamp T_s is associated with each element in a node N . This timestamp is updated at each visit of the corresponding element during a browse of the tree triggered by a query Q .

During the pruning phase, we use a depth-first search algorithm which browses the tree from left to right and with backtracking. The intuition behind this choice is that the traversing of the BSTree index structure is done from top to bottom (from the root to the leaves) in the tree and from left to right within each node, so for two successive elements of the BSTree structure, two cases may arise during the pruning process:

- Either the timestamp T_{si} of the current node is greater than the pruning threshold $tmpTh$, in this case we must continue the traversing of BSTree structure;
- Either the timestamp T_{si} of the current node is less than the pruning threshold $tmpTh$, in this case also two sub-cases are considered:
 - The timestamp T_{si} of the current node is less than the timestamp T_{si+1} of the next node, so there is a possibility to reach nodes with timestamps equal or superior to the pruning threshold $tmpTh$ and we should, therefore pursue the traversing of the BSTree structure and thus maintain this "bridge" of nodes that can lead us to the other nodes that do not obey to the pruning condition;
 - The timestamp T_{si} of the current node is greater than the timestamp T_{si+1} of the next node, in this case we must perform a backtracking then prune this branch before continuing the traversal of the tree.

Before detailing the pruning algorithm, it is important to note that:

- During the insertion Phase "BSTree_Insert" , any new insertion is performed at the leaf nodes, the timestamp T_s of the newly inserted element is initialized to zero, but in special cases where a balancing of the BSTree index structure is needed, the new element can be inserted in a non-leaf node. In this particular case, the timestamp T_s of the newly inserted element is initialized to the maximum of the timestamps of elements of the children node, and this in order to maintain the sort of the elements of each path in the BSTree structure initially based on the values of the timestamps;
- After each pruning phase, all timestamps T_s are set to zero.

The LRV-pruning procedure begins by recovering the root of the tree T . This value of the root of the tree T and the pruning threshold $tmpTh$ are passed as the effective parameters to the DFS procedure used to traverse the tree using the depth-first technique with backtracking.

The DFS procedure is composed of two phases:

- A search phase of the pruning target node,
- A pruning phase then continues the procedure by referring back to the first phase until pruning all the target branches.

The tree obtained after the pruning phase is not balanced. One of the solutions proposed to our BSTree structure to be balanced after the pruning procedure is to insert the unpruned elements in a new structure that will replace the old one, which will be destroyed at the end of the pruning phase.

3 Experimental Results and Performance Evaluation

The new indexing technique, BSTree, is compared to Stardust [1].

3.1 Evaluation of the Precision

The relationship between the precision and the variation of the radius for range queries is first studied, while setting the size of the basic window TW to 512 and the number of basic windows processed NW to 3600.

Figure 1 shows values of precision for BSTree which are higher than those of Stardust for radii ranging from 0.1 to 1, but still more many other better precision values for BSTree after the pruning stage, this is due to the elimination of unnecessary nodes during the pruning phase which will increase the quality of search operations.

In Figure 2, the relationship between the precision and the size of the alphabet used in the discretization phase (SAX) was studied. We note that Stardust performs more than BSTree for an alphabet of size $\alpha = 4$. For sizes of the alphabet greater than 4 ($\alpha = 6$ and $\alpha = 8$), the precision of BSTree is better than Stardust and can reach values close to 1. This can be explained as follows: If the size of the alphabet increases, the quality of SAX transformation becomes better (the number of windows of data streams that correspond to the same symbol SAX decreases) and vice versa, which will affect the quality of the results produced by BSTree.

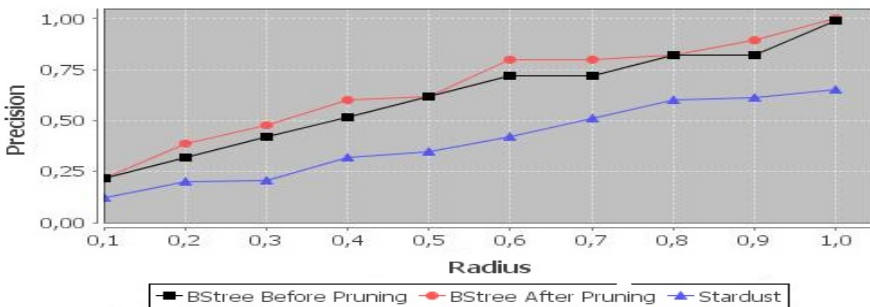


Fig. 1. Precision of BSTree before and after pruning phase VS Stardust by varying the radius of the queries for the "packet.dat"[3] dataset

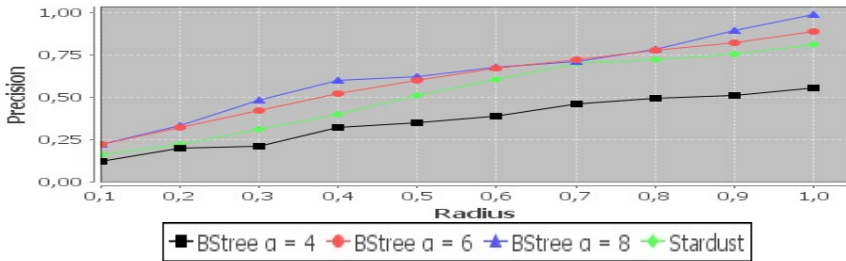


Fig. 2. BSTree precision for sizes of the alphabet SAX $\alpha = 4$, $\alpha = 6$ and $\alpha = 8$ VS Stardust by varying the radius of the queries for the synthetic dataset

3.2 Evaluation of the Recall

Closely to the precision, we also note that BSTree returns values of recall superior to those of Stardust.

4 Conclusion

In this paper, we have proposed a new structure of incremental indexing for similarity search and real-time monitoring of data streams that we have called BSTree. It is based on the technique of symbolic dimensionality reduction SAX and the famous Btree indexing structure.

References

1. Bulut, A., Singh, A.K.: A unified framework for monitoring data streams in real time. In: Proceedings of the 21st International Conference on Data Engineering, ICDE 2005, pp. 44–55 (2005)
2. Camerra, A., et al.: iSAX 2.0: Indexing and mining one billion time series. In: 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 58–67 (2010)
3. Keogh, E., Zhu, Q., Hu, B., Hao, Y., Xi, X., Wei, L., Ratanamahatana, C.A.: The UCR Time Series Classification/Clustering Homepage (2011), http://www.cs.ucr.edu/~eamonn/time_series_data/
4. Lin, J., et al.: Experiencing SAX: a novel symbolic representation of time series. Data Mining and Knowledge Discovery 2, 107–144 (2007)

Learning Teaching in Teaching: Online Reinforcement Learning for Intelligent Tutoring

Fangju Wang

University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract. Intelligent tutoring has started to, and will play an important role in education and training. A challenging task in building an intelligent tutoring system (ITS) is to create and maintain an optimal teaching strategy. In this paper, we present a new technique for addressing this challenge. We cast an intelligent tutoring system as a Markov decision process (MDP), and apply a reinforcement learning (RL) algorithm to learn the optimal teaching strategy through interactions between the system and students. This technique enables the system to teach a student based on his/her studying states, and allows the system to learn the optimal teaching strategy in an online fashion.

1 Introduction

Intelligent tutoring has started to, and will play an important role in education and training. Intelligent tutoring systems (ITSs) have remarkable strengths, including flexibility and low costs. With an ITS, a student may study a subject that is not taught in a local school. It may be easy for the student to incorporate the studies into his/her daily schedule, and to control the pace. Compared with human teachers, an ITS has major advantages. In many situations, studying with an ITS may be less expensive than attending a class. A well-designed ITS may offer more complete and up-to-date knowledge, and it is easier to update the knowledge base of an ITS.

A challenging task in building an ITS is to create and maintain a good tutoring strategy. We develop a new technique for building an ITS. This technique is based on reinforcement learning (RL). RL provides not only an effective way for online learning of the optimal tutoring strategy, but also a construction framework that well suits the nature of a tutoring and learning. We have implemented our technique in an ITS for tutoring calculus. In this paper, we discuss how we cast an ITS as a Markov decision process (MDP), then describe the learning algorithm, and then present the implementation and experiments.

2 Reinforcement Learning

Reinforcement Learning (RL) is a machine learning technique. In an RL algorithm, a learning agent learns knowledge through interactions with the environment [6]. A learning agent is also a problem solver. It learns and updates knowledge while it applies the knowledge to solve problems.

An RL algorithm can be represented as tuple (S, A, P, ρ) , where S is a set of states, A is a set of agent actions, $P: S \times A \times S \rightarrow [0,1]$ is the *state transition probability function*, and $\rho: S \times A \times S \rightarrow R$ is a *reward function*, where R is the set of rewards. Another major component in an RL algorithm is a *policy* denoted by π . The environment is characterized by a *Markov Decision Process (MDP)*.

At time t , the agent is in state $s_t \in S$. A state is an abstraction of a situation. The agent takes action $a_t \in A$ in s_t . The action causes a *state transition* into state s_{t+1} . After taking the action, the agent receives reward

$$r_{t+1} = \rho(s_t, a_t, s_{t+1}). \quad (1)$$

The choice of the action to take in s_t is guided by the policy:

$$a_t = \pi(s_t). \quad (2)$$

π is defined to maximize the long term benefit. Periodically, π is evaluated to see if it “always” chooses the best actions in actual problem-solving, according to certain criteria. If not, it is updated so that it will perform better in the future.

3 Related Work

In [1], the authors examined correlations between dialogue characteristics and applied RL to two corpora of spoken dialogues for tutoring: a human-human corpus and a human-computer corpus. Martin and Arroya used RL to improve the effectiveness of ITSs [4]. They introduced a method of increasing efficiency by way of customization of the hints provided by a tutoring system, Sarma and Ravindran proposed to use RL for building an ITS to teach autistic students [5].

Litman and co-workers developed a spoken dialogue system called ITSPPOKE using RL [3]. ITSPPOKE engaged the student in a spoken dialogue to provide feedback and correct misconceptions. In [2], the authors reported the use of an RL model that allowed the system to learn automatically how to teach to each student individually, only based on the acquired experience with other learners with similar characteristics.

An RL approach was developed for determining what dialogue features were important to a dialogue system for tutoring [7]. The experiments reported in [7] showed that incorporating dialogue factors such as dialogue acts, emotion, repeated concepts and performance played significant roles in tutoring and should be taken into account.

4 Architecture of the Intelligent Tutoring System

To apply RL for learning the optimal tutoring strategy, we build our ITS on the framework of MDP, which includes states, actions, and state transitions. In this section, we discuss how we define the states and actions in the system.

4.1 States

In designing an RL algorithm for an application, defining states is a core task. States and state transitions form the framework in which the learning agent conducts its jobs. The definition of states should be well suited to the nature of the application.

We define the states based on important concepts in the subject of tutoring. For example, in calculus such concepts include “variable”, “function”, “interval”, “limit”, “increment”, “difference quotient”, “derivative”, “differentiation”, and so on. Concepts are related to each other. To study a concept, one must first understand some other concepts. We call the latter *prerequisites*. Information about concept prerequisites is useful in defining states and eliminating invalid states.

For each concept, we define three conditions. Let C be a concept. We have

- the *understand* condition, denoted by C , indicating the system knows that the user understands the concept,
- the *not understand* condition, denoted by $\neg C$, indicating the system knows that the user does not understand the concept, and
- the *unknown* condition, denoted by $?C$, indicating the system has no information whether the user understands the concept or not.

A concept condition can be treated as a proposition. Therefore we can create conjunctive formula of concept conditions to represent the user's studying states with respect to the concepts. For example, we can use formula $(?C_1 \neg C_2 \neg C_3)$ to represent that the user does not understand C_2 and C_3 , and there is no information whether the user understands C_1 . We call the formula associated with a state the *state expression* of the state. In a state expression, the concepts are topologically sorted. A state represents a studying *state* of the student (user): What the student understands and what the student does not.

4.2 Actions

In a tutoring dialogue, asking and answering questions are the primary actions of the user and system. Other actions include those for confirmation, etc.

In our RL algorithm, there are two sets of actions: a set of *user actions* and a set of *system actions*. User actions mainly include the actions of asking questions about concepts, for example “what is a derivative?” In the following discussion, we denote a user action of asking about concept C by $[C]$. For example, when C represents “derivative”, $[C]$ is an action of asking about “derivative”.

The system actions mainly include answering questions about concepts, like “a derivative of a function is the limit of the difference quotient”. We use $\{C\}$ to denote a system action of explaining C . In the RL, a joint action is

$$a = (a_m, a_h) \quad (3)$$

where a_m is a system action and a_h is a user action. Since the ITS is a turn-by-turn system, the two actions have a temporal order: a_m is taken before a_h .

5 Initialization and Update of Dialogue Strategy

5.1 Policy Definition

The tutoring strategy is represented as the policy of the learning agent. A policy can work in the deterministic form of $\pi(s)$ or stochastic form of $\pi(s, a)$. $\pi(s)$ is used to choose the optimal action to take in s . It returns \hat{a} that maximizes function $Q(s, a)$:

$$\pi(s) = \hat{a} = \operatorname{argmax}_a Q(s, a). \quad (4)$$

$\pi(s, a)$ is defined as

$$\pi(s, a_k) = \frac{Q(s, a_k)}{\sum_{i=0}^n Q(s, a_i)} \quad (5)$$

where n is the number of all the possible actions that can be taken in s , and $0 \leq k \leq n$. $\pi(s, a_k)$ returns the probability that given π , a_k may maximize the long term benefit when it is taken in s .

Given policy π , each state $s \in S$ is associated with an *action-value function* $Q(s, a)$ and a *state-value function* $V(s)$. In the following we denote them by Q^π and V^π . $Q^\pi(s, a)$ can be defined as

$$Q^\pi(s, a) = \sum_{s'} P_{ss'}^a V^\pi(s') \quad (6)$$

where s' is the state that the agent perceives after it takes a in s , and $P_{ss'}^a$ is the probability of transition from s to s' after a is taken. V^π can be defined as

$$V^\pi(s) = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a (R_{ss'}^a + \gamma V^\pi(s')) \quad (7)$$

where $R_{ss'}^a$ is the *expected reward* when the agent takes a in s and perceives s' , and γ is a future reward *discounting factor* ($0 \leq \gamma \leq 1$).

5.2 Initialization of π

The Q and V functions define a policy, and the $P_{ss'}^a$ and $R_{ss'}^a$ define the two functions. The creation of $P_{ss'}^a$ and $R_{ss'}^a$ is the primary task in policy initialization.

To initialize π , we calculate $P_{ss'}^{a_m}$ and $R_{ss'}^{a_m}$ for all the system actions a_m . To initialize the $P_{ss'}^a$ and $R_{ss'}^a$, we use hand-crafted tutoring dialogue sessions as training data. The data is for bootstrapping the system. Each session is a sequence of ordered tuples of $(s_t, a_{m,t}, a_{h,t}, s_{t+1}, r_{m,t+1})$, where s_t is the state at t , $a_{m,t} \in a_t$ and $a_{h,t} \in a_t$ are the system and user actions taken in s_t , s_{t+1} is the new state the agents perceive after they take the actions, $r_{m,t+1}$ reward the agent receives after it takes the action.

For given s , a_m and s' , $P_{ss'}^{a_m}$ can be calculated as

$$P_{ss'}^{a_m} = \frac{|the\ system\ agent\ takes\ a_m\ in\ s\ and\ enters\ s'|}{|the\ system\ agent\ takes\ a_m\ in\ a|} \quad (8)$$

where operator $| \cdot |$ denotes “counting the times”. $R_{ss'}^{a_m}$ can be calculated as

$$R_{ss'}^{a_m} = \frac{\sum(\text{reward the system agent received after it takes } a_m \text{ in } s \text{ and enters } s')}{|\text{the agent takes } a_m \text{ in } s \text{ and enters } s'|} \quad (9)$$

In (8) and (9), the counts and sum are calculated from the training data. Once $P_{ss'}^{a_m}$ and $R_{ss'}^{a_m}$ are calculated from the tuples of $(s, a_m, a_h, s', r_m, r_h)$, V^{π_m} and Q^{π_m} are initialized and so is π_m . The policy is ready to choose actions for the system agent to take. The policy is then updated after the agent has started to interact with the user.

5.3 Policy Improvement Method

In the process of tutoring, the agent records the actual user actions and system actions for policy improvement. The recorded actions (raw data) are a sequence of tuples:

$$(s_1, a_{m,1}, a_{h,1}, s_2), (s_2, a_{m,2}, a_{h,2}, s_3), \dots (s_i, a_{m,i}, a_{h,i}, s_{i+1}) \quad (10)$$

In a tuple, for example $(s_i, a_{m,i}, a_{h,i}, s_{i+1})$, s_i is the ‘‘current’’ state, $a_{m,i}$ and $a_{h,i}$ are the actual actions taken in s_i , and s_{i+1} is the ‘‘next’’ state. Based on the recorded raw data, we generate the data used for improving π . The data are sequences of *updating tuples*.

Each updating tuple has a reward. We define *reward function* $\rho(s, a_m, a_h, s')$ for assigning the reward. The parameters are all from the same raw tuple. That is, s is the current state, a_m and a_h are system and user actions taken in s , and s' is the next state after a_m and a_h are taken.

$$\rho_m = \begin{cases} r' & \text{if } a_m \text{ is accepted by } a_h \text{ in } s, \text{ and leads to } s' \\ r'' & \text{if } a_m \text{ is rejected by } a_h \text{ in } s, \text{ and leads to } s' \end{cases} \quad (11)$$

where r' and r'' are two scalar values satisfying $r' > r''$. (We choose $r' = 2$ and $r'' = 1$ in our experiments.)

After the sequence of updating tuples is created, probabilities $P_{ss'}^a$ and expected rewards $R_{ss'}^a$ are updated by using the data. Then a policy iteration process [6] is executed on the updating tuples to update $V^\pi(s)$, and thus improve π , in order that π could guide the system agent to choose actions that are better accepted by the user.

6 Implementation and Experiments

We developed an experimental system, which consists of three modules responsible for tutoring, speech recognition, and input understanding. The latter two are not discussed in this paper. The module for tutoring implements the RL algorithm. Besides tutoring, it is also responsible for improving the tutoring strategy.

The experiments were conducted with 14 students who had different levels of knowledge about the subject. The system teaches a student at a time. The order for choosing the students was random. Each student interacted with the system for about 45 minutes, asking about 65 questions. The system responded to every question.

In evaluating the performance in choosing responses, the performance parameter is the times that the student ‘‘rejected’’ the system actions. It can be observed that at the

beginning the numbers of rejections were about half of the student questions, and it decreased to less than one third. Statistical analysis has indicated that with the system teaches more students, less and less answers were rejected. This implies that the system has learned better teaching skills when it taught students.

7 Summary

The novelty of our technique includes its definition of states based on the important concepts in the subject, and online learning optimal teaching strategy. Our state definition is different from those in the existing work, in which states are commonly defined by a number of “features”, such as correctness, certainty, concept repetition, frustration, percent correct, etc [7]. Our definition depicts the system's knowledge about the student's studying states. It has the advantages: A state defined by a state expression is Markovian. This property makes the state definition suitable for RL. Also, in tutoring a student, information about what the student understands is essential in choosing a response. Also, our technique is characterized by online learning.

Intelligent tutoring is an appealing education approach. However, its current application has been limited by challenging tasks in building practical ITSs. Our technique may contribute to wider application of ITSs in future education.

References

1. Forbes-Riley, K., Litman, D., Huettner, A., Ward, A.: Dialogue-learning correlations in spoken dialogue tutoring. In: Proceeding of the 2005 Conference on Artificial Intelligence in Education (2005)
2. Iglesias, A., Paloma Martinez, P., Fernando Fernandez, F.: An Experience Applying Reinforcement Learning in a Web-Based Adaptive and Intelligent Educational System. *Informatics in Education* 2(2), 223–240 (2003)
3. Litman, A.J., Silliman, S.: Itspoke: an intelligent tutoring spoken dialogue system. In: Proceedings of Human Language Technology Conference 2004 (2004)
4. Martin, K.N., Arroyo, I.: AgentX: Using Reinforcement Learning to Improve the Effectiveness of Intelligent Tutoring. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 564–572. Springer, Heidelberg (2004)
5. Sarma, B.H.S., Ravindran, B.: Intelligent Tutoring Systems using Reinforcement Learning to teach Autistic Students. In: *Home Informatics and Telematics: ICT for The Next Billion*, vol. 241, pp. 65–78. Springer, Boston (2007)
6. Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge (2005)
7. Tetreault, J., Litman, D.: Using reinforcement learning to build a better model of dialogue state. In: Proceedings of ECAL 2006 (2006)

Minimax Self Playing Game Alquerque

Ishtiaq Ahmed^{*}, Donghai Guan, Md. Nasir Uddin Laskar, and Tae Choong Chung

Department of Computer Engineering
Kyung Hee University, South Korea

{ishtiaq.khu,nasir,tcchung}@khu.ac.kr, donghai@oslab.khu.ac.kr

Abstract. Playing game in a computer is always fascinating and entertaining. But playing against a computer player is much more fun and challenging than a human player. We have proposed a simple algorithm for playing a repeated game named Alquerque in this paper. We show that by following this algorithm it is proved that the game playing strategy tends to minimum loss and maximum win against the human. Our analysis is the evidence of minimax search theorem and well-chosen combination of probability for solving this game. In this paper we have developed an intelligent computer player successfully that uses min-max search technique, strategy utility function which leads to win and draw in the worst scenario for playing the ancient game Alquerque.

Keywords: Minimax, utility function.

1 Introduction

In the last century, researchers focused on designing and implementing games which have artificial intelligence. Researchers also tried to implement game theoretic values which represent a game in win, lost and drawn condition from the perspective of the player who has to move first.

This paper is concerned about the problem of constructing a computer program for which it will be able to play Alquerque. Minimax, pruning techniques are used to build the artificial intelligence of the machine to play like human, which can think rationally and act according to the condition. In the last 20 years[2] significant steps have been done to solve large number of games. Among these games, different types of board games are in the leading portion where artificial intelligence has been applied in a regular basis. In our paper, one human player will play against the computer. After making one movement by the human player, our program will make a legal move by analyzing the human's turn and selects the best possible move which leads to win or at least a drawn condition by our proposed algorithm. Our algorithm is based on the basic principle of min-max strategy and pruning the sparse tree by its utility function, leads to optimal solution for each and every step.

The paper is organized as follows. Section 2 describes the background information and related work of the games and the rules to play this game rationally. Section 3

^{*} Corresponding author.

introduces our different types of searching strategies which leads to win or at least a drawn. Section 4 depicts our result section where different types of comparison are done according to different conditions.

2 Background and Related Work

Alquerque is a strategy based two players board game and is thought to be originated from the ancient Middle East around 3,000 years ago[1]. This game is originally known as Quirkat and it's believed to establish its current name during the Moorish reign of Spain. We will explain the rules of playing Alquerque bellow. The aim of the game is to capture opponent pieces. If one player captures the opponent player's all pieces, then he will win and lose vice versa. The game is drawn if neither player wins or loses.

Alquerque board consists of lines and intersections. The lines represent among which one piece can move through it during the game. There are different types of intersections. Someone has at most 5 connections to go though. On the other hand, someone has only three paths remaining. One piece can move horizontally, vertically or diagonally to its adjacent intersection. When the board was first arranged there are twelve pieces of each individual player. The black circle represents Computer player and the white circle represents the human player. The middle intersection is kept empty when the game is started.

The game is played by turns. Finishing the first players turn, the second player is allowed to make turn and vice versa. There are two kinds of moves which are consisted of capturing and non-capturing moves. The white piece which represents the human player pieces are allowed to turn his player first. Each piece can only move to the adjacent intersections except the backyard. In the non-capturing move a piece can move its adjacent empty intersection point whereas the capturing move, a piece can capture opponent piece by jumping over an opposite piece and keep doing unless there is no other capturing move. But these capturing moves are not mandatory which means this movement is fully optional. The goal of this game is to eliminate the opponent's pieces.

This game will be played as long as the opponent pieces are completely removed. The player who first removes opponent's entire pieces wins the game. The game will be drawn if none can able to win. This is achieved by the repetition of the same position with the same player to move.

3 Our Proposed Searching Strategy

The min-max algorithm is applied here for playing Alquerque with a computer against human. It is possible to know from a certain point, what are the next available moves. It means that the details information is available from a given point. So every player knows everything about the possible moves of himself and the opposite. At the beginning we construct the search tree which is required to represent searches. From a certain node in the search tree, the branches are constructed until no more decisions

are possible[4]. The computer player receives the maximum value of its children where as the human player selects minimum value[2, 3]. So, the computer player tries to select the move with the highest value whereas human player selects the move with the lowest value.

3.1 The Game Engine and Calculating the Best Move

In the core process there is a while loop, where the turns are altered one by one. Initially the turn goes for human and after successful movement the chance of movement is handled to computer. The loop will be continued unless there is no win condition or drawn condition. This game engine algorithm is depicted on the algorithm 1. Initially there is a 5×5 matrix is given[1] and 12 pieces of each player is organized on the board. Whenever a successful movement is done by a human then the turn goes to computer, analyzing each and every pieces of computer it selects the best move by which it executes its turn. So for each of the pieces of computer there is a value which is obtained by the function *calculateBestMove* and the computer selects the value which has the lowest one.

Algorithm 1. Game Engine	Algorithm 2. calculateBestMove
<p>Let the board is 5×5, computer's pieces 'C', human's piece 'H'.</p> <ol style="list-style-type: none"> 1. actP=human&&oppP=Comp 2. while (game continue)do 3. if P = human then 4. {from,to,inter} = getActfromHuman 5. Board[from] = empty; 6. Board[to] = 'H' ; 7. if action = canMove then Board[inter] = ϕ 8. else 9. for c \in C 10. calBestMove(pos, B) 11. executeBestMove(); 12. tempP = activeP 13. ActiveP = CP; oppP = tempP; 14. if canNotMove break; 	<p>Input: player P, Board B, position Pos Output: Solution to best-Move</p> <ol style="list-style-type: none"> 1. attPaths=findAttPath() 2. bestAttPaths = ϕ 3. for all attPaths 4. if attPaths= WIN_VALUE then 5. return attPath 6. else 7. attPath+= findOppAtt(oppP, B) 8. if P=human then return min(bestAttPaths) 9. else return max (bestAttPaths)

To interpret Algorithm 2, assume the 5×5 board is organized by the pieces of human and computer. Assume the piece on (i, j) on the board, from where we have to make choice a movement. This algorithm guides to choose the best movement among several possible choices, calculating proper utility value. At the beginning, the possible attacking paths are calculated. If there is any winning condition, then it returns that movement as the best one. Otherwise it returns the path which has lowest utility value when the player is human and maximum for the vice versa.

3.2 Possible Attacking Paths and Finding Opponent Attack

Finding proper attacking paths from a give point, this algorithm generates the possible paths. As we are already acknowledged that for removing an opponent piece the reward value is +1, for successful movement without removing any piece it simply added nothing with its reward value whereas killing attack [attacking paths] always add +1 with its reward value. In our proposed finding attacking path algorithm we only recurs the same procedure when it successfully removes opponent piece.

For a successful movement our proposed technique also examines whether its future outcome is relatively better than the current one. After examining the all possible attacking paths, it finds any winning condition then it returns the negative WIN_VALUE when player is human and positive WIN_VALUE vice versa[4]. Here the WIN_VALUE is a constant one and can't be changed during the program execution. But when this win condition is not satisfied it only returns the negative value of the lowest attacking path cost whereas the largest attacking path cost for human. This path cost means how many opponent pieces are occupied instead of itself.

Algorithm 3. findAttackingPaths	Algorithm 4. findOpponentAttack
<pre> Input: player P, Board B, position Pos output: all possible at- tPath from current Pos 1. if (!revAttack) then 2. if moveOnlyAttack 3. {from,to}=getAction 4. add attPath 5. end if 6. else if kilAttack then 7. {from,to, interme- diate}= getAction 8. attPath.value++ 9. add attPath 10. findAttPaths (oppo- nentPlayer, Board,to) 11. end if 12. return attPaths </pre>	<pre> input: player P, Board B output: total opp attack 1. attPaths = empty 2. for cp ∈ P 3. attackingPaths += cal- culateBestMove (P,B,pos) 4. for all attackingPaths 5. if attackingPath.value = WIN_VALUE then 6. if player=human then 7. return -WIN_VALUE 8. else 9. return WIN_VALUE 10. end if 11. if player=human then 12. return min(attPath) 14. else 15. return max(attPath) </pre>

These values depend on the utility function which started from the root node to its leaf nodes. For a single movement with occupying one opponent piece by computer, it adds +1 with its utility value and -1 for vice versa. But the movement without occupying any piece doesn't add any value. There are four different approaches to occupy opponent pieces. Here in 2A cell the black piece represents the computer player. It has two different approaches which produce four different movements to take over the control of the game.

Approach 1: Jumping over to cell 2C, it occupies one opponent piece. By this approach the utility function produces +1.

Approach 2: Jumping over to co-ordinate 4A then to 4C and again to 4E it occupies three opponent pieces. Then it doesn't find anything to occupy and thus it's stopped. But then the following step, the human player which is in 4E cell, can occupy 4E cell piece by jumping to 3E. So by this approach the utility function produces $+1+1+1-1=2$. So by selecting the first attacking path, second path, third path, fourth path which produces +1, second path +1, third path $-\text{WIN_VALUE}$, fourth path $-\text{WIN_VALUE}$ accordingly; whereas the traditional function produces $+1+1-1=1$ and $+1+1+1-1=2$ for third and fourth attacking path. If we would select the last attacking path that should be the lose condition.

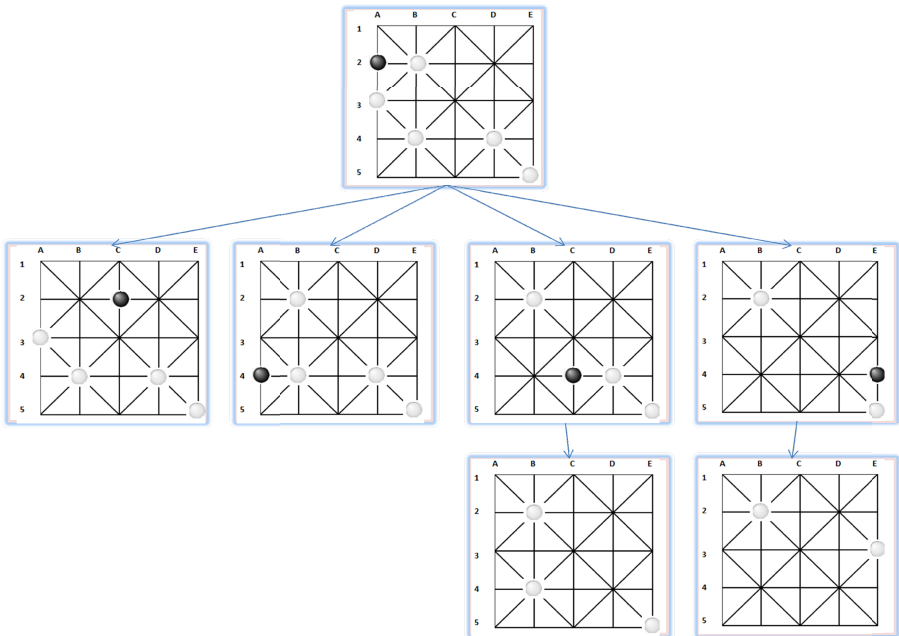


Fig. 1. Explaining a sample move for Computer turn

By typical utility function would consider the second approach as its value is greater than the first one. But it would be defeated by the opponent player by removing the only piece from cell from 5E to 3E. But we have considered this option in our proposed utility function that whenever it is going to lose by any movement that movement is never considered. In the traditional approach if we consider the following the figure.

4 Results and Discussion

All the experiments are performed in 5 x 5 cell standard environment. Our game has been played against 30 human and collected different data on different strategies. We have tested the algorithm in different depth level of the explored tree and noted the performance metric in terms of moves. Figure 2(a) describes the number of moves to win this game according to their depth level. As we notice that increasing the depth of the tree for a certain move condition, decreases the number steps to win the game.

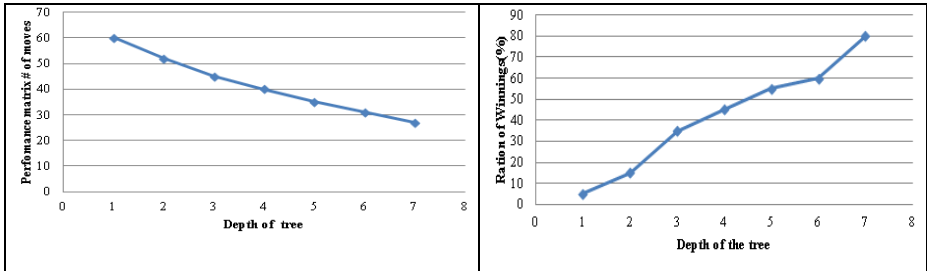


Fig. 2. (a). Avg num of moves VS Depth of the tree (b). Relation between winning rate (%) and depth of the tree

For a certain move condition if we analyze the data in recursive way that which move will predict good result among numerous ways, then it produces good result and output more winning ration. Figure 2(b) has depicts that by increasing the depth level of tree increases the probability of winning. We can also say alternatively that the ratio of winning percentage is proportional to the depth of tree headings.

Table 1. Comparison among difference procedures of playing game Alquerque

	1 Depth length	Normal Rec way	Our algorithm
matrix # of moves(avg)	60	45	27
Winning ratio (%)	20	70	90
Avg. time (in nano sec)	539105	1048285	832106
Completeness[find att.paths.]	No	Yes	Yes

Finally, we have compared among three different procedures in table 1. In our proposed algorithm we have already mentioned that we use minimax algorithm with pruning which outputs better performance than the other procedures. These are clear evidence that our algorithm tends to win with minimum number of moves with 27 whereas considering the current state evaluation needs 60 moves to win. Our proposed algorithm also pictures the winning ratio is quite high 90% than the other approaches. The average time requires for an every move are 539105, 1048285 and 832106 for first, second and our procedure accordingly. This time is recorded in an Intel Core™ i5 Processor and the simulation is done by Java. We can also estimate that the completeness of our algorithm is true.

Acknowledgement. This work was supported a grant from the NIPA (National IT Industry Promotion Agency) in 2013. (Global IT Talents Program).

References

1. Bell, R.C.: Board and Table Games from Many Civilizations, vol. 1, pp. 47–48. Dover Publications, New York City (1979) ISBN 978-0486238555
2. Russell, S., Norvig, P.: Artificial intelligence A Modern Approach, 3rd edn.
3. Zuhe, S., Neumaier, A., Eiermann, M.C.: Solving minimax problems by interval methods. BIT Neumerical Mathematics 30(4), 742–751 (1990)
4. Baker, A.B.: Intelligent backtracking on constraint satisfaction problems: experimental and theoretical result, Citeseer (1995)

Fractal Based Hardware Accelerated Technique for Graphical Rendering

Divya Udayan J.¹, HyungSeok Kim¹, Jun Lee², and Jee-In Kim²

¹ Department of Internet & Multimedia Engineering, Konkuk University, Republic of Korea

² Department of Advanced Technology Fusion, Konkuk University, Republic of Korea

Abstract. Rendering of natural scenes has been widely discussed by many researchers for a long time due to its numerous applications. But still the main challenge is complexity in geometry and memory unavailability in current hardware platforms. Natural scenes from real world contain a huge number of small details that are hard to model, take a lot of time to render and require a huge amount of memory. We address this problem by following the principle of self similarity or fractal geometry in the natural scenes. For evaluating the feasibility of fractal based image rendering in different dimensions, we have first considered the 2D structure, Mandelbrot set that has gained wide recognition both in mathematical and graphical domains because of its appeal and complex structure. In our work, we have examined the serial algorithm of this set and devised a parallel algorithm for the implementation on a massive parallel graphics processing unit (GPU) using the computer unified device architecture (CUDA) programming model. We have also extended our approach from 2D structure like Mandelbrot set to 3D real world example of terrain rendering. Performance is evaluated in terms of execution time and observed that a parallel implementation of the method on a GeForce GTS 450 GPU is on an average 2X times faster than its sequential implementation.

Keywords: Fractal, Mandelbrot set, Terrain rendering, Parallel computing.

1 Introduction

Recent technological possibilities of hardware platforms are replacing the manual work done in many modeling and rendering applications. Rendering of natural scenes is used in many applications including game content creation, cartography, video games, flight simulators, advertising etc. Nowadays, computer generated models are becoming more and more realistic with the improvement in computer hardware. The goal of our work is to create natural scenes that are very similar in appearance like trees, leaves, mountain range and many more. The main challenge is complexity in geometry and memory unavailability in current computers. Overcoming this complexity has been a challenging problem for many years. Therefore we follow the principle of self similarity [1] to model our natural fractal examples. This self-similarity is represented in the Fig. 1, using real world examples of tree and leaves. Fast rendering is another challenge that has to be solved due to the increasing demand for natural and

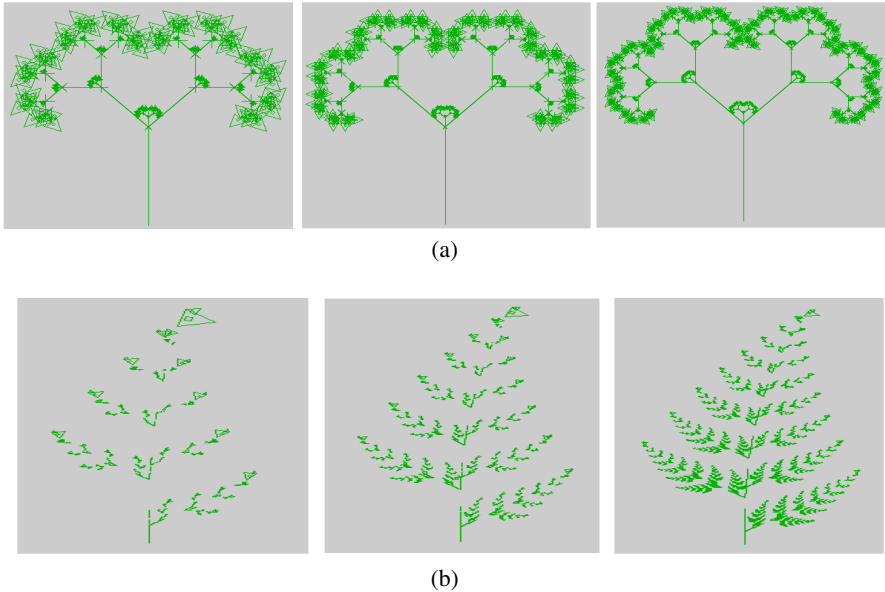


Fig. 1. Dynamic Modeling of natural scenes (a) Fractal based tree modeling (b) Fractal based leaf modeling

realistic rendering. Indeed, the targeted goal is to create images that are similar to what the viewer can see in real life with his/her eyes. The aim of this paper is to achieve visually convincing real time rendering of natural scenes using the principle of self similarity and to evaluate the feasibility of GPU for fractal based graphical rendering applications. In order to evaluate the GPU acceleration over CPU, we have considered the 2D structure, Mandelbrot set that has gained wide recognition both in mathematical and graphical domains because of its appeal and complex structure. We have also extended our work to 3D real world example of terrain rendering.

2 Related Works

In contrast to the human made objects, the natural objects are more irregularly shaped. Also, natural scenes exhibit self similarity across different spatial scales. Some researchers consider surface modeling and rendering as an optimization problem and suggested solutions through relaxation methods. For example, Grimson [2] suggested that given a set of scattered depth constraints, the surface that best fits the constraints passes through the known points exactly and minimizes the quadratic variation of the surface. He employed a gradient descent method to find such a surface. Extending this approach to use multi-resolution computation, Terzopoulos [3] proposed a method to minimize the discrete potential energy functional associated with the surface. In this

formulation, known depth and/or orientation constraints contribute as spring potential energy terms. Poggio et al. [4] reformulated these approaches in the context of regularization. Recently many image based modeling and rendering (IBMR) techniques are developed for terrain visualization [5]. The FlyAway system [6] uses a mipmap approach for a single, memory resident texture. The clipmap [7] provides a hardware supported method to handle extremely large textures. There are some shape grammar based methods for terrain rendering that uses context-sensitive rules[8]. CityEngine [9] is a well known commercial product based on shape grammar. Other terrain visualization approaches are limited with respect to the large texture data. Fast rendering is another challenge for fractal based image rendering. Issac K.Gang et al. [10] describe the parallel implementation and analysis of Mandelbrot set construction. They devised a parallel algorithm and implemented it using Message Passing Interface (MPI). In our work, we are focusing on terrain modeling using the self similarity principle from Mandelbrot set and devise a parallel algorithm on GPU using CUDA programming model.

3 Mandelbrot Set

The Mandelbrot set [11] is a fractal that has a very simple recursive definition (1):

$$M = \left\{ c \in C : Z_0 = c, Z_{n+1} = Z_n^2 + c, \sup_{n \in \mathbb{N}} |Z_n| < \infty \right\} \quad (1)$$

It is considered as a set of all complex numbers which do not tend to infinity when computing z_{n+1} . When computing the Mandelbrot set, one uses the fact that c belongs to M if and only if $|z_n| < 2$ for all $n \geq 0$. To demonstrate the potential for high-performance parallel computation on the GPU, we choose to implement the Mandelbrot set based on fractal rendering. First, the algorithm was implemented on CPU to verify the process flow using single-thread. Mandelbrot set rendering uses IFS (Iterated Function System), a method of constructing fractals; the resulting fractals are always self-similar. The typical process of generating a 2D IFS fractal involves selecting a two-input, two-output function at random and applying it to a point in the (X, Y) plane, plotting the result, and repeating the process. To represent the IFS function, we used the affine transformation, for representing 2D transformation. This transform can be represented by a transformation matrix (2):

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2)$$

where $X' = aX + bY + c$, $Y' = dX + eY + f$

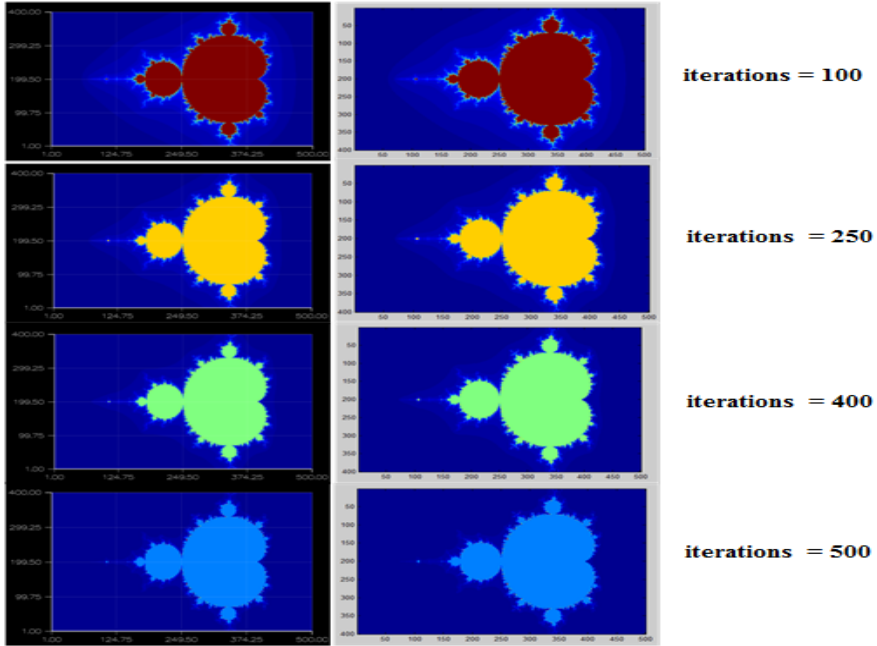


Fig. 2. Experimental results of Mandelbrot set color variations with increase in iterations. Left image shows the implementation in CPU and right image shows the implementation in GPU.

Finding the values of a, b, c, d, e and f , along with appropriate bounds on a window of X and Y is a time consuming process for CPU. We used the GPU implementation to reduce the time. The image is colored by the number of times a pixel is hit. Our aim is to draw the fractal by using two threads, one for each half of the drawing area. The method computes the fractal given a region in the complex plane. The implemented algorithm subdivides the complex plane into two regions and performs the classic Mandelbrot algorithm on each of them. Pixels are independent of each other and there is no writing to the same memory location. Color smoothing effect is necessary in order to provide an aesthetic feel to the rendered image. Coloring of exterior Mandelbrot set can be done using approaches based on continuous coloring, smooth color gradient or fractional iterations. In our method we used Normalized Iteration Count algorithm [12], which provides a smooth transition of colors between iterations. In the GPU implementation, there are dependencies between successive iterations of the algorithm. We used a strategy known as dependency relaxation to run as many iterations as possible concurrently with an insignificant accuracy loss. This strategy utilizes the computing power of parallel hardware more efficiently to achieve significant performance gain.

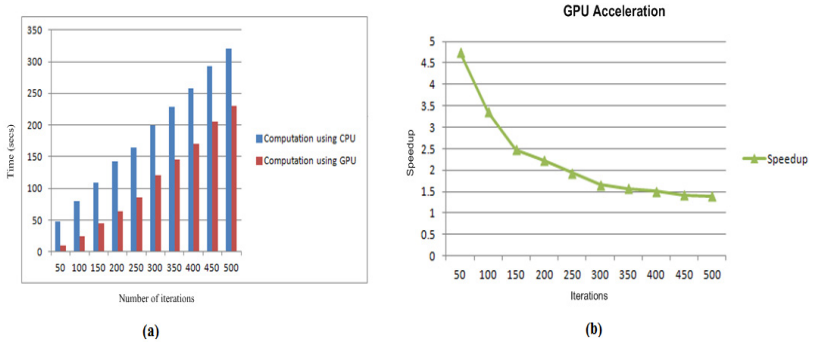


Fig. 3. (a) CPU vs GPU execution time with increase in number of iterations (b) GPU acceleration with increase in iterations

4 Terrain Rendering

The above study on Mandelbrot set can be extended to the creation of dynamic fractals. The whole universe is represented by dynamic topologically equivalent basic shapes of different scales. If the 2D structure is extended to 3D [13] then the expansion of the unifying interaction in two dense regions keeps them apart creating a less dense structure which is seen as void space between them. If we look closely at the Mandelbrot set, we can see that the tiny internal structures resemble the main Mandelbrot set. Therefore we follow the principle of self similarity to model our natural fractal examples. First, we generate a coarse, initial model. Then we recursively add additional random details that mimic the whole structure. The brief explanation of modeling mountain landscape is explained in this section. First we start with a grid of four points defining the corners. We then specify four random values that define heights at these points and scale the heights by a scaling factor ‘ s ’. Then we divide the grid into 6 triangles. We further divide each triangle into three new triangles. This defines three new points around the original triangle and also a point in the centre of it. Next we calculate the average of the heights of two neighboring corners and then add a random value. This random value is normally distributed and scaled by a factor ‘ S_n ’ related to the original scale factor ‘ S ’. We continue this procedure till we get a continuous fractal landscape which is given by equation (3) :

$$s_n = \left(\frac{1}{2}\right)^{H/2} s \quad (3)$$

where ‘ H ’ defines the smoothness of the landscape. This sequential approach can be applied to many application domains like 3D games, medical domain, film industry and 3D digital imaging. But the most challenging problem is fast rendering. We have proposed a parallel approach to generate terrain geometry in the background while

rendering takes place in the foreground. The solution is to use two threads. The main thread will render and update the terrain blocks and a worker thread will generate the geometry. But multithreading is very complex to synchronize efficiently. Solution to this is to use mutexes. Mutual exclusion objects are synchronization objects whose state is set to signaled when it is not owned by any thread and non signaled when it is owned. Thus mutexes can be used to schedule the threads. The thread which acquires the mutex will be executed whereas the other thread needs to wait until the execution is terminated. Our approach uses two mutexes, one for each thread. This can avoid the problem of infinite waiting due to mutex locking and unlocking in case of using only one mutex. So in our approach when the worker thread is generating the geometry, the main thread will be rendering and updating the terrain blocks, and no wait conditions would occur. Terrain texturing is implemented without a blend map by using multi-texturing blend operation as proposed by [14]. This approach provides reasonable blending for higher altitude scenes which lack finer level of detail due to the height of user viewpoint. However, there may be some level of inaccuracy due to the split calculations among physical processing units. But, if the scene rendered is aesthetically pleasing, then some inaccuracy in the operation may be acceptable.

5 Result Analysis

In the previous sections, we explained the implementation details of Mandelbrot set and its extension to real world terrain rendering. We will now briefly verify our results using some screenshots and images generated by both serial and parallel programs. Performance is evaluated in terms of execution time and overall GPU acceleration over CPU. The experimental results of Mandelbrot set are shown in Fig 2. The left side shows the implementation in CPU and right side shows the implementation in GPU. The terrain rendering application was evaluated based on the rendering time with increase in complexity of geometry. As the geometrical complexity increases, the rendering time also increases in sequential processing. The experimental results of GPU based parallel terrain rendering are shown in the Fig. 4. The performance shows that parallel computing for fractal based image rendering is 2X times faster than its sequential implementation. Fig. 5 shows the simulation of light effects on the terrain rendering at different time of the day and night. It should be noted that in fractal image rendering applications, the actual rendering is always done by the CPU and the GPU performs a secondary mechanism to speed up the computations involved in creating the geometry. The speedup and performance graph in Fig. 3 and Fig. 4 illustrate this fact graphically. After working with this problem, we are convinced that it is a good problem to parallelize. The main limiting factor for scaling this problem to larger problems is memory. For example, to generate larger size images, say 10,000 X 10,000 pixel image, over 500MB of memory is needed.

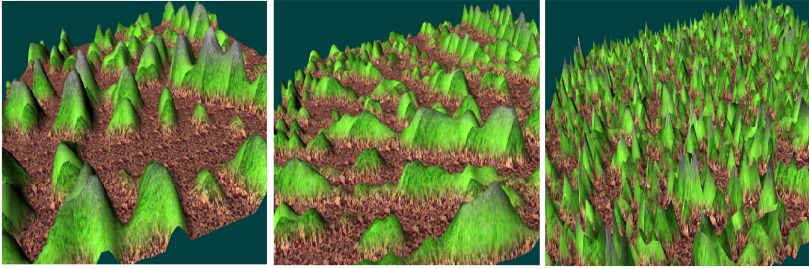


Fig. 4. Parallel implementation of fractal based terrain rendering with increase in complexity

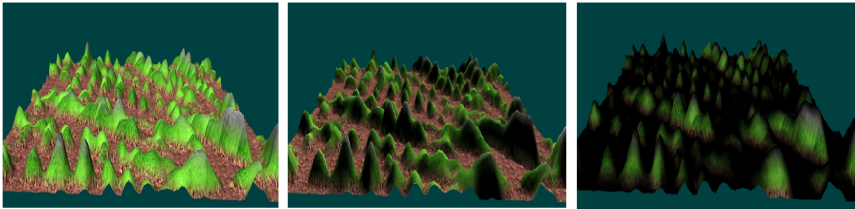


Fig. 5. Fractal based terrain rendering with varying lighting conditions

6 Conclusion

In this paper, we have discussed a hardware accelerated technique to increase the speed of computation in fractal based graphical rendering in both 2D and 3D dimensions. Our work can be extended to using fractals in many forms to create realistic fractal images of natural scenes like clouds, huge mountain ranges, coastlines and so on. This paper also opens new dimension to natural scene modeling, allowing us to mathematically define our environment with more accuracy than before. We have described the Mandelbrot set, which is a fractal and extended our approach to terrain rendering. We also discussed our implementation in both serial and parallel algorithms. We analyzed the performance and verified the results. As future work, we plan to consider this problem in a larger scale and improve our current implementation.

Acknowledgement. This research was partially supported by grant from Capture Korea Project funded by the Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2013 (50%). A part of this research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032182) (50%).

References

- [1] Barnsley, M.F.: *Fractals Everywhere*. Academic Press (1991)
- [2] Grimson, W.E.L.: *From Images to Surface: A Computational Study of the Human Early Vision System*. MIT Press, Cambridge (1981)
- [3] Terzopoulos, D.: *Computing Visible-Surface presentations*, Technical Report A.I. Memo No. 800, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts (1985)
- [4] Poggio, T., Torre, V., Koch, C.: *Computational vision and regularization theory*. *Nature* 317(6035), 314–319 (1985)
- [5] Edward Swan, J., Arango, J., Nakshatrala, B.K.: *Interactive, Distributed, Hardware-Accelerated LOD-Sprite Terrain Rendering with Stable Frame Rates*. In: *Proc. of SPIE*, vol. 4665, pp. 177–188 (2002)
- [6] Hüttner, T., Strasser, W.: *FlyAway: a 3D terrain visualization system using multiresolution principles*. *Computers & Graphics* 23, 479–485 (1999)
- [7] Losasso, F., Hoppe, H.: *Geometry Clipmaps: Terrain Rendering Using Nested Regular Grids*. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004)* 23(3), 769–776 (2004)
- [8] Muller, P., Wonka, P., Haegler, S., Ulmer, A., Gool, L.V.: *Procedural Modeling of Buildings*. In: *SIGGRAPH 2006: Proceedings of the 33rd Annual Conference on Computer Graphics and Interactive Techniques*, pp. 614–623 (2006)
- [9] Procedural, inc., *City Engine* (2009) <http://www.procedural.com>
- [10] Gang, I., Dobson, D., Gourd, J., Ali, D.: *Parallel Implementation and Analysis of Mandelbrot Set Construction*. In: *Proceedings of International Conference on Industry, Engineering and Management Systems* (2007)
- [11] Mandelbrot, B.B.: *Fractals, an art for the sake of art*. *Leonardo* 1989, vol. 7(suppl.), pp. 21–24 (1989)
- [12] Barrallo, J., Jones, D.M.: *Coloring Algorithms for Dynamical Systems in the Complex Plane*. *Visual Mathematics* (1999)
- [13] Savov, E.P.: *Dynamic Fractal Unifying Interaction Confirmed with Magneto-spheric Behavior and Orbital Data*. *Journal Complexity* 12(3), 61–76 (2007)
- [14] Luna, F.: *Introduction to 3D Game Programming with DirectX 9.0c: A Shader Approach*. Wordware Publishing Inc. (2010)

Topics on Public Game Art Using Media Façade

Hyoyoung Kim and Jin Wan Park

GSAIM, Chung-Ang University, Seoul, Rep. of Korea
the.kimyo@gmail.com, jinpark@cau.ac.kr

Abstract. Today, we can easily find game arts on media facade using a huge screen. There are many issues media artists should consider, such as publicity, theory of games, characteristics of the audience, spatiality, and etc. We will examine the impact of media facade work on the convergence of games and game characteristics with large-screen public art. This research and review will contribute for artists to make a large screen based media facade game that attract public efficiently.

Keywords: media façade, big screen game, public game, experimental game.

1 Introduction

Media façade created by digital LED walls or projection mapping process is an effective medium for delivering artist's ideas to general viewers. Many media artists like to use media façade since it is such a useful tool to deliver messages to the public. Like other installation pieces which aim to facilitate interactions, various contents on huge screens tend to adopt games in order to pull users into the content. However, it is not easy for artists who do not fully understand games to create content that leverages advantages of a supersized screen. This paper explores various topics on games and media façade in which a lot of media artists are interested. Also it reviews detailed information that could improve the quality of artwork. Moreover, this study aims to consider phenomenon that occurred where games and media façade meet.

2 Public Art and Interactive Games

In public art and games focusing on interaction realm, data on the media façade technology and games using media façade are abundant. The data could be reviewed to study origins of big screen games of today.

2.1 Public Art

Public art is a domain to categorize all installation works located in places everyone can unlimitedly reach, no matter which medium was chosen for the work. In other words, the most important feature of public art that make it different is its site: a site where the public can easily reach. The main factor to distinguish the site is its visitors,

and it's the visitors to characterize the place. Therefore, through understanding of the site and its visitors, user experience diversity could be predicted.

Public art often features artworks by artist groups or collaborations in attempt to lure more visitors. Art installations for specific events which were set temporarily are also common. These temporary art installations are divided into two: installations for authorized events and installations unauthorized events. Graffiti artists, for instance, usually keep a record of their works because graffiti is publicly unaccepted in many cultures so that graffiti artworks are often one-off. With the advance of the Internet, there is a new tendency to archive records of temporary art installations on websites.

Another characteristic of public art is that it needs cooperation of related district authorities. Of course there are some exceptions such as guerrilla graffiti art that inherently challenges authorities or performances such as flash mob that appear for a relatively short time. However, in general, public art mostly follow government rules or are affected by the government encouragement policies. For example, there is 'Percent for Art' act. This program requires that funding for public art be part of a construction budget.

Media façade is an art form using a medium that was unexpected in traditional public art. Traditional public art used to be limited to wall paintings or installations along with architecture. However, as media have diversified, these traditional limitations have faced radical changes. The fact that Korea changed the legal term 'art decoration' to 'artwork' in the Culture and Arts Promotion Act reflects this new flow of change.

Media façade is naturally labeled as public art since it uses architectural surfaces as its canvas. Therefore, media façade has to follow rules of public art, predict public experiences, and consider its value as a landmark. Especially LED architectural lighting designs such as Seoul Square in Seoul, Port authority in New York, and Graz Art Museum in Austria (Fig. 1) need to consider their public art aspects from the their architectural design stage.

Above all, what makes public art different from other artworks in museums is its viewers. Public art is open to the general public whereas museum visitors voluntarily



Fig. 1. Kunsthhaus Graz

come to museums intending to appreciate artworks. In this sense, media façade could be a very good example of public art. Audience of public art is the general public and they have lower motivations compared to handful of museum visitors with higher motivations. Therefore, some aspects of media façade overlap with those of commercially motivated popular arts.

2.2 Interactive art and Game Art

When it comes to games, it is important to consider aspects such as sports competition and game immersion, but direct interactions are seldom considered seriously. In fact, interactive art and games do not necessarily need computers. The reason we think of computers when we talk about interactive art and games is that the interaction computers could provide is much more immense than ever before. Interaction, which is considered as one of the most essential aspects of games these days, is in fact has been on the rise only after the advance of the computer technology.

Games in interactive art also reflect characteristics of computer games. It has not only unique characteristics of game including rule-following, competition in constrained environments, and victory of defeat, but characteristics of computers such as automated interaction. Attempts to apply game formats including game art or video game art to public art realm have resulted in alternative reality games using actual city spaces as game stages. *Urban Vibe: The Art of Playing City* by Art Center Nabi, Seoul could be a good example. Game aspects in these works do not appear clearly. They focus more on sharing group experiences rather than competition or outcomes.

What media art and game have in common is ‘interaction’. Interaction is a big part in both of them. Media art and games have definite similarities but it is still not easy to combine them properly and create satisfactory outcomes. Moreover, when it comes to the games that use large screens such as media façade, there are more to consider since they limit subjects of interaction to a fraction of their audience.

2.3 Experiments on Games

Convergence of art and game is not only for the artists who try to extend their territory and seek creative changes. The fact that ‘The Night Journey’, which is created based on video artist Bill Viola’s work, took the prize at the Independent Game Festival, and *Hush* won the Better Game Contest suggests, un-commercial and subjective game design of independent experimental games help games to expand their realms. However, user reactions to such games are quite different from experts’ opinion. Like critiques to ‘The Night Journey’ show, users do not approve of experimental games while experts think highly of their challenge. Un-commercial games do exist, but the question here is that is it appropriate to call them ‘games’ when most users do not empathize with them? Experimental games seem to sacrifice the essence of game itself to be in the realm of art.

Games using media façade look similar to experimental games in a sense. However, despite appearances to the contrary, big screen projects which have been implemented so far have excluded experimentation, and focused on classic game content.

That is because the big screen of media façade has its advantages and disadvantages: games are exposed to the general public so it has an extreme learning curve. As a result, the convergence of game and media façade causes suppression of an experimental mind, which is different than expected.

3 Issues of Media Façade Game

3.1 Goal Setting Issue

Academic and business views of the definition of game differ from one another since they have different angles for game categorization. For example, Johan Huizinga, as a historian and cultural theorist, defines game as something that has a significant function with a cultural quality. On the contrary, game of today is interpreted in a narrow sense due to the market that emphasizes computer games. Various games including board games and sports games exist in the market but it is difficult to find a definition to include all of them (Table 1).

Table 1. Varied definitions of game

Definition
“form of art in which participants, termed players make decisions to manage resources through game tokens in the pursuit of a goal” (by Greg Costikyan [13])
“context with rules among adversaries trying to win objectives” (by Clark C. Abt [14])
“exercise of voluntary control systems, where contest between powers, confined by rules to produce a disequilibrium outcome” (by Elliot Avedon & Brian Sutton-Smith [15])
“system that players engage in artificial conflict defined by rules that results in quantifiable outcome” (by Katie Salen & Eric Zimmerman [16])

Nevertheless, various expert opinions on the definition of game have essential aspects in common. First, games have to have an objective (goal, win, competition). Second, games have to have a rule (resource). Media façade games also have their goals. Setting a rule and forcing a wanted behavior are essential aspects of game whether they are for breaking user’s own record or for competing with other players. However, media façade games often do not have clear objectives, so it is sometimes hard to distinguish play from game. Not only media façade games but other media artworks adopting game aspects in an art-centric way show a similar tendency.

This ironic phenomenon has arisen since game designers’ definition and objective could be different from actual game players’ goal and expectation. It is the artists, not commercial game producers, who use media façade as a tool of expressing their art philosophy. From the outside, it seems like people enjoying and playing interaction

are the subject of the work of media art. However, to media artists, participants are more like just materials for the work rather than designer's fundamental goal. Artist-centered view like this could boil down to loss of game play by using games only as materials for art, which might disturb general user's game immersion.

3.2 Ambiguous Competition Issue

Artists who see users as materials for the performance do not think it is necessary to design a setting for competition. However, settings which are designed to facilitate competition are obviously necessary. Media artists do not consider these competition settings including the best score or competition among 3 or more people very seriously. Deciding winners and losers is crucial to encourage user immersion. Nevertheless there are a few problems for media art following the rules of ordinary games. First, there is no clear distinction between win or lose in public art which is designed for playing itself. Second, random users from the general public do not necessarily have the strong will for winning. Third, resources for example playing time are limited .

3.3 Game Watching Issue

One of the most distinctive characteristics of media façade is that only a fraction of its viewers experience interaction. It is like watching a sports game. Viewers do not actually play the game, but their flow level of sports game is high. Mass media such as televisions broadcast the games to the public, and viewers identify themselves with the players of the team they are cheering. This game watching is another form of classic one-way entertainments such as a play except the competition and rules and objectives. This form of game watching has been common starting from arcade games such as Gallaga (Namco, 1981) and Xevious (Namco, 1982) to Virtual Fighter (SEGA, 1993). It has evolved into enthusiastic cheering via broadcasting for Starcraft (BLizzard, 1998). This game watching tends to fall into the category of the traditional artwork-viewer relationship that media art has pursued rather than user's direct participation in game competition. Therefore, games are designed for viewers rather than for actual participants.

4 Conclusion

In general, even multi-user games are played on personalized screens. In other words, when users play general games, they face their own virtual world. However, media façade uses only one big screen that everyone could watch, which means non-participants are also forced to partially experience the game. Media façade in big cities especially need to be more cautious because forcing the public to put through unwanted experiences is like pollution to them. Majority of public could plunge into confusion when they unwillingly face so called game aspects such as reward or competition.

Game has various qualities hard to categorize according to one point of view. Media façade might not be the best medium to provide users ordinary game experiences such as challenge and purification while users solve problems and follow rules. Many artists who do not fully understand the advantage of game have proposed to restructure simple old games such as Pong and Break Out. Also, game designs efficiently using media façade and fully taking advantage of this efficient medium are hard to find. Previous game designs or media art designs have not considered various viewers, big screen, city environment, and interaction issues. However, careful consideration should be given to the issues above in order to offer interesting experiences to more viewers via big screen public games.

References

1. <http://www.publicartinpublicplaces.info/>
2. http://www.nabi.or.kr/project/.past_read.nab?idx=76
3. <http://www.thenightjourney.com/statement.htm>
4. <http://www.igf.com/>
5. <http://www.gamesforchange.org/play/hush/>
6. <http://www.bettergamecontest.org/>
7. Ludens, H.: *A study of the play element in culture*, Johan Huizinga. Beacon Press (1955)
8. Costikyan, G.: *I Have No Words & I Must Design* (1994)
9. Abt, C.C.: *Serious Games*, p. 6. Viking Press (1970)
10. Avedon, E., Sutton-Smith, B.: *The Study of Games*, p. 405. J. Wiley (1971)
11. Salen, K., Zimmerman, E.: *Rules of Play: Game Design Fundamentals*. MIT Press (2003)
12. <http://www.youtube.com/watch?v=rYAdtkZe90k>

A Physical Interactive Game for Learning English of Children

Sang-I Shin and Kyoungju Park

Dept. of Image, Graduate School of Advanced Imaging Science,
Multimedia and Film, Chung-Ang Univ.
sang.i.shin@gmail.com, kjpark@cau.ac.kr

Abstract. As English is the universal language used by international society with immense interest, we ought to seek for more effective teaching method. One way to reach such goal is to utilize a physical interactive game in teaching English. A physical interactive game encourages the spontaneous activity in learning without the psychological burden while promoting interest in English and basic usage of English. Also, physical movement interface can stimulate learners' motivation while inducing positive response in interest and immersion.

In this thesis, we will design interface using 3D depth camera to produce a physical interactive game for learning English of children by capturing the movement of children and creating a gesture that will trigger children's interest. Also, we will create a physical interactive game environment high in user interaction with the content and apply such method to introduce prototype of English learning game.

Keywords: Physical Interactive Game, Serious Game, Game Design.

1 Introduction

As international society grows, English became the universal language. Along with globalization, English gained lot of attention and influenced ethnicity, generation and even cultural aspects. The importance of English education became high resulting in learning age to be even lower and therefore, more effective English learning method at young age is in demand.[1]

A physical interactive game environment can offer enjoyable, interesting learning aspect as well as motivation to children learning English.[2] Children will use physical movement not only as the spontaneous activity in learning English but also as getting instantaneous feedback from the system while allowing subconscious learning afterwards. Therefore, such method of learning English can increase the learning efficiency.

In recent years, there were numerous researches on motion recognition such as using motion sensor to control television and game consoles but there have been only few researches that were actually related to games for education purpose.[3] Hence, more researches on utilizing physical activity data in modeling educational

environment and physical activity games focused on children who have relatively higher exercise level compared to adults.[4] These researches can promote user participation and produce prototype for sensible educational games that are interesting for children.

In this paper, we will use 3D depth camera to produce a physical interactive game for learning English of children prototype allowing children to have movement freedom in virtual game environment while establishing interest and fun gesture in games. Also, we will design interface contemplating children's thought process and movement and create a physical interactive game environment high in user interaction with the content to increase learning efficiency.

2 Related Work

Guan-Feng He(2012) suggested motion sensor based alzheimer's prevention game design establishing specific gesture for alzheimer's preventing calculation game and movement of character in memory games.[5] Suggested movements replace up, down, left, and right keys on keyboard. However, such gestures are complicated to learn naturally considering the age of users.

Connssynn Chye(2012) designed motion sensor based martial arts game for beginners to correct and teach posture allowing users to imitate proper posture.[6] It extracts specific points from geometrical data from the motion sensor to recognize joint points and calculate certain angle to again recognize joint points further ranging specific points to learn users movement by multiple feedbacks from above process.

3 The Proposed Method

3.1 Application of Natural Interface Design in Games

Serious game meant for children's physical interaction needs to be suitable for children and has to use gestures in game that can be easily recognized by 3D depth camera. Hence, such gestures have to be normal and familiar to children to promote children's participation and to prevent errors in motion recognition resulting in smooth play. According to the research done by Kyung-Ok Lee(2012) and Hye-min Won(2011), children had no problems in imitating simple gestures such as raising arm, moving to the sides and jumping. However, children had difficulties in imitating gestures such as spreading arms while jumping, sitting down and standing up, swimming, performing and flying and those movements had low recognition by motion sensor as well.[4][7]

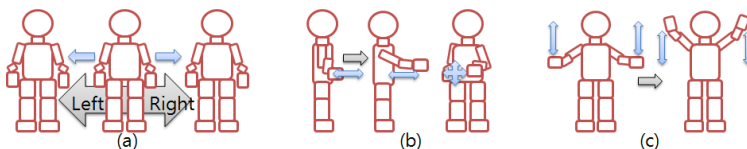


Fig. 1. Children's gesture applied in serious game (a)Body (b)Hand (c)Arm

Therefore, we suggest gestures that uses whole body and easy to imitate which can be obtained by using upper-body movement as well as hands and feet movement. Figure 1 below shows example of such movement: moving to the right and left, moving hands to the right, left, up, down, front and back and raising arm overhead and down. Such movements are easy to imitate and can be easily recognized. It does not limit to specific movements and allows creative movement by using whole body to increase interest level.[8]

To apply fixed movements, we use 3D depth camera to get user movement image and depth data input. Then, we use api(application program interface) to capture users body and skeleton information and further use such information to establish fixed gesture from gesture library.

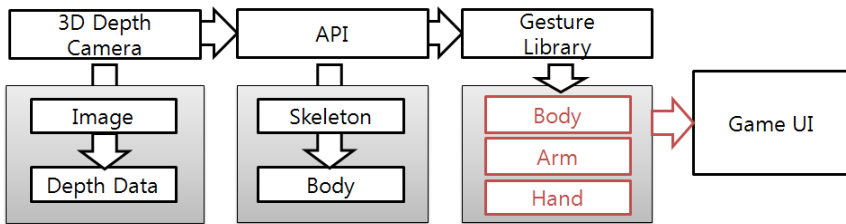


Fig. 2. Applying fixed movements process to Game User Interface

Established gestures in games are used in virtual game environment after setting position value process. Body gesture is calculated by change of x and y values in torso determining right and left distance differences and apply distance differences to user character to adjust the location of the user. Arm gesture uses location value of arm and head to launch a missile when location value of arm is equal to the y value of the head center. Lastly, hand gesture captures location of hand and calculates x, y and z value to use menu function.

3.2 Visual Interface Design

Visual interface includes everything related to designing user friendly virtual expression while increasing interest level in user interface. This research focuses on two specific interfaces which are menu construction method and game characters.

Game menu is simply composed of specific content of START, INFO and SCORE. Such menu construction method can be clearly perceived by children.[9] After choosing specific content from the menu, the picture of specific content changes to inform user that he/she has chosen such content from the main menu. Such exploration step is known to be liked by children.[10] By exploring simple content of the menu, children cognize the game and satisfy their curiosity while immersing to the game.



Fig. 3. Personified of Characters

Characters in game are not expressed as simple object but personified matter. According to the Piaget’s theory of intellectual development, children between the ages of 7 and 9 think human beings have created the nature and believe that every object has an emotion and life.[11] Because user character literally reflects users’ actual movement, character needs to apply children’s thought process as well as forming a strong bond with children.

3.3 Interaction Design

Interaction means expression response as a result of an action to children. [12] User utilizes movement to move character in game and confirms how ai character vanishes in attempting an attack realizing the result of his/her movement in game. the level of interaction increases even if user made wrong movement.[13] If user character failed to avoid ai character in game, user character will lose life level in game. Therefore, user will actively participate in game to avoid performing improper movement. Such physical interactive system effectively increases learning immersion and satisfaction.[14]

User can see his/her movement and result of specific movement in game. User can identify real time feedback of movement by moving depth Image content to below the game screen in starting the game. User can enjoy virtual reality by actually playing the game as well as identifying Depth Image content.



Fig. 4. Reflection of user’s movement in physical interaction system

4 Result

Prototype game is about attacking ai character which contains specific words related to user charter’s main theme. We used tree as main theme relating apple, leaf and wood as specific words related to the theme referencing children’s English dictionary CHANT CHANT.[15] Prototype game was manufactured for children’s ages between 7 and 9. we used a computer equipped with Intel i3 CPU 2.93Ghz, RAM4.00GB, Nvidia GT240 graphics hardware. Resolution is 1024x768. 3D Depth Camera, Kinect and application program interface by openni are expressed in java.



Fig. 5. Start page and play screen of prototype game

	Solar Scramble	Prototype Game
Genre	Board	Shooting
UI	Touch	Gesture
Interface	Repetitive type	Open type
Game Method	Creativity required	Creativity and originality required
Interaction	Medium	High

Fig. 6. Comparison between prototype game and solar scramble

Solar scramble suggested by Ashley R. Kelly(2010) is serious game meant for children’s education method utilizing communications multi touch table display.[16] Game is about recognizing start screen which consists of the orbits of planets and further correcting planets in wrong orbital and names of planets on screen. solar scramble requires creativity in game play interface but lacks in originality in terms of UI which is repetitive in solar scramble.[17] Prototype game on the other hand requires change of movements by physical interaction interface lacks repetitive process while high in creativity and originality interaction compared to repetitive interface game solar scramble.

5 Discussion

Development of interface reflecting children’s movement and thought process over cramming process is necessary for children because they are at critical age where

body balance and motor skill grow. It is possible to promote children's interest and further constructing learning game environment by physical interaction method.

Our research implemented a method to use educational game for fun while fulfilling educational purpose during children's developing period by using physical interaction. We anticipate our research benefits further UI contents development based on physical interaction focusing on children.

References

1. Lee, J.E., Choi, S.-Y.: The effects of using multimedia title on preschool children's English listening skills, vocabulary, story recall ability and affective domains. *Multimedia-Assisted Language Learning* 13(3), 237–252 (2010)
2. Cha, E.-M., Lee, K.-M., Lee, J.-W.: Developing Virtual Learning Environments for Improving Spatial Sense of Young Children. *The Korea Contents Association Journal* 7(6), 154–160 (2007)
3. Lee, H.J.: Change to Useful Game: Trends of Serious Game. *Electronics and Telecommunications Trends* 135(27-3), 43–50 (2012)
4. Won, H.-M., Lee, K.-M.: Interactive Game Designed for Early Child using Multimedia Interface: Physical Activities. *The Korea Contents Association Journal* 11(3), 116–127 (2011)
5. He, G.-F., Park, J.-W., Kang, S.-K., Jung, S.-T.: Development of Gesture Recognition-Based Serious Games. In: *IEEE-EMBS International Conference on Biomedical and Health Informatics*, pp. 922 – 925 (2012)
6. Chye, C., Nakajima, T.: Game based approach to learn Martial Arts for beginners. In: *2012 IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, pp. 482–485 (2012)
7. Lee, K.-O., Lee, K.-M., Lee, S.-H.: The Content Development of Interactive Children's Story Based on Movement Activity of Children. *The Korea Society for Children's Media* 11(2), 133–154 (2012)
8. Park, J.W., Song, D.H., Lee, C.W.: Interface of Interactive Contents using Vision-based Body Gesture Recognition. *Smart Media Journal* 1(2), 40–46 (2012)
9. Gilutz, S., Nielsen, J.: Usability of Websites for Children: 70 design guidelines based on usability studies with kids. Nielsen Norman Group (2002)
10. Dempsey, J.: Since Malone's Theory of Intrinsically Motivating Instruction: What's the Score in the Gaming Literature? *Journal of Educational Technology Systems* (1994)
11. Ginsburg, H.P., Oppen, S., Brandt, S.O.: Piaget's theory of intellectual development. Prentice Hall (1987)
12. Keegan, M., Olson, D.O.: Scenario educational software: Design and development of discovery learning. *Performance Improvement* (1998)
13. Kim, H.-R., Chang, H.-J., Park, S.-H.: Designing Tangible User Interfaces of Physical Interactive Game. In: *KIISE* (2002)
14. Choi, S.-H., Kim, K.-S., Yoon, S.-J.: A Case Study on Applications of Physical Interactive Systems in On-Line Games. *Korea Game Society* 4(2), 21–28 (2004)
15. I&book, English dictionary CHANT CHANT, I&book (2008)
16. Kelly, A.R., Wallace, J.R., Cerar, K., Randall, N., McClelland, P., Seto, A.M.: Solar scramble: an educational children's game for collaborative multi-touch digital tabletops. In: *SIGDOC 2010 28th ACM International Conference on Design of Communication*, pp. 27–32 (2010)
17. Aldrich, C.: Simulations and the future of Learning. Pfeiffer (2004)

Using Multiple Verifiers to Detect Sybils in a Social Network Graph

Kyungbaek Kim

Department of Electronics and Computer Engineering
Chonnam National University, Gwangju, South Korea
kyungbaekkim@jnu.ac.kr

Abstract. Detecting Sybil identities is important to operate a distributed system without losing its openness property. Recently, OSN(Online Social Network) based Sybil detection methods are proposed and an individual node can determine whether other nodes are Sybil or not. However, since the probabilistic properties of the previous methods, single verifier based Sybil detection may suffer from the wide variance in the performance of detecting Sybil nodes. In this paper, multi-verifier based Sybil detection method is proposed to mitigate the variance. The proposed method selects honest verifiers from a social network graph where honest and Sybil nodes are mixed. Then, the method determines whether a node is Sybil or not by comparing the likelihood of acceptance of the node and a given threshold. Through the extensive evaluation with the real-world social network sample graph, the proposed multi-verifier based Sybil detection outperforms the single verifier based Sybil detection in both aspects of accepting honest nodes and suppressing Sybil nodes.

Keywords: Online Social Network, Sybil Detection, Trustness.

1 Introduction

In a distributed system, detecting Sybil identities is an important issue. Sybil identities are fake identities belong to a malicious identity and exploited to obtain immoral gains from the system or subvert the system. In a P2P system, many Sybil identities join the system and they can gain the control of the system to hamper the operations of the system [1]. These Sybil identities also threat the openness of distributed systems by making the resources of the systems untrustworthy. Another example can be observed in a collaborative recommendation system. Many Sybil identities recommend the fake assertion of a malicious identity, and let other identities trust the fake assertion [2,3]. These malicious activities conducted by Sybil identities are called *Sybil attacks*.

A traditional way to defend the Sybil attack is increasing the complexity of generating identities such as CAPTCHA [4]. Another way is using a centralized authority which requires real-world identities such as social security numbers or credit card numbers. However, these approaches require expensive costs and cause another threats such as leaking the critical information of real identities to malicious identities.

Recently, OSN based Sybil detection methods have been proposed [5,6]. They use an online social network graph where the real world relationships are embedded. It is assumed that the online social network graph is composed of an honest region where honest identities reside and a Sybil region where Sybil identities reside, and these methods rely on the property that there are a limited number of cuts between an honest region and Sybil region. According to this, these methods let a single node in a graph determine whether a node is Sybil or not by using a probabilistic measure such random walks.

Generally, an individual node well determines whether a node is Sybil or not. However, since the previous methods use a probabilistic measure and the detecting process is conducted on each individual node, some nodes exhibit the misbehavior such that a node may consider many honest nodes as Sybil nodes or it may accidentally accept many Sybil nodes as honest nodes. That is, there is wide variance in the performance of detecting Sybil nodes by using a single verifier. Also, this wide variation of the performance causes a new kind of threat such as focused Sybil attack to a targeted honest node.

In this paper, the Sybil detection method using multiple verifiers is proposed to eliminate this variance in the performance of detecting Sybil nodes. While previous methods focused on that each individual node works as a single verifier, the proposed method uses multiple nodes as multiple verifiers which are collaborated to determine which node is Sybil or not. In a social network graph, multiple verifiers are selected and each verifier conducts an OSN based Sybil detection method, SybilLimit [5] to determine which node is Sybil or not. Based on the results of verifiers, each node in the social network graph obtains a value of likelihood that a node is accepted by a verifier. That is, the likelihood value of a node represents the likelihood that the node is an honest (non-Sybil) node. This likelihood can be normalized into a value between 0 and 1, and it can be used for determining whether a node is Sybil or not.

In the proposed approach, there are two challenges: 1) how to choose verifiers and 2) how to set a threshold to determine whether a node is honest or not. To choose the good verifiers, a few pre-trust nodes are used to gather a set of candidate nodes and verifier nodes are randomly selected from the set. This selection method is required because honest and Sybil nodes are mixed in a social network graph, and this careful verifier selection prevents a Sybil node from being a verifier node. To find out a reasonable threshold, the extensive evaluation with a sampled real-world social network graph is conducted. According to the results of the evaluation, it is shown that using multiple verifiers with a reasonable threshold accepts most of honest nodes and prevents most of Sybil nodes from being accepted. Also, through the evaluation, it is shown that the importance of choosing good verifiers to guarantee the performance of detecting Sybil node by using multiple verifiers.

2 Multi-verifier Based Sybil Detection

2.1 Background and Assumption

A social network graph, $G = (V, E)$, where $|V| = N, V = \{v_1, v_2, \dots, v_n\}$ and $|E| = M, e_{ij} \in E = v_i \rightarrow v_j$, can be viewed as a single strongly connected component. Each v_i is a node corresponding to an identity. If a node is corresponding to an

honest identity, it is called as an honest node. Otherwise, the node is called as a Sybil node. In a social network graph, an honest (non-Sybil) region where honest nodes reside coexists with multiple Sybil regions where Sybil nodes reside. Inside a Sybil region, Sybil nodes are easily generated and each of them can be connected to each other as many as possible. But, there are the limited number of attack edges between the honest region and each Sybil regions [5,8].

SybilLimit[5] is an OSN-based Sybil detection method, and it is used to determine whether a node is honest or not. Basically, SybilLimit is used for a single verifier node to determine whether a suspect node is a Sybil node or not. The verifier node, v , prepares the verification set of tails, S_v , which is composed of $r (= \theta(\log|V|))$ tails drawn from random routes of length $w (= \theta(\sqrt{|E|}))$. The suspect node, s , also prepares the sample set of tails, S_s , which is composed of r tails drawn from random routes of length w . If there is any common tail in both of S_v and S_s , the verifier node accepts the suspect as honest nodes.

However, the performance of SybilLimit may be different from each verifier node since the process of preparing the verification set and the sample set relies on a probabilistic measure. According to this reason, some approaches have tried to use multiple verifiers and assign the Sybil-resistant trust value which indicates the likelihood that a node is honest [3,7]. But, these works still did not consider the challenges such as how to choose good verifiers and how to set a threshold to accept a node as honest.

2.2 Algorithm of Using Multiple Verifiers

The main idea of using multiple verifiers is gathering the results of detecting Sybil nodes from multiple verifiers and determining whether a node is Sybil or not based on the gathered result such as how many verifiers accept the node. To do this, a system needs a coordinator to gather/examine the results. Usually the possible coordinator can be the OSN provider which knows the entire social network graph. Also, this coordinator has a few number of trust nodes which are considered as known honest nodes.

The basic algorithm of using multiple verifiers to detect Sybil nodes is shown in Fig. 1. Algorithm 1 shows the algorithm of choosing good verifiers. At first, we use a known honest node to collect the nodes which may be honest nodes as the set of candidate verifier nodes, $S_{candidate}$. Since both honest nodes and Sybil nodes are mixed in V , during this step we need to add the nodes which is verified by the known honest node as the candidate verifier nodes. Then, l verifier nodes are randomly selected from $S_{candidate}$. According to this algorithm, we can choose honest nodes as verifier nodes to some extent and prevent that many Sybil nodes are selected as verifier nodes.

After the good verifiers are chosen by conducting Algorithm 1, multi-verifier based Sybil detection is performed like Algorithm 2. Firstly, each verifier node conducts SybilLimit based Sybil detection for a node and the likelihood value (t_i) of the node is calculated by dividing the number of verifier nodes accepting the node by the total number of verifier nodes. Then, if the likelihood value (t_i) of the node is greater than a given threshold value (T_{thres}) the node is considered as an honest node, that is, h_i becomes true.

Algorithm 1 : multiple verifier selection algorithm

Require: $G(V,E)$: a social network graph
Require: v_p : a known honest node
Require: l : number of verifiers
 $S_{candidate} \leftarrow \emptyset$
 $S_{verifiers} \leftarrow \emptyset$
 $S_{v_p} = \text{obtainVerifierSet}(v_p)$
for each v_i **in** V **do**
 $S_{v_i} = \text{obtainSampleSet}(v_i)$
 if $((S_{v_i} \cap S_{v_p}) \neq \emptyset)$ **then**
 $S_{candidate}.\text{add}(v_i)$
while $(|S_{verifiers}| < l)$ **do**
 $S_{verifiers}.\text{add}(\text{pickRandomNode}())$

Algorithm 2 : Multi-verifier based Sybil Detection Algorithm

Require: $G(V,E)$: a social network graph
Require: $S_{verifiers}$: a set of verifier nodes
Require: T_{thres} : a threshold to be accepted
for each v_i **in** V **do**
 $\text{accept} \leftarrow 0, h_i \leftarrow \text{false}$
 $S_{v_i} = \text{obtainSampleSet}(v_i)$
 for each p_i **in** $S_{verifiers}$ **do**
 if $((S_{v_i} \cap S_{p_i}) \neq \emptyset)$ **then**
 $\text{accept} \leftarrow \text{accept} + 1$
 $t_i \leftarrow \text{accept} / |S_{verifiers}|$
 if $t_i > T_{thres}$ **then**
 $h_i \leftarrow \text{true}$

Fig. 1. Algorithms of multi-verifier based Sybil detection

3 Evaluation

To understand the limitation of single verifier based Sybil detection and evaluate the proposed multi-verifier based Sybil detection, we conducted both methods with a sample social network graph obtained from Facebook [2]. The sample graph has 50k nodes and 905,004 edges, and it is considered as an honest region. Sybil regions are generated artificially. There are 25 Sybil regions and each Sybil region has 100 Sybil nodes. A Sybil region is generated as a single strongly connected component where the average number of edges is 15, and it has 2 attack edges which are connected to honest nodes randomly.

Fig. 2(a) shows the performance of single verifier based Sybil detection, including minimum and maximum performance. According to the figures, we note that there is very wide variance in the performance of detecting Sybil nodes. Especially, when the length of random route is smaller, we can observe wider variance.

Fig. 2(b) and Fig. 2(c) represents the performance of multi-verifier based Sybil detection with various threshold, T_{thres} . In these figures, we can observe that multi-verifier based Sybil detection outperforms the single-verifier based Sybil detection in both aspects of accepting honest nodes and suppressing Sybil nodes. Also, we note that there is a tradeoff between the threshold value and the performance of multi-verifier based Sybil detection. With smaller threshold value, multi-verifier based Sybil detection accepts more honest nodes but it may accepts more Sybil nodes as well. On the other hand, with bigger threshold value, it can suppress Sybil nodes aggressively but it may not accept some honest nodes.

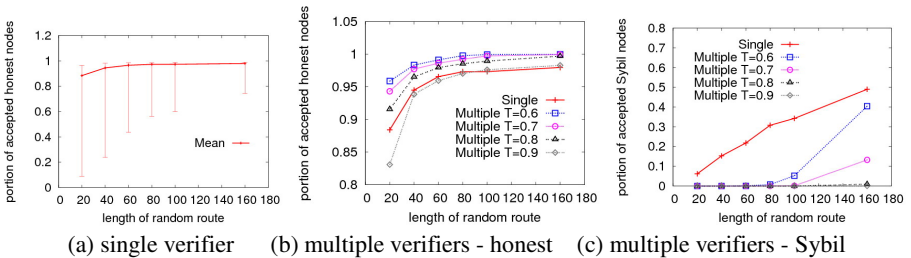
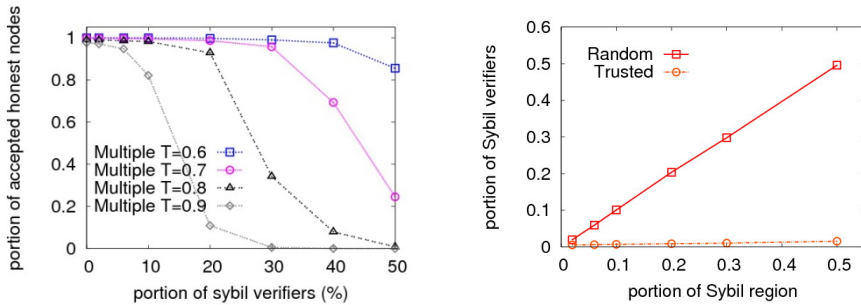


Fig. 2. Portion of accepted nodes as a function of the length of random route

Fig. 3(a) represents the impact of Sybil verifiers. As we can expect, when there are more Sybil verifiers, the performance of multi-verifier based Sybil detection is significantly degraded. That is, choosing honest verifiers is important to guarantee the operation of multi-verifier based Sybil detection. Fig. 3(b) shows the comparison between random verifier selection and the proposed verifier selection (Algorithm 1). In the figure, we can observe that the proposed algorithm works very well to choose honest verifiers among a social network graph where honest and Sybil nodes coexist.



(a) Portion of accepted honest nodes as a function of the portion of Sybil verifiers. Length of Random Route = 100. (b) Portion of Sybil verifiers as a function of portion of Sybil region in a social network.

Fig. 3. Importance of trusted verifier selection. As the portion of Sybil verifiers increases, the probability of accepting honest nodes decreases.

4 Conclusion

The previous OSN based Sybil detection methods can be used for each individual node to determine whether other nodes is Sybil or not. However, it is easily observed that there is the wide variance in the performance of single verifier based Sybil detection. To eliminate this variance, this paper proposes the multi-verifier based Sybil detection, which is composed of two algorithms: 1) Selecting honest verifiers from a social network graph where honest and Sybil nodes are mixed and 2) Determining honest nodes by comparing the likelihood of acceptance of a node with a given threshold. Through the evaluation, we note that the importance of selecting honest verifiers and choosing a reasonable threshold to guarantee the performance of the proposed multi-verifier based Sybil detection method. Currently, we are working on how to choose a reasonable threshold adaptively to an arbitrary social network graph and how to improve the performance of Sybil detection.

Acknowledgement. This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Danezis, G., Lesniewski-laas, C., Kaashoek, M.F., Anderson, R.: Sybil-resistant DHT routing. In: de Capitani di Vimercati, S., Syverson, P.F., Gollmann, D. (eds.) ESORICS 2005. LNCS, vol. 3679, pp. 305–318. Springer, Heidelberg (2005)
2. Sirivianos, M., Kim, K., Gan, J.W., Yang, X.: Assessing the Veracity of Identity Assertions via OSNs. In: Proc. COMSNETS 2012, Bangalore, India, January 3-7 (2012)
3. Sirivianos, M., Kim, K., Yang, X.: SocialFilter: Introducing Social Trust to Collaborative Spam Mitigation. In: Proc. IEEE INFOCOM 2011, Shanghai, China, April 10-15 (2011)
4. von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: CAPTCHA: Using Hard AI Problems for Security. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 294–311. Springer, Heidelberg (2003)
5. Yu, H., Gibbons, P.B., Kaminsky, M., Xiao, F.: SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. In: Proc. IEEE S&P 2008, Oakland, CA (May2008)
6. Viswanath, B., Post, A., Gummadi, K.P., Mislove, A.: An Analysis of Social Network-Based Sybil Defenses. In: Proc. SIGCOMM 2010 (2010)
7. Kim, K.: Sybil-Resistant Trust Value of Social Network Graph. In: Proc. the First International Conference on Smart Media and Applications (SMA 2012), Kunming, Yunnan, China, August 21-24 (2012)
8. Mohaisen, A., Hopper, N., Kim, Y.: Keep your friends close: Incorporating trust into social network-based Sybil defenses. In: Proc. IEEE INFOCOM 2011, Shanghai, China, April 10-15 (2011)

The Impact of Addiction to Twitter among University Students

Shahnil Asmar Saaid, Nalisa Alia Amin Al-Rashid, and Zaridah Abdullah

Faculty of Communication and Media Studies,
Universiti Teknologi MARA (UiTM) Malaysia
{asmarsaaaid, zaridah255}@salam.uitm.edu.my,
aleaamin@gmail.com

Abstract. Twitter has made a big impact on the social environment especially when it comes to young adults that it has become decisively addictive. The objective of this paper was to examine the impact on the addictive usage of Twitter among undergraduates. The questionnaire survey research was conducted on Universiti Teknologi MARA (UiTM) students. The results indicated that of the 100 responses, 34% tended to be heavy Twitter users when they tweeted 5-10 tweets per day. In more extreme case, 10% lost count on how much they tweeted per day. Majority of them (71%) also proclaimed that they were addicted to Twitter with the intention to gain followers (45%), updating news (30%), to know what's happening around them (15%) and to kill their free time (10%). Future studies should assess on the advantages and disadvantages that government agencies or phone service providers may gain based on the findings.

Keywords: Twitter, Addiction, University students.

1 Introduction

For the past few years, microblogging has relatively become a new phenomenon. It allows people to publish short text messages (sometimes a photo, a video, or an audio fragment) in order to update their status (“what are you doing?”), share information, ask questions, or communicate between two or more person. It is a broadcast medium in a form of blogging. Although it can be used for many other purposes, status updates (updating one's own status and reading the status updates of others) is still one of the main uses. Twitter, the most well-known application, was made available in 2006 [10]. Founded by Jack Dorsey, Evan Williams and Biz Stone, Twitter became the communication tool it today by essentially combining three existing technologies: real-time delivery notification dispatch software invented by Dorsey, instant messaging, and text messaging. Twitter allows any of its users to contribute to the market of free ideas on its network and to share their interests, opinions and perspective about life and important societal or political events [3].

The social network has now passed the half-billion account mark — specifically 517 million accounts as of July1, 2012, with 141.8 million of those users in the

United States. Just as most of Twitter's users are coming from outside the U.S., so are the tweets: the top three cities in terms of tweets are Jakarta, Tokyo and London [11]. It is also ranked number 20 in popularity among all social networking sites globally, with it being ranked the most popular microblogging service. Twitter attention and use is proliferating, with estimates that web traffic to the Twitter.com site has grown over 600% from November 2007 to November 2008 [6].

Twitter is a microblogging tool in the internet that provides a "real-time nature" for the users and the followers. According to Junco, *et.al*, [8] even though blog users update their blogs every other day in a long sentences and paragraphs but with Twitter users have the access to update their tweets and their whereabouts several times a day. Users can know how other users are doing and often what they are thinking about now, users repeatedly return to the site and check to see what other people are doing. The large number of updates results in numerous reports related to social events such as parties, baseball games, and presidential campaigns. They also include disastrous events such as storm, fire, traffic jam, riots, heavy rainfall, and earthquakes.

Twitter is the main player when it comes to microblogging in this era especially when it comes to the new generation. Majority of people own a smart phone like Iphone, Blackberry, or Samsung and these types of phones have a Twitter account. The user tends to spend their time consuming on Twitter and also to gain information about current news and issues by Twitter as Twitter implements the 'real-time nature' movement where users update and get information on the spot.

2 Problem Statement

The rise of Twitter has skyrocketed ever since it was launched in 2006 and it has given an overwhelming response including celebrities taking part of having a Twitter account and tweet their babbles to keep in touch with their fans. Some of the users are avid users of Twitter as they will tweet whatever they are doing, thinking and where their whereabouts for the public to see. Though the intention of Twitter is to make people be informed of current issues, some users abused the purpose of Twitter by spreading false news and rumors all over the microblog. Not only it effects negatively on the microblog, the avid Twitter users tend to depend on Twitter as their daily consumption by giving constant personal updates in their Twitter timeline and glue to their laptops or busy fiddling with their smart phones instead of communicating face to face. Users, especially youngsters depend on Twitter to gain information of the current situation and to update their personal thoughts and whereabouts. They have depended on Twitter as a social foundation to be noticeable amongst strangers without knowing who they really are. Some users of Twitter have also been labeled as 'Twitter Famous' in order to gain thousands of followers due to their constant updates of their ongoing lives or tweeting about interesting things that are relatable to the lives of people. The objectives of this research thus to indentify the impact of addiction to Twitter among university students and to assess the factors that influence them.

3 Users and Gratifications Theory

This study employs the basic model of Uses and Gratifications Theory to understand why people actively seek out specific media outlets and content for gratification purposes. The theory discusses how users actively search for media that will not only gratify the need they needed but to enhance their knowledge and social interactions. The theory places more focus on the users instead of the actual message itself by asking “what people do with media” rather than “what media does to people” [9]. In this research, this theory helps to explain why majority of young adults now choose Twitter as a source of information and interaction when it comes to networking and gaining information especially about the current issues. Answers and opinions will be sought out from of the avid Twitter users to concur the research on the role, the impact and the consumption of Twitter.

4 Methodology

For this research, 100 questionnaires were sent to young adults who had signed for a Twitter account in the Faculty of Communication and Media Studies. They represented eight different programmes in the faculty (Journalism, Public Relations, Broadcasting, Advertising, Publishing, Instructional Communication, Interpersonal Communication and Communication Management). These subjects represented a broad spectrum of the population in terms of the usage levels and age. They were self-identified as the problematic users (more or less) of twitter. Subjects were highly animated about the topic. Guided from the existing literature, the results from the research questions including common characteristics of users, situational factors and negative consequences, were built into a conceptual framework for addicted consumption of the twitter. A Statistical Package for the Social Sciences (SPSS) software was used in this study to analyze all data gathered from the survey.

5 Results and Discussion

5.1 Demographic Data

From the 100 respondents 55% were female and 45% were male. 79% were from the age group of 21 to 25 years. This followed by those from the age group of 18 to 20 years old (15%), 26 to 30 years old (5%) and 30 years old and above (1%). The questionnaires were distributed to Semester 2 until Semester 6 students of eight different programmes in the faculty. The majority of the respondents was from MC222 (Public Relations) with the amount of 30%, followed by MC224 (Advertising) with 26%, MC225 (Publishing) 24%, MC223 (Broadcasting) 11%, MC221 (Journalism) 6% and MC227 (Interpersonal) and MC228 (Management & Policy) 1% each.

5.2 The Usage of Twitter and Characteristics of It Addictive Users

A total of 100% respondents had a Twitter account. In terms of financial support, it is found that the majority of the students were relying on their parents for their daily income (54%) while others (35%) were depending on loan or scholarships. All respondents own a Twitter account with the majority (55%) chose smart phones as their ideal device to tweet while other used Ipad (18%), tablet (17%) and laptop (10%). The study showed that most respondents owned a Twitter account for more than a year (54%) with the majority (35%) tweets 1 – 5 a day while 34% responded that they tweeted around 5 – 10 tweets a day. Whereas other respondents (21%) tend to be heavy users when they tweeted 11 – 20 a day and 10% have lost count on how much they tweeted per day. The majority of students (43%) used Twitter to express their feelings, thoughts and situations that they were under while others (31%) used Twitter to keep them updated with the latest issues and events. Meanwhile, 17% used Twitter to keep in touch with their friends on their updates as Twitter is the cheapest way to communicate. Other than that, 5% used Twitter to make new friends, and 4% prefer to use Twitter for other things, such as for entertainment and to fill up their time.

5.3 Factors Influenced the Use of Twitter

This study revealed that 55% of the respondents agreed that friends were the strongest influence for them to join Twitter with the persuasion to update on the latest happenings and to communicate each other easily. Whereas 24% chose to join Twitter out of curiosity since Twitter is basically everywhere. The remaining respondents (15%) felt the peer pressure from being left out if they were not joining Twitter, and 5% joined Twitter with different types of reasons of keeping themselves updated with the latest news of their favourite musicians, celebrities or football news.

5.4 Consequences of Addicted Twitter Usage among University Students

In seeking the respondents biggest impact or consequences of addictive used of twitter, this study reveals that the students used either broadband or wifi to accommodate them to continue tweeting every month. 12% of the respondents spent RM150 – RM200 monthly for their selected bills that noted as the highest amount of payment in the category. 34% of them spent RM100 to RM150 per month while 28% had to allocate the bill around RM50 – RM100 for each month. The remaining 26% represented those who paid around RM35 – RM50 on their monthly bills.

The majority (71%) proclaimed themselves to be addicted to Twitter, with the amount of 45% claimed that they addicted to Twitter to be ‘in the game’ and gain followers. 30% claimed that they addicted due to them spending more than eight hours in Twitter alone to constantly reading and updating tweets in their timeline. With the addiction getting severe, 15% responded that their life revolved around Twitter and would tweet about everything and anything that runs in their life to gain followers and keep their followers updated about them. Meanwhile, 9% were addicted where they used Twitter as a form of communication tool amongst their friends and family since Twitter is cheap and fast. The remaining, (2%) claimed that the reason on why they were addicted to Twitter because they enjoyed reading other people’s tweets and updates.

Talking about the other consequences of addictive twitter users, most of the students agreed that Twitter have benefitted them or the way they live. This is because through Twitter they have gained a lot of friends (29%) and some (22%) have won free giveaways from Twitter due to concert tickets and movie premieres vouchers that were organized by certain company as a marketing and promotion strategy. Other than that, 22% professed Twitter benefitted them by making them happy due to being entertained by the tweets and keeping them updated with the latest events, happenings and news that is not on the mainstream media. The remaining 5% represent those who have benefitted by getting endorsement deals, this is especially to those who are famous because of their tweets and are labelled as 'Twitter Famous' or 'Twitter Celebrity' and have thousands of followers under their belt.

The majority of the students (56%) searched for the comfort from Twitter when they were feeling down, lonely or depressed. They depended on Twitter for comfort by expressing their frustration or just by reading other people's tweets that are comforting, motivational and funny to cheer them back. In Twitter, there are a lot of parody or jokester accounts to tweets jokes every minute and there are even accounts specializes on motivational quotes and advices to the Twitter users. However, 44% disagree about the stated fact.

Based on the findings, 83% agreed that Twitter has a great positive impact on their lives where it has becoming the new media on spreading information and news, as it is fast and easy way to do so by using smart phones or tablets. However, 17% disagreed with the stated fact when they responded that Twitter has contributed to anti-social atmosphere where people tend to neglect real time conversations.

When asked about the medical consequences of using Twitter constantly, the majority (97%) claimed that they are not fully aware of any medical symptoms that can be caused by excessive use of Twitter. In addition, only 6% of the respondents have encountered difficulty in falling asleep when they are in excessive use Twitter for the night. In addition, the majority (97%) disagreed that Twitter have caused them to be more susceptibility to stress and fatigue in any way. This support the view of Griffith (1996), whereby Twitter have extended the concept of addiction from substance dependency of students to problematic behaviours such as excessive internet usage or technology dependence that leads to stress and fatigue due to the minimum amount of sleep since Twitter never stop updating the timeline unless everybody stops tweeting.

6 Conclusion and Recommendation

The famous microblog, Twitter, has definitely made a huge impact in the digital world that it has become a viral use among the new generation especially the young adults. Not only Twitter is used for socializing, marketers and organizations have taken the advantage of using this medium as a tool to spread information, campaign and even to promote their product in the Twitter timeline for free [2].

With the vast popularity of Twitter to average people, whoever that are an avid user or receive a lot of RTs by people, they can gain a lot of followers and able to rise up to their 15 minutes of fame by being 'Twitter Famous'. With local and international news organization joining Twitter, it is no wonder that young adults prefer Twitter more than reading newspaper or any other medium to gain fast information [7].

Celebrities also promote many people to join Twitter as they can follow their favourite celebrities and keeping in touch with their latest updates and information. Same goes to politicians where they can engage the public easier by using Twitter with their updates and their debates on the upcoming elections to gain the public's interest [1].

Though Twitter has its positive impact, there is also a drawback. In the Twitter world, users are shown to be good in socializing but only in a 'virtual' world, in reality, they have trouble communicating to people and are labelled as an 'anti-social' or an 'outcast' [8]. With users being too addicted to Twitter, some users have also receive some medical symptoms on where they have difficulty on falling asleep due to being active on Twitter and wanting to tweet and receive updates. [4].

As recommendation for future research, the reproduction and development of the same study, a comparative analysis study of the addictive usage of Twitter among University students could be made with a developed country such as The United States of America or countries that receive the most Twitter feeds like in Indonesia or Japan and compare those findings. By doing so, the researcher would have bigger insight on the topic. Apart from that, further study can be conducted by using target group such as working adults or school adolescents to determine if similar results are made. Studies on the implication of how Twitter used can be linked to university students in relation to academic outcomes should also be interesting.

References

- [1] Anderson, S.: The Twitter toolbox for educators. *Teacher Librarian* 39(1), 27–30 (2011)
- [2] Bellin, J.: Facebook, Twitter, and the Uncertain Future of Present Sense Impression. *University of Pennsylvania Law Review* 160(2), 331–375 (2012)
- [3] Brock, A.: From the Blackhand Side: Twitter as a Cultural Conversation. *Journal of Broadcasting & Electronic Media* 56(4), 529–549 (2012)
- [4] Davis Jr., C.A., Pappa, G.L., de Oliveira, D.R.R., de L. Arcanjo, F.: Inferring the Location of Twitter Messages based on User Relationships. *Transactions in GIS* 15(6), 735–751 (2011)
- [5] Griffith, M.: Gambling on the internet: A brief note. *Journal of Gambling Studies* 12(4), 471–473 (1996)
- [6] Hughes, A., Palen, L.: Twitter adoption and use in mass convergence and emergency events. In: Landgren, J., Jul, S. (eds.) *Proceedings of the 6th International ISCRAM Conference*, Gothenburg, Sweden (May 2008)
- [7] Junco, R., Heiberger, G., Loken, E.: The effect of Twitter on College student engagement and grades. *Journal of Computer Assisted Learning* 27(2), 119–132 (2011)
- [8] Junco, R., Elavsky, C.M., Heiberger, G.: Putting twitter to the test: Assessing outcomes for student collaboration, engagement and success. *British Journal of Educational Technology* 44(2), 273–287 (2013)
- [9] Katz, E.: Mass communication research and the study of culture. *Studies in Public Communication* 2, 1–6 (1959)
- [10] Lin, F.L., Hoffman, E., Borengasser, C.: Is Social Media Too Social for Class? A Case Study of Twitter Use. *TechTrends: Linking Research & Practice to Improve Learning* 57(2), 39–45 (2013)
- [11] Lunden, I.: <http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012>

Design and Prototype Implementation of Smart-Phone Voice Locker Using Voice Recognition

Won Min Kang, Ki Won Lee, Ji Soo Park, Jong Hyuk Park*

Department of Computer Science and Engineering
Seoul National University of Science and Technology (SeoulTech),
172 Gongreung 2-dong, Nowon-gu, Seoul, 139-743, Korea
wmkang0@gmail.com
{ainves, jisoo08, jhpark1}@seoultech.ac.kr

Abstract. Recently the interest in biometrics rises; voice recognition is highlighted in point that it does not restrict convenience, place and time of its use and is under the development by many organizations. This paper recommends the voice lock that provides the safety and the convenience in the use of smart-phone by utilizing the voice recognition technique and materializes its prototype. Also, it analyzes the current issues and discusses the use case scenario, etc.

Keywords: Voice Recognition, Smartphone.

1 Introduction

Voice recognition is an imputing means making it possible for the machine to recognize the voice which is the basic human communication tool to obtain the wanted information or to perform the wanted behavior. Voice is superior than the other interfaces in respects of convenience and natural of use; research for voice which started from research for the vocal sound dependent on speaker on a small scale of words in respect that it is possible to input and verify the information through telephone wire at a distance and to overcome the restraints of place and time, at present, through the stage in which recognizes the word sound independent on speaker on a large scale of vocabularies, reaches the continuous voice recognition phase in which uses the large vocabularies. [1].

This voice recognition technique is mainly used in mobile or smart building and is regarded as a safer authentication method than the pin number input method or pattern recognition method which is in use with voice recognition technique. The authentication method is certainly required to prevent the individual information and data being used in smart phone from leaking. This paper designed and materialized the methods enabling to perform the self-authentication and the more convenient functional conversion using the voice recognition technique to provide the safety of smart phone.

* Corresponding author.

The constituent elements of this paper are as follows: Chapter 2 is concerned with the research, chapter 3 with the materialization of voice locker, and chapter 4 with the conclusions and the further research way.

2 Related Works

2.1 Voice Recognition HMM Algorithm

The speaker recognition can be categorized into text dependent and text independent. Text dependent means the determined terms, that is the determined words or sentences. In case of text dependent system, due to its features, DTW (Dynamic Time Warping) algorithm is mainly used; therefore it is worrying that the different person can eavesdrop and impersonate. Text independent means that the user speaks the undetermined words. In case of text independent system, HMM (Hidden Markov Model) algorithm is mostly used to decrease the drawback of text dependent system [2].

HMM algorithm is a method to find the most similar model in comparison with each other of the established HMM models after initially generating the hidden Markov models for individual voice signal specific parameter intended to recognize through the recognition learning and generating the similar hidden Markov model for the newly input voice signal specific parameter also. It has a weakness area in which HMM model should be generated for the input voice and compare it with all models each time resulting in decreasing the recognition speed but has a strong area also in which the initial learning speed is high and especially it requires to generate HMM model only for the new word sound to add the word sound of new voice recognition [2, 3].

Fig. 1 shows the training process of HMM algorithm. Training process generates the output stream using voice data to estimate HMM for it by use of the maximum possibility method. The estimation of codebook is performed through the same training for the numerous words to be used in the recognition of new voice.

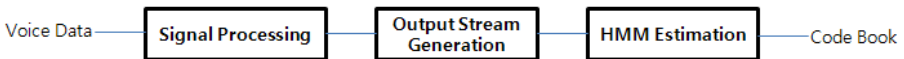


Fig. 1. Training of HMM Algorithm [4]

The signal processing to generate the output stream for voice data is shown on the following Fig. 2. The voice signal in type of analogue is digitalized and fractionated in the small time table. The fractionated time table is called as frame. After leveling the voice signal of frame using windowing, suit the autoregressive model to estimate twelve (12) coefficients. The spectral analysis leads to estimate twelve cepstral coefficients. Twelve cepstral coefficients indicating the voice features should be obtained through the spectral analysis and two energy values should be obtained to simplify all of twenty six values to be vector indicating the voice signal feature of each frame. Vector quantization process assigns one cluster number for twenty six dimensional vectors using the cluster analysis and the discrimination analysis. One value is countered to the frame [4].

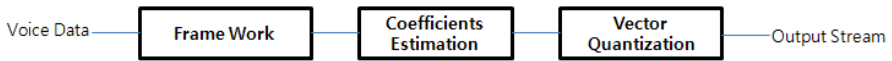


Fig. 2. Process of Generating the Output Stream of the HMM Algorithm [4]

2.2 Google's Voice Recognition System

Google's voice recognition engine does not teach the language directly to the machine but analogizes the user's intention by analyzing the human sound waveform in a statistical manner.

Google progressed the research to increase the speed and the accuracy of voice recognition service using the cloud computing which uses the servers dispersed in the various places as one computer by using HMM algorithm and adopts the method in which the voice data input in smart phone is transmitted to the cloud server and the server recognizes the voice and re-transmit the recognition result to the smart phone [5].

Google's voice recognition system configures as Fig. 3. The system is divided into two large parts of preprocessor and recognition unit; when the user inputs the voice through voice recognition, the system extracts the recognized segment for the sound waveform, performs the noise processing to increase the success probability of recognition, and transmits the voice data to the recognition unit. Recognition unit extracts the features of data and compares them with the data stored in the cloud server. Thereafter, the compared optimum data are shown to the user.

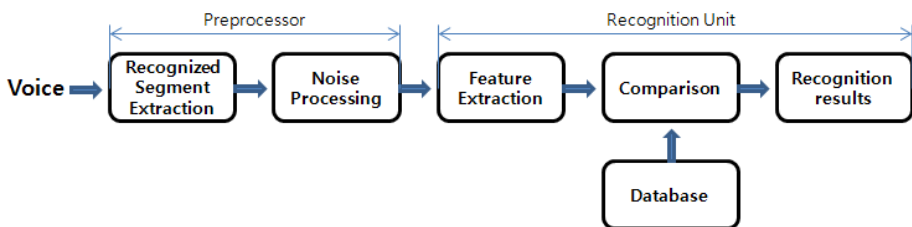


Fig. 3. Google's Voice Recognition System Configuration [5]

The recent research utilizing the voice recognition technique uses HMM algorithm like Google's voice recognition technique mostly in the development process. In some cases, the voice recognition technique is used in the security system.

Voice recognition security system materialized in this paper authenticates the user intended to access the system based on the algorithm by which extracts the features from the input voice signal of speaker and the authenticated user obtains the authority to use the system as allowed by the level authorized to access the system. Recognizer of voice recognition security system enables to recognize the user by performing word matching and sentence matching with HMM model-based tutorial DB (SQL DB); In respect of authorization, the system is materialized so that the user can obtain

the first authority with voluntary access control and the authorized level of system with the forced access control to achieve the voice recognition-based security system to secure the safety of information system [6].

• **Important source code**

```
private void showSelectDialog(int requestCode, Intent data){
    String key = "";
    if(requestCode == GOOGLE_STT)
        key = RecognizerIntent.EXTRA_RESULTS;

    mResult = data.getStringArrayListExtra(key);
    String[] result = new String[mResult.size()];
    mResult.toArray(result);

    str = mResult.get(0);
}
```

Fig. 4. Source Code

Fig. 4 represents a part of Google’s API. In that, the voice recognized through the voice recognition interface is matched to the word stored in Google’s DB through voice understanding and dialogue management to find and return the optimum word. The searched words are input in the form of list and shown to the user.

3 Implementation and Design of Voice Locker Prototype

3.1 Voice Locker Prototype Design

Fig. 5 shows the overall configuration of voice locker system. As the user transmits his/her voice to the smart terminal, the words consisted of sound waves through voice recognition process are transmitted through the language understanding module. Language understanding module generates the semantic representation expressing the speaker’s intention by processing the speaker’s voice to be language and analyzing it. In the dialogue management of the generated meaning representation, the best dialogue strategy is calculated in consideration of the dialogue stream and intention, the meaning representation necessary for the response is generated to match to DB, and the optimum word is selected. For the present DB, a Google DB interlocked

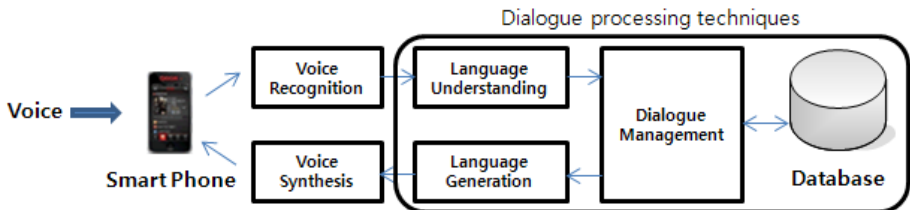


Fig. 5. Voice Locker Prototype Design

to 3 G grid is used. In DB, the optimum word is matched and transmitted to the language generation module in which the word, the sentence, or representation to be responded by the system is generated. Thereafter, it is transmitted to the user for verification.

Fig. 6 shows the scenario in which the voice recognition process is progressing in the voice locker. Actuations are categorized into eight (8) types and functions two times in a large way.

Table 1. Terms explanation

Terms	Explanation
VL_User	Delivered to the voice at VL
VL	Voice is processed and stored in the database
VL_DB	Compare the received words from the VL and the stored words in database

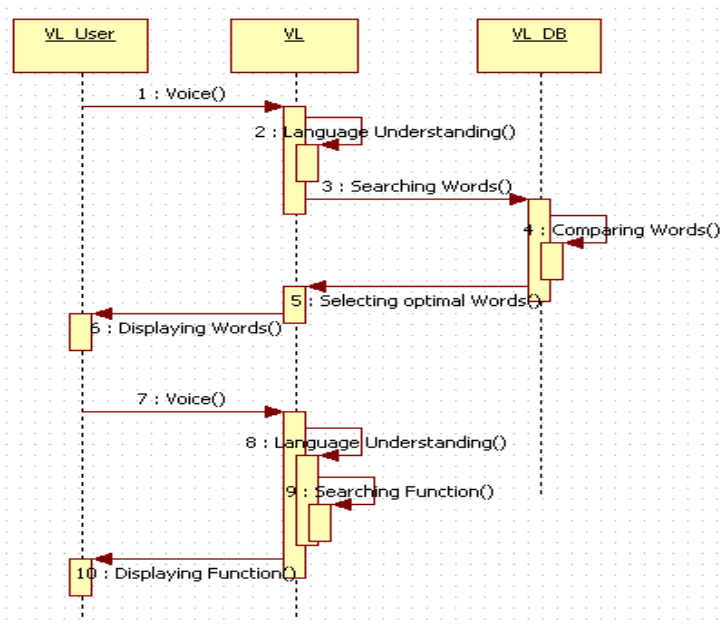


Fig. 6. Voice Locker’s Detailed Scenario

- STEP 1** VL_User → VL, Passing voice
- STEP 2** VL → VL, Language understanding
- STEP 3** VL → VL_DB, Search for a word in the database
- STEP 4** VL_DB → VL, Compare Received words from the VL and stored words in the database

- STEP 5** VL_DB → VL, Select the optimal words
- STEP 6** VL → VL_User, Display the selected words
- STEP 7** VL_User → VL, Passing voice
- STEP 8** VL → VL, Language understanding
- STEP 9** VL → VL_DB, Search for a function
- STEP 10** VL → VL_User, Display the function

3.2 Voice Locker Implementation

This system is designed and materialized to enable to perform the locker function through matching the words registered in Google DB to the words input by user. Voice recognition is authorized only by the words that the user knows and the existing pin number recognition method can also be set for possible use. In point of interface, Google interface is used and the source is materialized so as to be interlocked to the Google’s DB through Google’s API.

Fig. 7 shows the interface of systems and the screen in which voice recognitions are materialized.

- **Interface and voice recognition screen**



Fig. 7. Interface and Voice Recognition Screen

This system includes, by utilizing the voice recognition technology, not only the function of authentication but also the function of functional conversion. The system, by using the function of functional conversion, guarantees to provide users with the convenience and is characterized by it that the only voice enables to convert fast and easy. In the system, by speaking the word of phone and camera through voice recognition, the function is converted and shown.

Fig. 8 shows the screen that the function is being converted. Clicking man-shaped icon on upper screen will have the voice recognition interface appear and speaking the wanted function will have the function appear enabling to use the function.

- **Function switch**

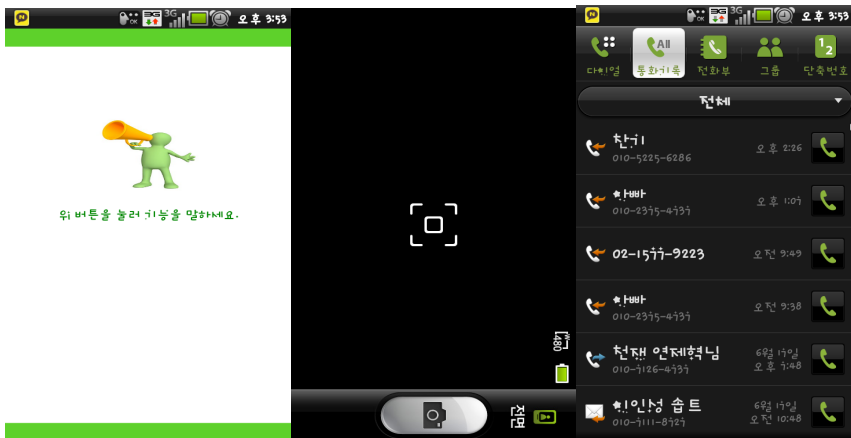


Fig. 8. Function Switch

4 Conclusions

This paper describes how the voice recognition technique materializes the smart phone voice locker. The voice recognition technique is used in the various areas, but if used in the security system, it will provide the higher safety than the existing method of inputting the pin number. Also, the function of functional conversion can be performed more conveniently than the existing method of touching or buttoning way; however, the voice recognition technique that is being used now significantly lowers the capability of recognizing an individual. For the security function, the authentication is essential. In other words, the authenticated user can only perform the function or process the data. In this context, the voice locker that has been materialized up to date is limited in respect of function. In order to recognize an individual in a safer way, it should be possible to find the ways to extract, recognize, and authenticate the individual information; therefore, it is necessary to research the methods of authenticating and extracting the individual information. Doing this will provide the high level of security to prevent the individual information from leaking. For the future research, it is recommended to develop a tight authentication method for and apply to the smart phone user utilizing the voice recognition technique.

Acknowledgments. This study was financially supported by Seoul National University of Science & Technology.

References

1. Lee, S.-W.: Voice Technology. IBM SW Training Data Package, 12 (December 2001)
2. Jang, S.-S.: Analysis of HMM Voice Recognition Algorithm. Journal of Advanced Engineering and Technology 3(3) (October 2010)

3. Kim, J.-H., Ryu, H.-S., Kang, J.-M., Kang, S.-I., Lee, S.-B.: A study on the Speech Recognition Module's Design Using HMM Speech Recognition Algorithm. In: Proceedings of KFIS 2001 Fall Conference (December 2002)
4. Son, G.-T., Jung, S.-H., Bak, M.-W.: A study on HMM for speech recognition. *The Korean Communications* 5(1) (1998)
5. Culture Technology in-depth Reports, Korea Creative Contents Agency (November 2011)
6. Lee, M.-G.: Implementation of Voice Awareness Security Systems. *The Institute of Electronics Engineers of Korea* 29(1) (2011)

Extended RBAC Model with Task-Constraint Rules

Li Ma, Yanjie Zhou, and Wei Duan

School of Mathematical and Computer Science,
Jiangxi Science & Technology Normal University, Nanchang, China
liima@sina.com,
{zyanjwm, wduan1221}@163.com

Abstract. RBAC model supports the principle of least privilege by the appropriate combination of roles assigned to users. However, the minimum role set is hard to find. Role hierarchy and inheritance can result in aggregating lots of permissions. To solve this problem, previous work mainly studies the approximate least privilege set by trying to find the minime role sets. However, to reduce the unnecessary privileges is another key issue to solve the problem. To this aim, the concept of task is taken as a kind of constraints introduced to the RBAC model, which contains four constraint rules that can flexibly control the permission inheritance and role activations. An example is showed the four rules can effectively be used to reduce redundanct permissions.

Keywords: Access control, RBAC, Task, Role.

1 Introduction

The principle of least privilege (LP) states that only the minimum access rights that are necessary to perform an operation should be granted to users[1], which is one of the security principles that RBAC model supports. However, the concrete implementation of this principle in the security design area is often problematic due to the lack of a precise definition[2]. Sandhu et. al think LP can be supported by RBAC model because the components of the model can be configured so that only those permissions required for the tasks conducted by members of a certain role are assigned to the role[3]. In fact, when the number of both roles and permissions are large, it is almost impossible to find a completely accurate set of roles that has the appropriate set of permissions required by a user. Therefore, researchers try to study algorithms to find the minime role sets[4,5]. For example, Dong et al brought forward a Least Privilege Mining Problem(LPM), and algorithms to get an approximating solutions to LPM[4]. Least privileges can be approximately found, which is the opinion of previous work. However, how to reduce the unnecessary privileges is a key issue to solve the problem. As we know, there are many constraint mechanisms in RBAC model to restrict the privileges a user can acquire. For example, the RBAC standard uses static and dynamic separation of duties constraints to restrict the number of activated roles[6]. Since these constraints can reduce the permissions, role hierarchy and inheritance will still result in aggregating lots of permissions in senior roles[7]. Therefore, neglecting

these factors by using the mathematical method to find a least combination of roles is not an effective way. This paper provides another approach to reduce the redundancy of permissions.

The LP principle in RBAC is related closely to the concept of task, which controls the activation of roles as Sanhdu et al. mentioned in RBAC96 model. Since the task is an important element, the concept of task is not defined explicitly in RBAC model, which makes RBAC be granted as a passive access control model[8]. Oh and Park proposed an improved access control model, T-RBAC, by adding the task element as a connection between role set and permission set in RBAC model[9]. In order to remain the core concept of RBAC model, this paper holds that the factor of task is implied in LP mechanism of RBAC model, and we can let it be an explicit constraint to RBAC model, thus, RBAC is both a passive and active access control model. Therefore, this paper proposes a task-constrained RBAC model to enforcement the LP principle.

2 RBAC Model with Task-Constraint Rules

RBAC contains many constraints, such as dynamic and static separation of duties, cardinality constraints of user to role assignments and so on. To analyze the way RBAC supporting LP principle, we think the activation of roles shouldn't be decided by users discretionarily, but by the constraint of specific tasks assigned to users, so we can add a task-constraint to the RBAC model. Task is a fundamental unit of business work or business activity. Oh and Park discussed the task concept in access control model, and defined four classes of tasks: Class *P* (Private), Class *S* (Supervision), Class *W* (Workflow) and Class *A* (Approval for activity), where Class *P* means the access rights for the tasks are not inherited by senior roles; tasks in Class *S* are related to management or supervision and inherited to senior roles; tasks in Class *W* are not inherited to senior roles and belong to a business process; and Class *A* has characteristics of Class *S* and Class *W*.

On the basis of Oh and Park's work, we can further call those tasks in Class *P* and *S* as routine tasks because they are related to users' routine work according to their jobs or duties in organizations, and call tasks in Class *W* and Class *A* as workflow tasks for they belong to workflow. The characteristics of four classes of tasks can be taken as constraints to apply to *RH* and *PA* relations. If we put these factors into RBAC model, we can further understand the flexibility of RBAC model.

Therefore the RBAC model (here we take RBAC96 model as the orginial model)can be extended to task-constrained RBAC model showed in Figure 1.

Definition 1(Task-constrained RBAC model). $M=(U, R, P, S, UA, PA, RH, user, roles, C)$, where: U, R, P, S stand for the set of users, roles, permissions and sessions respectively, and UA, PA, RH stand for the set of user-role assignment, permission-role assignment, and role hierarchy relations respectively; and $user$ is a mapping function from a session s to a user; $roles$ stands for a function that each session s activates the set of roles; C stands for the set of constraints, and $C=(C', T, TR, TP, TC, TH)$, where: C' represents the set of all previously existing constraints, including separation

of duties, cardinality constraints, and so on; T stands for the set of tasks; $TR \subseteq T \times R$, stands for the set of many-to-many assignment relation between tasks and roles, and one task can be completed by multiple roles, and also one role can do many tasks; $TP \subseteq T \times P$, stands for the set of inclusion relation between tasks and permissions, which is a many-to-many relation, i.e. one task contains multiple permissions, and one permission can be contained by many tasks; TC stands for the set of task constraints to RH and PA relations; $TH \subseteq T \times T$, is a inclusion relation between tasks, for example a task t can be decomposed to two sub-tasks t_1 and t_2 , then $(t, t_1), (t, t_2) \in TH$.

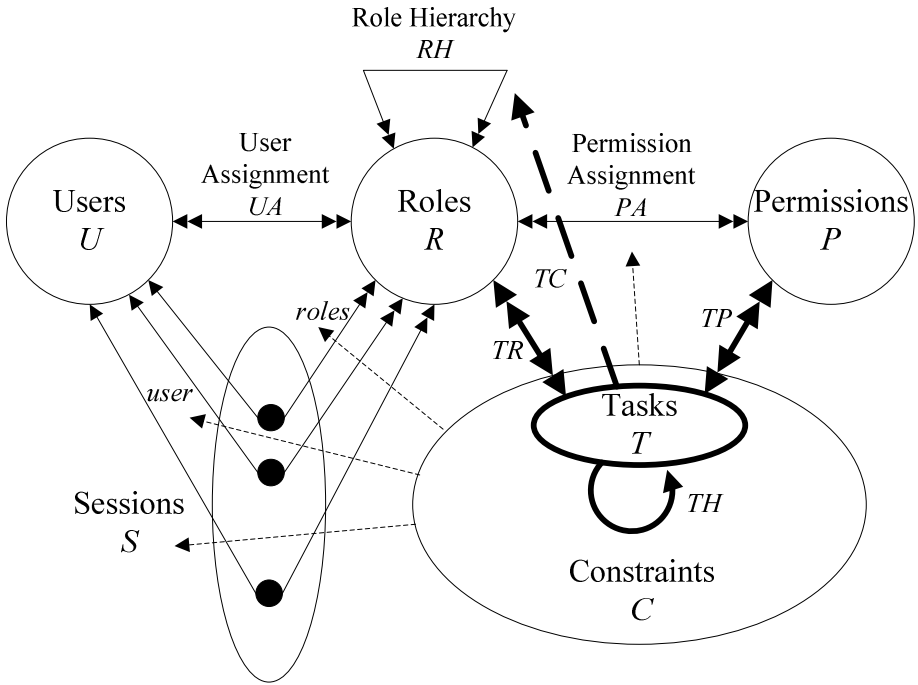


Fig. 1. Task constrained RBAC model

Task-constrained RBAC model is different from the T-RBAC model proposed by Oh, since we don't take task T as an element between roles and permissions, but as a constraint, because the task works as a constraint in the runtime of the access control system. If there are no tasks, the access control system is also complete. In access control systems, the number of designed roles and their assigned permissions are decided by the security strategy of organizations. The task can be decomposed to many sub-tasks, and each task is associated with permissions via TP relations. Then the TR relation can be resulted from the union of TP and PA . However, how to find an appropriate set of roles that owns a minimal set of permissions from TR relation is a troublesome problem. We can add constraints to reduce the redundancy of permissions. Before we get that, we first introduce the concept of permission types.

Definition 2 (Classification of permissions). Permissions can be classified to two types, READ type and WRITE type, where READ refers to the set of permissions that can read but not change any information, and WRITE refers to the set of permissions that can change information.

Definition 1 gives no concrete specification of *TC*. Here we consider several factors. Since there are hierarchy relations between roles, and senior roles can inherit permissions of junior roles, so, the assignment of tasks may difficultly explain which roles should fulfill the task, thus, we give some constraint rules:

- Rule 1: All tasks that can be completed by junior roles can not be assigned to senior roles.
- Rule 2: To those task that completed by junior roles, senior roles can only inherit the READ permissions, not WRITE permissions.
- Rule 3: The permissions for the private task of junior roles can not be inherited by senior roles.
- Rule 4: In any session, if there is no task assignment, the roles can not be activated.

where Rule 1 constrains the *TR* relation, and is also called junior role priority rule, which meets the requirement of organizations that if a task can be fulfilled by a junior role, it need not be assigned to senior roles; Rule 1 and Rule 3 are constraints to *RH* relation, to reduce the number of permissions. And rule 2 states that to those tasks that are already finished, senior roles can not re-modify system's information; Rule 4 is the constraint to *roles* function, i.e., task assignment is the pre-request conditions, and only when the task is assigned can the roles be activated.

With the task constraints, RBAC is also an active permission management model.

3 Example and Analysis

This section describes an example of a sales department (the example is originally from [9] with a little change). The role *sales_director* is a superior to all the roles, and *sales_manager* is senior to *sales_clerk* and *sales_man*, where *sales_clerk* and *sales_man* are mutual exclusive roles. The *TP* assignment is as Table 1.

Table 1. Task-permission relation

Task	Permission
T ₁	Read order sheet P ₁
T ₂	read customer information P ₂
T ₃	fill price form P ₃ , read sales price P ₇
T ₄	read customer information P ₂ , fill customer verification information P ₄
T ₅	read order sheet P ₁ , read customer verification information P ₅ , fill approval sheet P ₆
T ₆	read order sheet P ₁ , read sales price P ₇ , fill approval sales P ₈
T ₇	read stock sheet P ₉ , read out of stock sheet P ₁₀ , fill out of stock sheet P ₁₁ , make delivery plan P ₁₂
T ₈	write customer information P ₁₃
T ₉	make marketing plan P ₁₄
T ₁₀	review employee record of sales P ₁₅ , fill employee evaluation form P ₁₆
T ₁₁	review all customer's information P ₁₇

We can classify the above tasks as follows: Class *W*: {T₁, T₂, T₃, T₄, T₇}; Class *A*: {T₅, T₆}. The routine task of *sales_man* includes making plan, writing customer’s information, and so on. The routine task of *sales_director* is to review employee’s achievement, make development plan and so on. Therefore, we can define Class *P*: {T₈, T₉}; Class *S*: {T₁₀, T₁₁}. The instance of the workflow task can be T₁→T₂→T₃→T₄→T₅→T₆→T₇. Thus, *TR* relation can be represented as follows: *sales_director* is assigned tasks T₆, T₁₀, T₁₁, *sales_manager* is assigned tasks T₅, *sales_clerk* is assigned tasks T₄, T₇, and *sales_man* {T₁, T₂, T₃, T₈, T₉}. Here we use Rule 1 to restrict *TR* relation. For example, both *sales_manager* and *sales_clerk* can execute T₄, while since *sales_clerk* is lower than *sales_manager*, the T₄ can be assigned to *sales_clerk* first.

Before the task is assigned, permission-role assignment *PA* and role hierarchy relation *RH* can determine the permission that every role can be assigned. RBAC standard takes the role hierarchy as role inheritance, i.e., permissions of the junior role can be transferred to senior roles. We can compare the permissions assigned before and after we use task-constraints. Table 2 shows all permissions each role assigned.

Table 2. Permission assignment with inheritance

Role	Permission
<i>sales_director</i>	P ₈ , P ₁₅ , P ₁₆ , P ₁₇ , P₁~P₇, P₅, P₆, P₂, P₄, P₉, P₁₀, P₁₁, P₁₂, P₁₃, P₁₄
<i>sales_manager</i>	P ₅ , P ₆ , P₂, P₄, P₉, P₁₀, P₁₁, P₁₂, P₁, P₃, P₇, P₁₃, P₁₄
<i>sales_clerk</i>	P ₂ , P ₄ , P ₉ , P ₁₀ , P ₁₁ , P ₁₂
<i>sales_man</i>	P ₁ , P ₂ , P ₃ , P ₇ , P ₁₃ , P ₁₄

where all inherited permissions are represented in **bold** type.

Permissions can be classified to READ={ P₁, P₂, P₅, P₇, P₉, P₁₀, P₁₅, P₁₇}, WRITE={ P₃, P₄, P₆, P₈, P₁₁, P₁₂, P₁₃, P₁₄, P₁₆}, and the set of permissions of private task={ P₁₃, P₁₄, P₁₅, P₁₆, P₁₇}, which can not be inherited w.r.t. Rule 3.

According to the four rules, we can reduce some unnecessary permissions. Since tasks T₁ to T₅ have already been completed, *sales_director* can only inherit all READ permissions from junior roles according to Rule 2. And as to rule 3, *sales_director* can not inherit P₁₃ and P₁₄ from *sales_man*. And P₁₁ and P₁₂ of *sales_clerk* in T₇ can not be inherited yet w.r.t. Rule 1. Thus **P₁, P₂, P₅, P₇, P₉, P₁₀** are inherited permissions of *sales_director*. The set of permissions of *sales_director* are {**P₁, P₂, P₅, P₇, P₉, P₁₀**, P₈, P₁₅, P₁₆, P₁₇}. Therefore, the permission number is reduced from 17 to 10. Similarly, the set of permissions of *sales_manager* is {**P₁, P₂, P₇, P₉, P₁₀**, P₅, P₆}, which is reduced from 13 to 7. The permissions of other two roles, *sales_man* and *sales_clerk*, are still unchanged since those role don’t inherit any other permissions.

With the task-constraints, users can get less permissions without affecting the completement of tasks. However, the result permissions are not the least permissions yet. For example, *sales_man* is assigned permissions {**P₁**, P₂, P₃, P₇, P₁₃, P₁₄} to execute task T₁, but only **P₁** is useful. To task T₄ and T₅, *sales_clerk* has permissions { **P₂, P₄**, P₉, P₁₀, P₁₁, P₁₂}, where useful permissions are **P₂** and **P₄**, and *sales_manager* has permissions { **P₁, P₂, P₇, P₅, P₆**, P₉, P₁₀}, where **P₁**, **P₅** and **P₆** are useful. The permissions assigned to some role could be too coarse-grained for the existing tasks.

4 Conclusion

RBAC can support LP principle by activating parts of roles, and users of senior roles can selectedly activate many junior roles. However, the way to activate the roles should not be decided by users discretionarily. To restrict the permission inheritance and role activation, this paper proposes a task-constrained RBAC model to control the activation by adding four constraint rules. With the limit of space, this paper only gives an example to explain how these rules can reduce the number of permissions. The rigorous analysis will be given in future work.

References

1. Saltzer, J.H., Schroeder, M.D.: The Protection of information in computer system. Proceedings of the IEEE 63(9), 1278–1308 (1975)
2. Byuens, K., Scandariato, R., Joosen, W.: Least privilege analysis in software architectures. *Software and System Modeling* (November 2011)
3. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based Access Control Models. *IEEE Computer* 29, 38–47 (1996)
4. Dong, L.J., Kang, X.J., Wang, M.C.: How to Find a Rigorous Set of Roles for Application of RBAC. *Journal of Software* 7(2), 398–407 (2012)
5. Wei, L., Cai, J.-Y., He, Y.-P.: Implicit Authorization Analysis of Role-Based Administrative Model. *Journal of Software* 20(4), 1048–1057 (2009) (in Chinese)
6. ANSI INCITS 359-2004. Role Based Access Control. American National Standards Institute (2004)
7. Sandhu, R., Bhamidipati, V.: The ASCAA Principles for Next-Generation Role-Based Access Control. In: Proceedings of 3rd International Conference on Availability, Reliability and Security, Barcelona, Spain (2008)
8. Deng, J.B., Hong, F.: Task-Based Access Control Model. *Journal of Software* 14(1), 76–82 (2003) (in Chinese)
9. Oh, S., Park, S.: Task-role-based access control model. *Information Systems* 28, 533–562 (2003)

Historical Data Recovery from Android Devices

Yitao Yang, Zhiyue Zu, and Guozi Sun

College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, China
{youngyt, B10040132, sun}@njupt.edu.cn

Abstract. The popularity of Android devices has brought convenience to the life of people. However, it also has brought new information crime. In order to master the evidence of crime, accurate and highly efficient digital forensics technology is urgently needed. Data recovery is an important part of digital forensics. A large number of Android devices use YAFFS2 file system. The traditional file undeletion technology can hardly meet the data recovery requirements of current smart devices. Based on the characteristics of YAFFS2, this paper proposed a data recovery solution and fulfilled a forensics tool named `yaffs2_recovery`. It not only realized the file undeletion but also recovered the historical data of the files. The experimental results showed that the recovered historical data can provide valuable clues to digital forensics analysis.

Keywords: Digital Forensics, Data recovery, Historical data, Android, YAFFS2.

1 Introduction

Android is a fast growing mobile operating system with rich functions. Mobile devices equipped with Android can work as personal computers. People can install various applications and access the high-speed wireless network. According to statistics, by the end of 2012, Android has occupied the biggest market share of smart mobile devices system. At present it seems that this trend will continue.

Android devices bring convenience to the life of people. However, it also brings information crime activities by aid of smart mobile devices, such as spam and scam messages, privacy leakage and password theft. In these crimes, the mobile devices of the criminal suspects store a lot of important information, which will provide beneficial evidence and clues on criminal investigation. Digital forensics is one of effective means to fight against information crimes. The purpose of forensics is to reveal information as much as possible for investigation [1].

A large number of cases indicate that most related information about the crime will be deleted timely. It is hard to obtain the important information through the general forensics means. Therefore data recovery is a key technology in digital forensics science.

The traditional recovery methods are file undeletion technologies. They aim at the deleted files and try to recover them to the state before deletion. Android devices use NAND flash memory as storage media and install specific file system adapting to the

physical features of flash memory. YAFFS (Yet Another Flash File System) is one of the file systems designed for flash memory [2]. Several file undeletion technologies for this file system have been developed [2][3].

In many cases, a file will often be modified instead of being deleted. For example, SMS (Short Message Service) database file *message.db*. Due to the high frequency use of SMS, the message records will be frequently inserted and deleted in the database file. However, the file still exists in the memory. File undeletion technologies cannot recover anything about the deleted message records. Paper[4][5] proposed database record recovery technologies for SQLite, which is the most common database used by smart mobile devices. Nevertheless, they may work- ineffectively when the deleted records have been erased from the flash memory by file system. Consider the following scenario: the file size of *message.db* is reduced from 100KB to 80KB by file system because of garbage collection. The schemes proposed in [4] or [5] cannot recovery records belong to the reduced 20KB data because they don't exist in *message.db* any more. For this situation, a recovery scheme is proposed in this paper. This scheme traverses the entire file system image and recovers all the historical data of files in the image. The recovered data will be valuable for further forensic analysis. The paper is arranged as follows. In the Section 2, the features of YAFFS2 file system will be discussed. Section 3 describes the recovery scheme including pseudo code. An experiment will be demonstrated in Section 4 and Section 5 is brought to a conclusion.

2 YAFFS2

Flash memory is commonly known as the solid-state memory, which is suitable for serving as the storage of mobile devices. It is constituted by a series of fixed size blocks and the number of write operation on these blocks is finite. *Wear leveling* is a mechanism used to make the write operations on all blocks evenly. Since the traditional desktop file systems failed to take wear leveling into consideration, these file systems are unsuitable for flash memory. YAFFS is the first file system designed for flash memory. YAFFS2 is the latest version, which is compatible to YAFFS1 and supports greater storage capacity[5]. The other features are listed as below.

- YAFFS2 has a true log structure and has abandoned the deletion marker. This means YAFFS2 writes sequentially.
- YAFFS uses *chunk* as a data unit. Each chunk has an OOB (Out-Of-Band) area for extra description. If a chunk has a size of 2 KB, its OOB area is 64 Byte in size.
- *chunk* has two types: data chunk and Object Header chunk.
- Each file in YAFFS2 has its own unique number, called Object ID, which is similar to the *inode* number in ext3.
- To take *chunk size* of 2KB as an example, every 64 chunks belongs to the same *block* and shares a common Block ID.

Following example describes some YAFFS2 key features which our recovery scheme is based on.

- Action 1. Creation of a new file in 3K Bytes. (State of chunk is depicted in Table 1)
- Action 2. Modification of the file in action 1. The file is added 2K Bytes in addition. (See Table 2)

Table 1. State after creation of a file

Block ID	Object ID	Type	File Size
10	125	Object header	0
10	125	Data	2K
10	125	Data	1K
10	125	Object header	3K

Table 2. State after modification of the file

Block ID	Object ID	Type	Size
10	125	Object header	0
10	125	Data	2K
10	125	Data	1K
10	125	Object header	3K
11	125	Data	2K
11	125	Data	2K
11	125	Data	1K
11	125	Object header	5K

As can be seen in Table 1 and Table 2, it is not difficult to find that the modification will write the new content in the next block. Finally an Object Header will be appended. The only difference of data versions is the Block ID, so we can identify historical versions of a file according to its Block ID.

Obviously, the historical data of files are stored temporarily in flash memory until YAFFS2 starts garbage collection mechanism. So the time of garbage collection will directly affect the effectiveness of historical data recovery.

3 Recovery Scheme

According to the design principle of YAFFS2 in Section 2, the Object ID of all history versions of a file is identical. Since the chunks with the same Object ID might cross several blocks, these chunks can be collected into a group. Each group has three members that are Object ID, Block ID and file point offset. The file data can be easily read with offset so storing file data into group is unnecessary.

Since all historical versions of a file will belong to one group, different files will belong to different group. A custom data structure *super* is used to organize these groups. Fig 1 describes the process of group collection.

As shown in Fig. 1, group collection is a chunk traversal in file system image. *super* is defined as a *map*, one class of STL(Standard Template Library). The primary key in *super* is Object ID and the value of the key is another *map*, denote by *map'*. Block ID is the primary key of *map'*. A *vector*, another class of STL, is the value of *map'*. *offset* in *vector* means the data offset in image. The entire definition of *super* is described in C++ style .

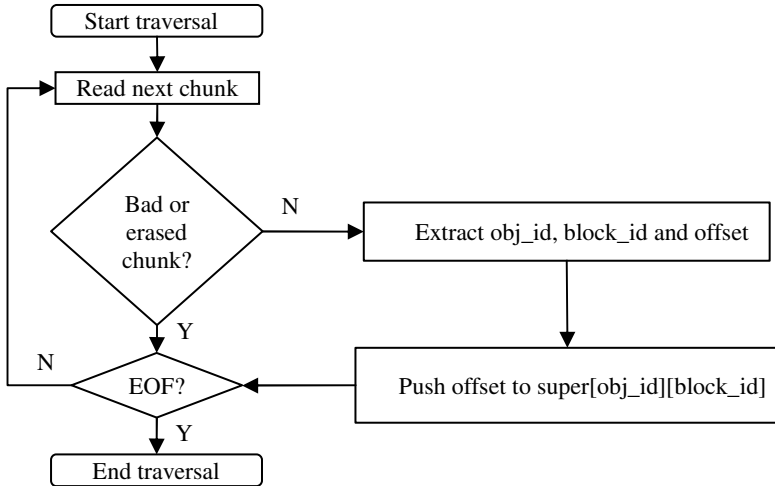


Fig. 1. Flow diagram of group collection

```

super: map<objid_t, map'>
map': map<blockid_t, vector<offset_t>>
  
```

The recovery process reads chunk according to offset in *super* at first. Then recovers the historical versions of files according to the chunk type and copy them to the result directory. Each historical version of file is attached data and time of modification as suffix. Finally, the historical versions of all files which can be recovered are stored in a specific directory. Each version of same file can be distinguished from their file name.

A digital forensics tool named *yaffs2_recovery* is developed under the scheme mentioned above. Recovery experiments using this tool are conducted in next section.

4 Recovery Experiment

This section will carry out a demonstration of recovering file historical data on HTC Desire smart phone. Before obtaining the device image, root authority of the device has been acquired.

4.1 Tool Preparation

ADB (Android Debug Bridge) is commonly used to connect with Android devices. *nandread* runs on the device and can obtain the NAND flash memory image with OOB data. Fig.2. shows the OOB data organized by *nandread*.

	block id	object id	type	chunk id	file size	
0000800:	1032 0000	0204	0000	a805 0000	0008 0000	.2.....
0000810:	ffff ffff	ffff	ffff	ffff ffff	ffff ffff
0000820:	ffff ffff	ffff	ffff	ffff ffff	ffff ffff
0000830:	ffff ffff	ffff	ffff	ffff ffff	ffff ffff

Fig. 2. OOB data organized by *nandread*

4.2 Acquisition of YAFFS2 Image

The flash memory in Android devices is divided into some partitions. These partitions are managed by the MTD (Memory Technology Device). In order to see them, the command *cat* can be executed: *adb shell cat /proc/mtd*. The result of executing this command is showed as Fig.3.

```
$ adb shell cat /proc/mtd
dev:   size  erasesize  name
mtd0: 000a0000 00020000 "misc"
mtd1: 00480000 00020000 "recovery"
mtd2: 00300000 00020000 "boot"
mtd3: 10e00000 00020000 "system"
mtd4: 02800000 00020000 "cache"
mtd5: 07fa0000 00020000 "userdata"
```

Fig. 3. The result of command *cat*

In order to obtain the *userdata* partition image, the following *nandread* command can be executed:

```
adb shell nandread-d /dev/mtd/mtd5 -f/sdcard/mtd5.nandread.
```

After pulling the image file *mtd5.nandread* to the computer, the acquisition of YAFFS2 image is finished.

4.3 Experimental Results

yaffs2_recovery is a Linux executable program developed with the principles mentioned in Section 3. It is used to analyze the image file obtained in Section 4.2. Now the open source is available at <https://github.com/TissueFluid/yaffs2-recovery/>.

The result data generated by *yaffs2_recovery* has 77.4M Bytes in size and 1466 items of file historical data. All of them are stored in a specific directory. In order to test the performance the recovery scheme, two versions of a SQLite file are selected

from result data. In the recovery process, the attribute mtime was attached to the each file name so the modify time of the files can be recognized easily. As shown in Fig.4, the two historical versions of message.db were respectively generated at 9:30am and 9:35am on January 22, 2013.

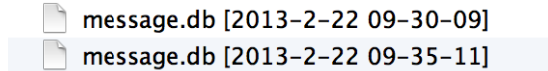


Fig. 4. The historical data list of a specific file

A part of records in these two files are listed by SQLite tool in Fig.5 and Fig.6. This example shows a fact that the user deleted one message of which target is Tim between 9:30am and 9:35am on January 22, 2013. It might possibly be a useful clue for investigation.

• target	content	time
Jasey Wang	Call me at 10.	2013-1-19 07:56:04
Magared	I have no time, maybe next Friday	2013-1-21 10:11:42
Tim	It really sucks.....	2013-1-21 20:08:10
Tom	Tomorrow is the deadline	2013-1-21 22:09:01

Fig. 5. Records in message.db [2013-2-22 09-30-09]

• target	content	time
Jasey Wang	Call me at 10.	2013-1-19 07:56:04
Magared	I have no time, maybe next Friday	2013-1-21 10:11:42
Tom	Tomorrow is the deadline	2013-1-21 22:09:01

Fig. 6. Records in message.db [2013-2-22 09-35-11]

5 Conclusion

The historical data of files represent the modification records of them. The experimental results show that the historical data is very important and useful for the forensic analysis and investigation. This solution not only makes up for the defects of the file undelete technologies but also is a significant complement to the data recovery technologies of Android devices.

References

1. Hoog, A.: Android Forensics: Investigation, Analysis and Mobile Security for Google Android. Syngress (2011)
2. Zimmermann, C.: Mobile Phone Forensics: Analysis of the Android Filesystem (YAFFS2). Diss. Master's thesis, University of Mannheim (2011)

3. Jovanovic, Z., Redd, I.D.D.: *Android Forensics Techniques* (2012)
4. Jeon, S., et al.: A recovery method of deleted record for SQLite database. In: *Personal and Ubiquitous Computing*, pp. 1–9 (2012)
5. Wu, B., Xu, M., Zhang, H., Xu, J., Ren, Y., Zheng, N.: A Recovery Approach for SQLite History Recorders from YAFFS2. In: Mustofa, K., Neuhold, E.J., Tjoa, A.M., Weippl, E., You, I. (eds.) *ICT-EurAsia 2013*. LNCS, vol. 7804, pp. 295–299. Springer, Heidelberg (2013)
6. Manning, C.: How YAFFS works, 6: p. (2011) (retrieved April 2010)

A Static Semantic Model for Trusted Forensics Using OCL

Zehui Shao¹, Qiufeng Ding¹, Xianli Jin^{1,2}, and Guozi Sun^{1,2}

¹ College of Computer, Nanjing University of Posts & Telecommunications, Nanjing, China

² Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks,
Nanjing, China

{1011041126, Y004091410, jxl, sun}@njupt.edu.cn

Abstract. According to the features of various properties of digital data, a static semantic model of features for trusted digital data using OCL (Object Constraint Language) is proposed. These features obtained from the forensic domain of digital data are hierarchically decomposed and merged based on FODA (Feature Oriented Domain Analysis) modeling process. Then a feature tree is built with semantic logical relation in order to get the overall semantic description of features in the forensic domain of digital data, meanwhile, formally describing the features of various attributes of digital data by OCL which has a rigorous mathematical semantics and is easy to understand. The features of digital data are classified with the concept of set in OCL, and the relevance and dependence among various features are described with the operations of set in OCL. Finally, a feature model is built in digital data of Windows system with the use of OCL operations.

Keywords: information security, trusted forensics, OCL, feature, FODA, digital forensics.

1 Introduction

The development and popularization of computer and network technology has provided much convenience for the individual's daily life and work. However, the number of the crime with the use of computer and network is increasing day by day, and there are more and more ways to commit this kind of crime [1]. It is the main responsibility of computer forensics to survey, acquire, analyze, and show the relevant digital data in court [2]. Computer forensics is to acquire, save, analyze, and display the evidence of computer crime, in accordance with the norms of law [3]. Nowadays, besides the further development of the traditional technology, the main research of computer forensics is dynamic computer forensics and formalized forensic analysis technology [4].

As far as the computer forensics is put into practice, the basic way to take the evidence from digital data mainly depends on forensic officers and forensic tools [5]. There are two problems existing in this "individual and tools" model. The first one is that the experience and character of forensic officers can deeply affect the consequence

of forensic. The second is that the reliability and precision of the forensic tools can also change the result. As a conclusion, the extended model can be described as “experience-share, formalized model, and testimony”, that is to say, when focusing on obtaining the evidence, we should share as much knowledge and experience as possible based on the experienced forensics officers, and construct a formalized forensic model based on logic, automated, and the related mathematic theories. At the same time, the validity of this model should be testified theoretically so as to construct a trusted forensic testimony system, which leads the forensics system of digital data more mature, and the data more persuasive [6].

According to the expanded model “experience-share, formalized model, and testimony”, we can conclude the formalized model as the formalized system model of trusted digital forensics. It includes three aspects below:

- 1) The concept of trusted digital forensics.
- 2) The static semantic model of features for trusted digital data.
- 3) The dynamic semantic model of description for trusted digital data.

This paper mainly studies the first and second aspects.

2 The Formalized System Model of Trusted Digital Forensics

2.1 The Concept of Trusted Digital Forensics

The trusted computing is the action of trusted components, and its operation or process is predictable under any operating condition [7]. It can as well resist the destruction from application software, virus, and physical interference. The ultimate purpose is to ensure the information integrity of the entity [8]. In the process of digital forensics, we can establish a trusted evidence theory system referring to the thought of trust and trusted computing [9]. The digital forensics can be divided into two aspects: one is the trust of static property, namely to ensure that the static semantic features on process of trusted discovery and trusted fixation is trusted; the other one is the trust of dynamic behavior, namely to make the process or trusted behavior of digital data transformed into evidence trusted from the following three aspects: trusted extraction, trusted analysis, and trusted expression. Regarding all the above as the foundation, we can realize the trusted extraction and trusted analysis of designated digital data. Fig.1 shows the framework of the trusted digital forensics.

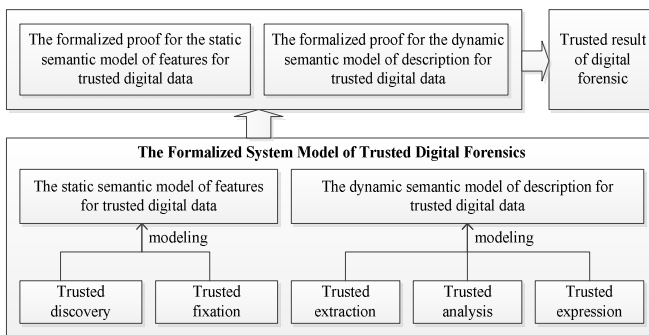


Fig. 1. The framework of the trusted digital forensics

2.2 The Research Method of the Static Semantic Model of Features for Trusted Digital Data

As the digital data has all kinds of features such as paralleling [10], asynchronous [11], distributive [12], and randomness [13], certain relationship and dependency exists between these digital data, for example, the destruction of one piece of digital evidence can influence the validity of other related digital data [14]. Accordingly, how to preserve the digital data is one of the most important aspects in trusted digital forensics. This paper would build a static semantic model of features for trusted digital data with the use of FODA (Feature Oriented Domain Analysis). These features obtained from the forensic domain of digital data are hierarchically decomposed and merged based on FODA [15] modeling process, and a feature tree about semantic logical relation is built to obtain the overall semantic description of the features in the domain of digital forensics. The tree structure can help to describe the overall field of digital forensics in a graphic and mathematic way. Meanwhile, the paper adopts the understandable OCL (Object Constraint Language) with rigorous mathematical semantics to describe all kinds of features of digital data formally, uses the concept “collection” in OCL to classify the digital data, and utilizes the “collection operations” to describe the relationship and dependency between these features.

3 Building the Static Semantic Model of Features for Trusted Digital Data

3.1 Object Constraint Language

Object Constraint Language, OCL for short, is a kind of declared language which is used to constrain the informed model element [16].

OCL is such an enquiry language that any operations in OCL cannot influence the model itself. The “collection” in OCL is referred to a kind of data type with four concrete collection types [17]. It includes:

- 1) Set: the mathematic collection, random ordering, without repetitive elements.
- 2) Ordered Set: it is a kind of mathematic set, whose elements' order is determined by where they are located in the set, thus, there is no duplicate element.
- 3) Bag: it can contain duplicate element.
- 4) Sequence: the elements in it are ordered bags.

OCL defines the operations of some collections as follows:

- 1) Select: it filters out some elements according to some informed rules in the collection to form a new collection.
- 2) Reject: it is opposite to the first operation, which filters out elements that don't meet the demand to form a new collection.
- 3) Collect: it creates a new collection which extracts interesting information from one or more collections.
- 4) For all: it specifies the constraint applied to every element in the collection.

- 5) Exists: it makes certain whether a value exists in the collection.
- 6) Iterate: it has access to every member of the collection and operations according to certain rules.

Iterate is a kind of general operation; while select, reject, for all, exists, collect can be expressed by one of iterate expressions.

These operations are to generate new collections from the current collections by using flexible and powerful means. By using collections and collection operations, features can be classified, and the relation and dependency between features can be expressed.

As OCL is query language, any operations on OCL will not influence the model, which means the current status of the system will not change with the OCL constraint expression [18]. Just as a result of the characteristic, OCL can be applied to the field of digital forensic, and conform the forensic rules (it does not deal with the raw data).

By using collections and collection operations, features can be classified, and the relation and dependency between features can be expressed.

3.2 The Feature Model of Digital Data

Feature is perception of the system from certain perspectives, which can be regarded as an information collection of describing domain models [19].

FODA modeling is related to two conceptions: abstract and refining. Abstract describes the common requirement in the field of application system, namely, abstracts the generality. Its main purpose is to acquire the field model. Refining can be used to develop the specific applications. When modeling in the process of FODA field, and by decomposing and merging the abstracted features in the field of application system, a feature tree of semantic logical relation is built. Then the overall feature semantic description of the forensic domain of digital data is obtained, and DFM (Domain Feature Model) [20] is modeling.

Definition 1. FM (Feature Model) can be expressed as a triple group < feature, relation, constraint >. Feature is the feature collection, whose elements are the feature items in the feature tree; relation is the logical relation collection of feature; constraint is the constraint collection of features and their relations.

Definition 2. DDFM (Digital data Feature Model). DDFM is a kind of construction model of feature models, it should satisfies the following constraints:

```

context DDFM inv :
self.stereotype = FM and
self->ownedElement->forAll (p.p. ocl Is KindOf (Feature) or p.p.ocl Is KindOf (Relation) or
p. oclIsKindOf (Constraint)) and
self.Feature->includes (name) and
self.Feature->includes (type) and
self.Feature->includes (size) and
self.Feature->includes (location) and
self.Feature->includes (createTime) and

```

```
self.Feature→includes (visitTime) and
self.Feature→includes (modifyTime) and
self.Feature→includes (content)
```

feature, relation, and constraint is the elements of DDFM; name, type, size, location, and content is the name of digital data (the identification of digital data), type (such as txt, jpg, pdf, rmvb, etc), size, path, and content; createTime, visitTime, modifyTime is the creating time, the last access time, and the last modifying time of the digital data.

In the DDPM, the element Constraint is the feature-constraint description of the components. It includes the constraint of inherent features, as well as the constraint of extended features. According to the different relations between kinds of features, we can divide them into four classes: self-constraint, “father-son” feature relation constraint, dominant relation constraint, and recessive relation constraint.

1) Self-constraint

Definition 3. Self-constraint. Self-constraint is the rule that digital data features should meet for some constraint of their own properties.

In the DDPM inherent feature, for example, name has the typical feature of the constraint. Its uniqueness should satisfy the constraint as follows:

```
context name inv :
self.ocl Is KindOf (Feature) and
self.allInstances.value→isUnique (nl)--self. allInstances→forall (f1, f2|f1<>f2 implies
f1.value<>f2.value)
```

2) “Father-son” Feature Relation Constraint

Definition 4. Father feature and son feature. Assume that there are two features fa and fb. If fa is the refinement of fb’s direct semantic level, we can conclude that fa is fb’s son feature, and fb is fa’s father feature.

In the digital data feature tree, “father-son” feature can be divided into mandatory constraint, optional constraint, and multiplicity constraint.

Definition 5. Mandatory. If a son feature cFeature and father feature pFeature satisfy the constraint:

```
context cFeature, pFeature inv mandatory :
pFeature.subFeature→notEmpty( ) implies pFeature.subFeatures→includes(cFeature)
```

We can say that cFeature is more mandatory, compared to pFeature, remember it as mandatory(pFeature, cFeature).

Definition 6. Optional. If a son feature cFeature and father feature pFeature satisfy the constraint:

```
context cFeature ,pFeature inv option:
pFeature.subFeatures→notEmpty( ) implies (pFeature.subFeatures→includes(cFeature) or
pFeature.subFeatures→excludes(cFeature))
```

We can say that cFeature is more optional, compared to pFeature, remember it as option (pFeature, cFeature).

Definition 7. Multiplicities. Multiple constraint is the constraint of repeat numbers that father to son.

UML2.0 rules show four kinds of repeat numbers, as Table 1 shows.

Table 1. Multiplicity

Symbol	Lower bounds	Upper bounds	Explanation
0..1	0	1	At most one instance of the sub-feature
1	1	1	Only one instance of the sub-feature
0..*	0	infinity	Arbitrary
1..*	1	infinity	At least one instance of the sub-feature

In DDPM, the multiplicity between “father-son” features is a positive integer or interval. As a result, we can expand UML repeat number so as to support more repeat numbers with more form. As the repeat numbers are nonnegative integers, repeat numbers can be divided into 3 classes: N , $N_1 .. N_2$, and $N_3 ..*$; ($N > 0$, $N_2 > N_1 = 0$, $N_3 = 0$). The multiplicity constraint between “father-son” features can be described by OCL below:

```

context cFeature ,pFeature
inv :
  pFeature.subFeatures→includes(cFeature) and cFeature.allInstances→size () = N
context cFeature ,pFeature
inv :
  pFeature.subFeatures→includes(cFeature) and cFeature.allInstances→size () >= N1
and
  cFeature.allInstances→size () <= N2
context cFeature ,pFeature
inv :
  pFeature.subFeatures→includes(cFeature) and cFeature.allInstances→size () >= N3

```

3) Dominant Relation Constraint

Dominant relation constraint means that two features between non-“father-son” relationship exists somehow direct relation: alternative, depends, or excludes.

Definition 8. Alternative. There exist feature-a and feature-b. If they satisfy the constraint:

```

context feature-a ,feature-b inv alternative :
  DDPM.Feature→excludes(feature-a) implies
  DDPM.Feature→includes(feature-b) and
  DDPM.Feature→excludes(feature-b) implies
  DDPM.Feature→includes (feature-a)

```

We can conclude that feature-a and feature-b are alternative; remember it as `alternative(feature-a, feature-b)`.

Alternative makes it a rule that at least one feature must be chosen from the two features feature-a and feature-b, which not only reflects the indispensable but multiple parts in the field application, but also represents the different process and expression of the same feature in different applications.

Definition 9. Depends. If these two feature-a and feature-b meet the constraint:

Context feature-a, feature-b inv depends :
 DDPM.Feature→includes(feature-a) implies
 DDPM.Feature→includes(feature-b)

We can conclude that feature-a depends on feature-b; remember it as depends (feature-a, feature-b).

In DDPM, depends always exist between feature and element feature(description of feature).

Definition 10. Excludes. If these two feature-a and feature-b meet the constraint:

context feature-a, feature-b inv excludes :
 DDPM.Feature→includes (feature-a) implies DDPM.Feature→excludes(feature-b) and
 DDPM.Feature→includes (feature-b) implies DDPM.Feature→excludes(feature-a)

We can conclude that feature-a and feature-b are mutually exclusive; remember it as excludes (feature-a, feature-b).

Exclude constraint shows that two features are mutually exclusive and cannot exist in the same field application.

4) Recessive Relation Constraint

Recessive relation constraint is a kind of constraint that exists among features caused by the definition of digital data, the environment of digital data, and the internal relationship between essential features. For an example, in DDPM, createTime is smaller than visitTime and modifyTime, namely:

Context createTime,visitTime,modifyTime
 inv: visitTime >= createTime and modifyTime >= createTime

We can conclude the four classes of constraint between digital data features based on analysis of feature field method on combining digital data semantic features. These constraints are not only suitable for describing the relation of inherent features in DDPM, but also suitable for describing the relation of extended features in DDPM.

Next, horizontal and vertical analysis can be launched on practical cases through DDFM, the feature tree can be built, and the dependency and relevancy between features can be expressed by the related operations on OCL.

4 Practical Examples and Analysis on It

The digital data feature tree from Windows has been built.

First, analyze the feature of files which are picked up in Windows system. Besides the basic characters of DDFM, the digital data features also include documental basic

information, documental application environment, source of documents, and document information that depend on the file types as extended features. Fig.2 shows the WDDFM (Windows DDFM).

In Fig.2, name, type, size, location, createTime, visitTime, modifyTime, and content are the eight inherent features of digital data. These features are all the same in all kinds of systems.

BasicInfo is the basic information of digital data, including access rights (read only and hide), language, creator, owner, and user group. These features may be defined by creators or users who operate the system.

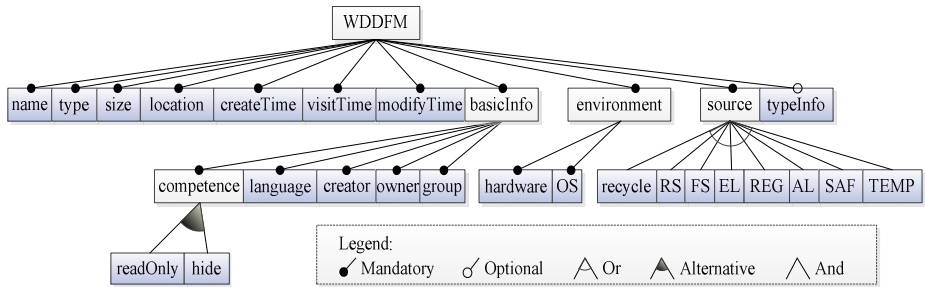


Fig. 2. WDDFM (Windows DDFM)

Environment is the environment feature of digital data, including hardware info, and operation system info. It depends on the system used, while Windows is the main operation system studied in this paper.

TypeInfo is the basic information of type’s digital data, which relies on the inherent character of type. If type is a pdf file, TypeInfo would describe the title, theme, author, keywords, creator, edition, pages, encryption or not, etc.

Source describes the source of digital data. The source of digital data from Windows can be divided into 8 parts below.

- 1) RS: Relax Space. The information of unrecoverable deleted files can be obtained from it.
- 2) FS: Free Space. The deleted files can be obtained from it, including damaged and not accessible clusters.
- 3) EL: Event Logs.
- 4) Reg: Registry.
- 5) AL: Application Logs that are not administrated by Windows time log server.
- 6) SAF: Special Application Files, such as the Internet history of Internet Explorer, and cache.
- 7) Temp: lots of temp files created by applications.
- 8) Recycle: Recycle bin.

On the basic of digital data feature tree, OCL operations can realize our related request.

Below are some practical examples.

Example 1:

context WDDFM

inv: WDDFM->select(plp.type='txt')->notEmpty()

The result is to find the subset whose digital data type is “txt”.

Example 2:

context BasicInfo

inv: BasicInfo->collect(plp.owner)

The result is a bag (it may has more than one same owner) derived from BasicInfo, which only has one feature of owner.

Example 3:

context environment

inv: environment->forAll(plp.OS='Windows')

As a result of the digital data feature models from Windows, the OS values should be Windows. The OCL sentences above require all the OS values in Set environment are Windows.

Example 4:

context WDDFM

inv: WDDFM->exists(plp.location='d:\hide\') and WDDFM->exists (p.modifyTime-p.createTime)

If the location value of digital data is ‘d:\hide\’, the return value is true, or it false. That is to explain whether there is digital data under d:\hide\. If exists, the second statement just judges whether this data has been modified. If the value is false, it has not been modified since created, which also means that this digital data is trusted.

5 Conclusions

This paper uses FODA to build a static semantic feature model of trusted digital data, and formally describe all kinds of features of digital data in OCL. It classified those features through “collection” and “collection operation”, and describes the relationship and dependency between them.

The work on next stage is to discuss the formalized verification method of trusted discovery and trusted fixation. By using OCLE (the environment of OCL) tools, the test and verification in grammar and semantic can be launched on the static attributes dependency of digital data, such as originality, completeness, and anti-tamper.

Acknowledgments. The authors would like to thank the reviewers for their detailed reviews and constructive comments, which have helped improve the quality of this paper. This paper is supported by the Chinese National Natural Science Foundation (No. 61073114), the Foundation of Nanjing University of Posts and Telecommunications (No. NY212059), and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- [1] Wang, S.J.: Measures of retaining digital evidence to prosecute computer-based cyber-crimes. *Computer Standards & Interfaces* 29(2), 216–223 (2007)
- [2] Lim, K.S., Savoldi, A., Lee, C., Lee, S.: On-the-spot digital investigation by means of LDFS: Live Data Forensic System. *Mathematical and Computer Modelling* 55(1-2), 223–240 (2012)
- [3] Kleve, P., Mulder, R.D., Noortwijk, K.: The definition of ICT Crime. *Computer Law & Security Review* 27(2), 162–167 (2011)
- [4] Ayers, D.: A second generation computer forensic analysis system. *Digital Investigation* 6(S), S34–S42 (2009)
- [5] Lee, S., Savoldi, A., Lim, K.S., Park, J.H., Lee, S.: A proposal for automating investigations in live forensics. *Computer Standards & Interfaces* 32(5-6), 246–255 (2010)
- [6] Sun, G.Z., Geng, W.M., Chen, D.W., Shen, T.: One validity model of digital forensics based on trusted probability. *Chinese Journal of Computers* 34(7), 1262–1274 (2011) (in Chinese)
- [7] Smith, M., Friese, T., Engel, M., Freisleben, B.: Countering security threats in service-oriented on-demand grid computing using sandboxing and trusted computing techniques. *Journal of Parallel and Distributed Computing* 66(9), 1189–1204 (2006)
- [8] Tao, Y.C., Jin, H., Wu, S., Shi, X.H.: Scalable DHT- and ontology-based information service for large-scale grids. *Future Generation Computer Systems* 26(5), 729–739 (2010)
- [9] Ding, Q.F.: A formalized model of digital data for trusted forensics. *Nanjing University of Posts & Telecommunications* (2011) (in Chinese)
- [10] Stoica, P., Sandgren, N.: Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches. *Digital Signal Processing* 16(6), 712–734 (2006)
- [11] Wöllmer, M., Al-Hames, M., Eyben, F., Schuller, B., Rigoll, G.: A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing* 73(1-3), 366–380 (2009)
- [12] Zhou, S.Q., Chin, K.S., Xie, Y.B., Yarlagadda, P.K.: Internet-based distributive knowledge integrated system for product design. *Computers in Industry* 50(2), 195–205 (2003)
- [13] Li, C.K., Wong, D.S.: Signcryption from randomness recoverable public key encryption. *Information Sciences* 180(4), 549–559 (2010)
- [14] Kenneally, E.E., Brown, C.L.: Risk sensitive digital evidence collection. *Digital Investigation* 2(2), 101–119 (2005)
- [15] Schobbens, P.Y., Heymans, P., Trigaux, J.C., Bontemps, Y.: Generic semantics of feature diagrams. *Computer Networks* 51(2), 456–479 (2007)
- [16] Jin, X.L., Ma, H.D.: A Semantic Description Model of Features for Service Component Using OCL. *Journal of Computer Research and Development* 44(12), 2112–2121 (2007) (in Chinese)
- [17] Object Constraint Language (EB/OL), <http://www.omg.org/spec/OCL/2.0/PDF>
- [18] Marković, S., Baar, T.: Refactoring OCL annotated UML class diagrams. *Software and Systems Modeling* 7(1), 25–47 (2008)
- [19] Jung, C.S., Seo, H., Kang, H.G.: Estimating redundancy information of selected features in multi-dimensional pattern classification. *Pattern Recognition Letters* 32(4), 590–596 (2011)
- [20] Reinhartz-Berger, I., Sturm, A.: Utilizing domain models for application design and validation. *Information and Software Technology* 51(8), 1275–1289 (2009)

A Novel Hybrid Cellular Automata Based Cipher System for Internet of Things

Mouza Ahmed Bani Shemali, Chan Yeob Yeun,
Mohamed Jamal Zemerly, and Khalid Mubarak

Khalifa University,
Electrical and Computer Engineering Department,
P.O. Box 573, Sharjah, UAE
{mouza, cyeun, jamal, mubarak}@kustar.ac.ae

Abstract. High volume products for Internet of Things have a constricted cost that limit the cryptographic implementation in such devices. Thus, there is a need to design a lightweight cryptography stream cipher that can fit such low computation devices with low cost. This paper proposes a lightweight stream cipher that can enhance Linear Feedback Shift Register (LFSR), Feed Carry Shift Register (FCSR) within the Shrinking Generator stream cipher and used Cellular Automata (CA) to update the cipher stream.

Keywords: IoT, LFSR, FCSR, CA, Cipher, SG, SSG.

1 Introduction

Stream ciphers design has been abandoned for several years but recently increasing improvements of the low computation devices such as IoT that in need for a lightweight cipher design encouraged recent activities in this area. The stream cipher usually consists of three phases which are initialization, main phase to produce the key, and finally update phase.

This paper proposes a novel stream ciphers with lightweight properties. The aim of the proposed stream cipher is easy to implement and hard to cryptanalyze. Thus the main building block of our proposed stream cipher is simple as the simplicity of the LFSR and FCSR. Where, the update function depends on the Cellular Automata (CA) mathematical model.

This paper is organized as follow. Section 2 introduces LFSR, FCSR and the main block of our stream cipher which are the Shrinking Generator (SG) family and the Cellular Automata (CA). Section 3 comparing between the SG family to choose the best main block among them. In Section 4, our proposed stream cipher is proposed. Section 5 shows the security analysis of the proposed stream cipher. Finally, section 6 concludes this paper.

2 LFSR and FCSR

Designers would prefer register such as LFSR [1, 2] or FCSR [3] as a main block of their design in designing the Pseudo Random Numbers (PRN) because it could be fast, simple and is associated with algebraic structures that make it easy to analysis. The main feature within the FCSR is the addition of the carry register that provide non-linearity properties and increase the linear complexity. Thus, it increases the difficulty in finding the initial state of the register. That feature adding more computation time for the cryptanalysis in finding the carry register value and finding the initial state of the register. Thus, in cryptographic view of point analyzing FCSR is much harder than LFSR.

Besides the output of the LFSR is linear that is easy to break while the output of the FCSR is quadratic. The quadratic output of the FCSR is more resistance to the algebraic attack and the correlation attacks than the LFSR output since it is updated nonlinearly.

On the other hand, the implementation of the LFSR is lighter and less in cost than the FCSR since it needs to add a carry register memory. Nerveless, using LFSR or FCSR as main block for generating pseudorandom number cannot be impalement without any enhancement since the output can be predicted using Berlekamp-Massey algorithm. The register in the FCSR and LFSR is clock in regular manner which lead to ability to predict it easily. As a result research find out that irregular clocking of the register can make the cryptanalysis harder. Shrinking Generator (SG) is an algorithm that used irregular clocking on the register.

The SG algorithm used two register which are register A for the data and register S for the sampling with clock. Then a condition is applied at each clock in order to shrinking the output of the register S. Each time two bits is clock together one from A and the other one from the S. If the output of S is 1 then the output stream cipher is A. However, if the output of S is zero then the output of A is discarded and another clock is applied. These steps repeated till it reaches to the desirable key length. Thing to notice from the SG algorithm is that the attacker will find difficulty in finding the gap between the output and the discarded values of register A. Also, the algorithm called Shrink because the output of register A is shrinking due to the discarding value that applied when the value of register S is zero [4].

Under the SG there is another algorithm that called Self Shrinking Generator (SSG) [5] that uses the same idea of the SG. However, SSG algorithm used only one register instead of two register. Thus the rule to produce the output stream here is a lite bit different. Each time two bits of the register is clocked if the output bit is 01 then the output is zero. If the output is 11 then the output is one. Other than that the output is discarding.

There is another part of SG family which is the cascade algorithms. The cascade algorithm is a combination of the two generators ideas with the combinations of LFSR and FCSR. The cascade algorithm can enhance the period length of the stream and the statistical properties of random sequences in [6]. Fig. 1 explains four cascade cipher streams.

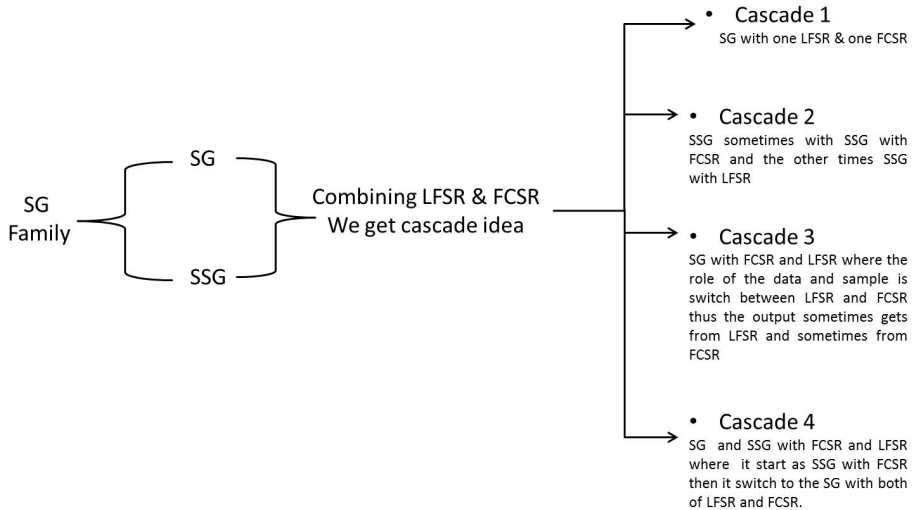


Fig. 1. SG family summaries

SG family algorithms can produce PNG with good properties that what is shown within the Statistical Test (ST) that we applied in section 3. Thus after applying ST one algorithm can be used as a main block of our cipher.

Our stream cipher is combining both of the Shrinking Generator (SG) and the Cellular Automata (CA) in order to provide a stream cipher that is hard to predict. CA are discrete mathematical models of finite state machine that can be arrange in lattice way. Each state called cells that can update itself synchronously according to local rules at each discrete time step. The concept of CA was first introduced by Stanislaw Ulam and John von Neumann to Facilitates the study of the biological processes such as self-production by converting it to a mathematical model [7]. An example of the used of CA is the game of life that is invited by John Conway, the mathematician [8]. Game of life considers as an example of self-organizing systems.

The idea of using CA in cryptography was first suggested by Stephen Wolfram at 1985[9]. The paper is based on studies of having random behavior given a random input in the physical system. Wolfram suggests a simple one-dimensional CA that arranges the N input stream in a lattice arrangement. Each N bit is considered as a cell that can be updated in discrete time steps depending on the neighboring cells and according to the following rule:

$$a'_i = a_{i-1} XOR (a_i OR a_{i+1})$$

The output of this rule is a random sequence that can be used as a key and can be used to encipher the plaintext P as follow:

$$C_i = P_i XOR a^i$$

In our proposal CA is used to update the key in the last step. Next section compares between the SG family algorithms in order to find the most suitable main block for our proposed stream cipher.

3 Comparing between the SG Family Algorithms

The main properties that can firms the pseudorandom binary sequences are mainly having some statistical properties such as low cross correlation values and a good distribution between zeros and ones. Thus, in order to compare between the ciphers algorithms of the SG family a comparison is done within the three properties [10]. The first property is the distribution between zero and one if it is equal or not. Second one is the run for the stream cipher the number of $\frac{1}{2}$ run have to be of length 1, $\frac{1}{4}$ of the run have to be of length 2, $\frac{1}{8}$ of the run have to be of length 3 and so on. Where, in each case the distribution between zero and one need to be equal. For example, the following stream 01001011010100 is 14-bits in length and contains eleven runs: 0 1 00 1 0 11 0 1 0 1 00. The expected number of runs within stream cipher of length i is $n/2^i$, where n is the sum of both zeros and ones runs. Finally autocorrelation test is applied to find if there any repeated pattern with the stream cipher. Thus, if the autocorrelation value is zero then that means that the pattern is random. Actually, it is so difficult to find stream cipher with zero autocorrelation. If the correlation value one then that means that the pattern is the same. Where the negative value means that the pattern is the same however it is shifted 180 degree. To calculated the autocorrelation excel program is used.

All of the SG family stream ciphers are implementing using JAVA code. Also, the algorithm that can test the previous mention statistical test is implementing in JAVA. For each stream 5 samples are taken to find in the final the average of these five samples for each stream cipher. A table 3 that is presented in section 5 illustrates the results of the 3 tests on the SG family ciphers with our proposed stream cipher. In order to find out the best solution each test is ordered from the best to the worst depending on the nearest to the ideal value in each test. The weight is given from 10 to 3. The 10 value is given to the best result and 3 value is given to the worse result. Thus, Table 3 shows also, the result weight values to the discussed stream ciphers. The best solution is the cascade 1 that gets the highest weight among the rest ciphers. Accordingly next section proposed our stream cipher using cascade 1 as a main block.

4 The Proposed Stream Cipher

The proposed stream cipher that is addressed in this paper is a major enhancement to its predecessor in [11]. This paper adds the update function to the stream cipher which consider as an important element within the stream cipher design. Based on the statistical test analysis the proposed cipher uses the cascade 1 as the main building block. The cipher stream works through three main stages which are initialization stage, output key stage and update stage. Each stage is explained in more detailed below:

Stage 1. Initialization Step

The first step is the initiation step that uses a key and an Initial Vector (IV) to produce an output stream. Both of the key and the IV have 128 bits in length. This step

produces bits and updates the registers of the LFSR and FCSR each with 128 bits without outputting all the bits fed back to the registers. Thus, after reaching 256 bits in the state this step is finished.

Stage 2. Produces the Secret Key

After the Initialization step, the generator can be used to produce a stream cipher. The steps to generate the stream are as follow:

- Both of the FCSR and the LFSR are clocked together
- If the output of the FCSR is 1 then the output of the state is the output of the LFSR at that state.
- If the output of the FCSR is 0 then the output stream of the LFSR at that state is discarded.
- The output stream cipher is the sum of the state and the carry value modulo 2.
- The sum value is divided by 2 to update the value of the carry value.

The generator generates a stream that can be used as a key that can be XORed with the message.

Stage 3. Update the Key

The Update Stage used the CA as a basis that is considered as a non-linear filter. However, to choose the suitable rule we applied CA rule with more than one dimension depending on the cell locations. In order to apply our rules first the 128 bits are divided on to 16 x 8 matrices. Thus our rule has 8 rows with 16 cells in each row and either two or three dimensions. Row one and 8 and the first and the last cell in each row have three neighbor cells where the rest which start from row 2 to row 7 and from cells 2 to 15 have four neighbors cells. Thus the rule is divided into rule for 3 cell neighbors or for 4 cell neighbors. The cells with 3 neighbors use 2 dimensional rule that uses its row and either the above or the below row. For example row one has 3 cell neighbors and uses its row and the row which is below row 2 to update its cells. Also row 8 has three neighbors cells and uses its row which is above row 7 to update its cells. Where the rest update their cells using three dimensional row depending on the above and the below rows. In order to choose the best rule all of the elementary CA rules are applied and changes to the rules are also applied to reach the rule that can provide a balance distribution between the zeros and ones with our stream cipher. The rule can update the key each time randomly and can produce a new key each time. The rule is as follow:

Each cell checks the surrounded cells and applies the following rules:

- If cell surrounded by 3 cells if one =2 or if zero >=3 then cell =1
- If cell surrounded by 4 cells if one 2 or 3 or if zero >=4 then cell=1
- Else cell = 0

For example if the input key is

010010011010111001101101111010110011001101011110001110000111100111
 01100110100100110101110011011011110101100110110101111000111000

The input will be organized within a matrix of 8×16 as seen in Table 1.

Table 1. Our rule demonstration

0	1	0	0	1	0	0	1	1	0	1	0	1	1	1	0
0	1	1	0	1	1	0	1	1	1	1	0	1	0	1	1
0	0	1	1	0	0	1	1	0	1	0	1	1	1	1	0
0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	1
1	1	0	1	1	0	0	1	1	0	1	0	0	1	0	0
1	1	0	1	0	1	1	1	0	0	1	1	0	1	1	0
1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	1
0	1	0	1	1	1	1	0	0	0	1	1	1	0	0	0

Then, the rows are shuffled in random ways. Thus the output stream will be as follow:

11011001111011000010000101001011100100000010100001011010
 010100011010110100100011110101100100111101010000001100111
 01010000110110

The proposed hybrid CA based stream cipher is implemented using Java code and ST is applied on the output stream. The result of the ST can be shown in Table 3 section 5. Next section provides security analysis of the proposed hybrid CA based stream cipher.

5 Security Analysis of Our Proposed Cipher

There are several attacks that stream ciphers usually suffer from such as Time Memory Trade-offs (TMTO), Correlation attack and Algebraic attack. TMTO [12] pre-computes a large amount of data in order to lower the computation complexity of the cipher. For this attack the attackers pre-compute as many different states as they can with their corresponding key stream then sort and store these pre-computed data. Then, they obtain a key stream and compare it with the pre-computed data looking for the collision. If the collision is found then the secret state can be recovered. To avoid TMTO attack the state size of the designed cipher should be at least twice the key size, and the IV size should be at least as large as the key size. Our proposed stream cipher used initial state with 256 bits size which is larger than the size of the key and the IV (both of them are 128 bits). Beside the size of the key and the IV is the same. Thus, a collision between the states cannot be accrued leading to exploiting the secret state and recovering it. For the correlation attack [13] it is applied when there is a correlation between the key stream and the output sequences. The key stream k is

correlated with the output sequences z if the probability between $k = z$ is equal .To avoid this attack a redundancies in the stream cipher need to be reduced by updating each state bit frequently.

Thus, in our proposed stream cipher AC is used to update all the state of the register simultaneously. Other than the correlation there is a fast correlation attack that used much faster algorithm than the exhaustive search over all initial states of the generator register by means of applying parity equations produced from the feedback polynomial. Fast correlation attack is proven to be infeasible against long LFSRs if they have larger than ten taps. Consequently, the number of taps in our proposed stream is 66 taps (the polynomial equation can be found in [11]) that allow us to avoid the fast correlation attack and to have carry register (cr) up to 6 digits. Having cr value as explains before can increase the linear complexity of our stream cipher.

Finally there is algebraic attack [14] on stream cipher which can recover the secret key by analyzing the multivariate relation between the key and the output. Thus the attacker uses his/her knowledge of some output stream and combines it to find the initial state of the stream cipher. In order to avoid such attack the register length L needs to be large thus the attacker needs to have $2L$ bits in order to be able to compute the key. Accordingly, the update function in our proposed stream cipher update the register content each 200 bits. The CA updated function is difficult to be reverse by attacker thus that makes the stream cipher more resistant to algebraic attacks. In addition, ST applied on the proposed stream cipher to prove it is randomness. The result can be seen in Table 3. Combing both of the SG and CA provide a PRN that is difficult to crack. Also, combing FCSR and LFSR provide longer period and a stream cipher with much complex linearity. All these factors work together in order to provide a secure key that can be used for cryptographic purpose. The main feature within the design is that it consists of sample element that proved to be easy to implement and hard to crack. As a result, our aim in providing simple cipher stream easy to implement hard to crack is accomplish here.

Table 2. Comparison of statistical tests against the proposed stream cipher

	L-SG	W	L-SSG	W	F-SG	W	F-SSG	W	CAS1	W	CAS2	W	CAS3	W	CAS 4	W	Pro-posed	W
Dist	2.594	6	2.958	4	1.258	8	2.820	5	1.595	7	7.070	2	4.708	3	0.419	10	1.095	9
Run	5.4E-13	3	6.0E-1	2	5.0E-16	6	5.1E-1	5	4.1E-1	10	4.9E-1	8	4.8E-19	8	5.1E-1	4	4.9E-1	7
Auto	2E+15	5	2E+14	8	2E+16	2	-2E+13	10	-3E+15	7	7E+15	3	6E+14	6	4E+15	4	-1E+14	9
W	14		14		12		20		24		13		18		18		25	

Dist=Distribution Test, AutoC=Autocorrelation Test, W=Weight, L=LFSR, F=FCSR, CAS=Cascade.

6 Conclusion

This paper proposed novel SG CA based stream cipher that intend to be used in constrain environment such as IoT. The proposed stream cipher combined both of the LFSR and FCSR with SG algorithm and used AC as an updated function.

The combination of these elements is to ensure that the stream cipher can provide a secure cryptographic key that is hard to crack. Thus, ST is done on the stream cipher to prove its randomness and that it can be used as a cryptographic key.

References

1. Alfke, P.: Efficient Shift Registers, LFSR Counters, and Long Pseudo- Random Sequence Generators, Xilinx. Tech. Rep., Version 1.1 (1996)
2. Koeter, J.: What's an LFSR? Texas Instruments Incorporated (December 1996)
3. Klapper, A., Goresky, M.: 2-adic shift registers. In: Anderson, R. (ed.) FSE 1993. LNCS, vol. 809, pp. 174–178. Springer, Heidelberg (1994)
4. Coppersmith, D., Krawczyk, H., Mansour, Y.: The shrinking generator. In: Stinson, D.R. (ed.) Advances in Cryptology - CRYPTO 1993. LNCS, vol. 773, pp. 22–39. Springer, Heidelberg (1994)
5. Meier, W., Staffelbach, O.: The self-shrinking generator. In: De Santis, A. (ed.) EUROCRYPT 1994. LNCS, vol. 950, pp. 205–214. Springer, Heidelberg (1995)
6. Couture, R., L'Ecuyer, P., Tezuka, S.: On the distribution of k-dimensional vectors for simple and combined Tausworthe sequences. *Math-132 L'Ecuyerematics of Computation* 60(202), 749–761 (1993)
7. Von Neumann, J.: Theory of Self-Reproducing Automata. Univ. of Illinois Press, Urbana (1966)
8. Hjelle, G.A.: Conway's Game of Life. *Math. Circle* (February 22, 2009)
9. Wolfram, S.: Cryptography with cellular automata. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 429–432. Springer, Heidelberg (1986)
10. Mitra, A.: On Pseudo-Random and Orthogonal Binary Spreading Sequences. *International Journal of Information and Communication Engineering*, 447–454 (2008)
11. Shemali, M.A.B., Yeun, C.Y., Zemerly, M.J., Mubarak, K.: A New Lightweight Hybrid Cryptographic Algorithm for the Internet of Things. In: Proc. ICITST 2012, London, December 10–12, pp. 87–92 (2012)
12. Biryukov, A., Shamir, A.: Cryptanalytic time/memory/data trades-of for stream ciphers. In: Okamoto, T. (ed.) ASIACRYPT 2000. LNCS, vol. 1976, pp. 1–13. Springer, Heidelberg (2000)
13. Penzhorn, W.T.: Correlation Attacks on Stream Ciphers: Computing Low-Weight Parity Checks Based on Error-Correcting Codes. In: Gollmann, D. (ed.) FSE 1996. LNCS, vol. 1039, pp. 159–172. Springer, Heidelberg (1996)
14. Courtois, N.T., Meier, W.: Algebraic Attacks on Stream Ciphers with Linear Feedback. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 345–359. Springer, Heidelberg (2003)

An Efficient Computer Forensics Selective Imaging Model

Waleed Halboob^{1,3}, Khaled S. Alghathbar^{1,2}, Ramlan Mahmood³,
Nur Izura Udzir³, Mohd. Taufik Abdullah³, and Ali Deghantanha³

¹ Center of Excellence in Information Assurance,
King Saud University, Riyadh, Saudi Arabia

² Departments of Information Systems,
Collage of Computer and Information Sciences,
King Saud University, Saudi Arabia

³ Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
{wmohammed.c, kalghathbar}@ksu.edu.sa,
{ramlan, izura, taufik, alid}@fsktm.upm.edu.my

Abstract. Selective imaging is a new concept in computer forensics. It is used for collecting only the data that is relevant to the crime and helps in improves the scalability of the investigation process. However, the current selective imaging approaches directly image the identified data without considering their offsets on the targeted user storage. This paper investigates the impact of the relevant data offsets on the efficiency of the selective imaging process. A practical selective imaging model is presented which includes a digital evidence ordering algorithm (DEOA) for ordering the selected relevant data items. The proposed selective imaging model has been implemented and evaluated in different types of storage devices. The evaluation result shows that even if our proposed algorithm has a small efficiency negative impact before the imaging process starts; it has a large positive effect on the efficiency of the selective imaging process itself.

Keywords: Computer forensics, digital evidence, selective imaging, efficiency, ordering algorithm.

1 Introduction

Computer forensics is a computer security discipline focused on *identifying, collecting, preserving, analyzing* and *presenting* digital evidence on digital systems so that the presented digital evidence is acceptable in a court of law. The traditional procedure for collecting the digital evidence is making a bit-by-bit physical image (called a raw or 'dd' image) from the user storage device and later on - at the computer forensics lab - the image is analyzed. However, this procedure is not scalable any more as the increased amount of user data and storage size will increase the required imaging and analyzing costs (resources and time). As a solution, selective imaging concept is introduced [1].

The idea behind the selective imaging concept is to collect only pre-selected relevant data. A pre-imaging analysis is used for selecting first the data items that seem to be relevant to the crime. According to Turner [2], the data items can be selected through *manual*, *semi-automatic* and *fully automatic* selections. Using the *manual selection* method, an investigator manually selects the relevant data items from, for example, a folder tree. With the *semi-automatic selection* method, the relevant data items are selected using tools enabled with search engines. Finally, the *fully automatic selection* uses intelligent methods for deciding which data items are relevant and according to some parameters given by the investigator. Several selective imaging methods have been proposed such as *risk sensitive digital evidence collection* [1] and *digital evidence bags* [2, 3, 4, 5, 6, and 7]. However, these approaches image the selected data items directly without considering their offsets (physical address or position) on the user storage.

This paper studies the impact of the digital evidence offset on the efficiency of the selective imaging process. A practical selective imaging model is proposed which includes a *digital evidence ordering algorithm (DEOA)*. The DEOA orders the relevant evidence according to their offsets. The proposed imaging model has been implemented and evaluated with several cases in different storage types.

The structure of this paper is as follows. Section 2 presents a brief review of the literature on selective imaging in computer forensics. Section 3 presents the proposed selective imaging model and its implementation. Section 4 presents the discussion and critical evaluation of the proposed model. Finally, Section 5 concludes the paper with future directions to enhance the proposed work.

2 Related Works

The rapid increase in the size of user storage and data makes a scalability challenge to the forensics investigation process. As a result, the cost (in term of the required time and resources) of making a bit-by-bit image (raw or 'dd' image) from the user storage device is increased from time to time. As a result, researchers have tried to address this issue first by reducing the required amount of storage space using what is called *block based compression* [8, 9] to the data stream. Another solution is *hash-based disk imaging* [10] in which the amount of the collected data is reduced using data de-duplication and reduction technologies. Other possible solutions consider the cost of both the required time and storage. These solutions use *effective and efficient analytical* and *selective imaging* [11]. The *effective and efficient analytical* concept is applied to the digital evidence analysis step. In other words, a bit-by-bit image is made from the whole user storage device during evidence collection and then the collected image of data is analyzed selectively or in a distributed manner. Researchers using the *effective and efficient analytical* concept have applied *distributed evidence analysis* [6], *data mining search process* [12], *file classification* [13], and *clustering text-based search* [14]. However, reviewing these effective and efficient analytical approaches is outside the scope of this research paper.

The idea behind selective imaging concept [2, 4, 5, 7, and 15] is to image or collect only relevant data to a crime, instead of making a physical bit-by-bit image from whole user storage device. Researchers on selective imaging concept have proposed several methods such as *risk sensitive digital evidence collection* [1] and *digital evidence bags* [2, 4, 5, 7, and 15]. The *risk sensitive digital evidence collection* method is introduced by Kenneally and Brown [1]. They discussed the benefits and costs of using both a bit-by-bit imaging and selective imaging. They also proposed their method on how selective evidence can be collected during the digital evidence collection phase in dead and live systems without compromising forensics standards.

Turner [2, 4, 5, and 7] proposed the concept of the *digital evidence bags* (DEB). The DEB is a new forensics format for saving the collected digital evidence along with their metadata. The DEB has three files: .tag, .index and .bad. The first file is used for storing some metadata (e.g., timestamps, investigator, location) about the investigated case. The .index file also stores some metadata but about each collected evidence. Finally, the .tag file stores the exact data (raw data bit stream) of the collected evidence.

As discussed above, several research efforts have been directed to resolve the problem of voluminous scalability, but the implications of the selected data items offsets on the efficiency of the selective imaging process have not been studied.

3 The Proposed Selective Imaging Model

To study the impact of the relevant data offsets on the imaging efficiency, the proposed process includes a *digital evidence ordering algorithm* (DEOA) that orders the relevant data items according to their offsets or physical positions. However, the proposed process includes three steps (Figure 1): *digital evidence selection*, *digital evidence ordering*, and *digital evidence selective imaging*. Due to the limited space provided here, the comprehensive design of the selective imaging model is leaved as our extended work.

3.1 Digital Evidence Selection

This step is used for selecting the relevant data items to the crime. In this research, we assume the use of the semi-automatic method as this method is the most practical solution as discussed in Section 1. An investigator runs a forensic recovery tool (or tools) from an external machine to scan and analyze the targeted storage device in a read-only mode. The targeted device must be analyzed first to determine all existing data even active or deleted. The targeted device is accessed in a read-only mode to ensure that the content of the targeted device will not be altered. After the storage has been scanned and analyzed, the investigator searches for relevant evidence and saves the search results in a report. For this purpose, several computer forensics tools - such as Winhex (X-Ways), FTK Access Data, and CnWRecovery – have been tested and used. These tools enable the investigator to search for and select the relevant data items using different methods which can be categorized as manual and semi-automatic

selection methods. In addition, the investigator can report the search result in a standard file format files mostly in a *Common Separated Values* (csv) file format report. As shown in Figure 1, the output of this step is several .csv files for different search tries in which, each .csv file contains several hits for one search result. These .csv reports contain metadata (file name, path, size, etc.) of all found data.

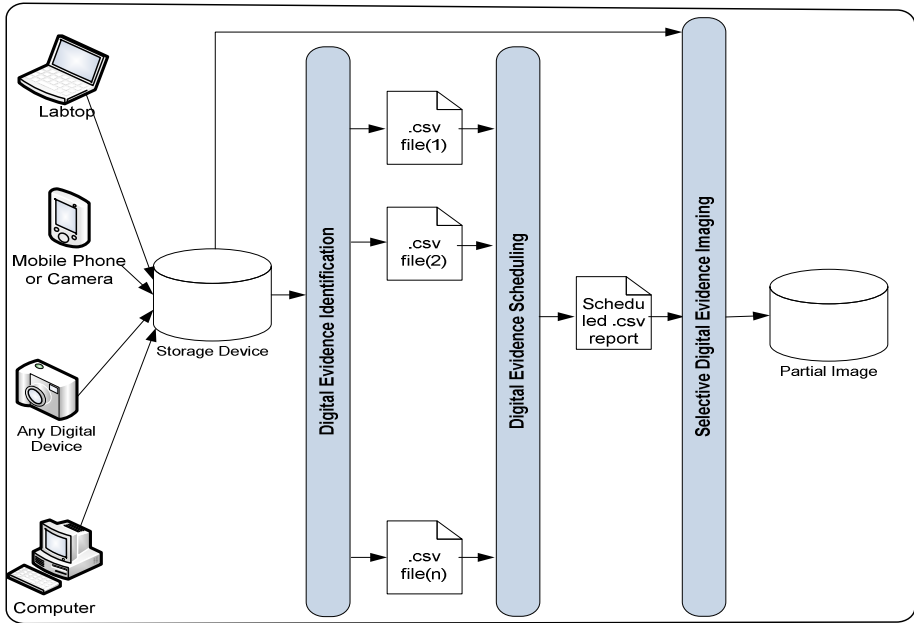


Fig. 1. The architecture of the proposed selective imaging model

3.2 Digital Evidence Ordering

After selecting the relevant data into .csv files, the search results (found inside the .csv files) are passed to the digital evidence ordering algorithm (DEOA) which simply first merges the .csv files together into one .csv file called *scheduled .csv report*. During the merging process, the data items that share the same name and physical address are filtered out. Then, the final data items found inside the *scheduled .csv report* are ordered according to their offset inside the targeted user storage.

3.3 Digital Evidence Selective Imaging

The evidence imaging step, is used for imaging all relevant data items to a partial forensic image. The partial forensic image contains all relevant data items along with their metadata. For each data item that must be imaged, the data item is read from the user's storage device, hashed and then the data item with its metadata are written into the partial image. Where, the non-relevant data items are already filtered out during

the digital evidence ordering step. In terms of the partial forensic image used, the selective imaging process requires a forensic image that supports at least two futures which are: i) multi objects streams as each data item is selectively imaged and, as a result, needs to be stored inside an image as a separated object stream; and ii) storing metadata of each object stream. Several existing forensics image formats have been proposed and used such as RAW/DD, SGZIP, EO1, AFF3 and AFF4. The proposed model uses the AFF4 since this format supports the required futures [3]. Finally, the integrity of the collected partial AFF4 image is ensured by comparing the hash value of the imaged data items inside the image with the hash value generated during reading data items from the data storage device. If not matched, then the whole process has to be re-executed. If matched, the hash value – called *imagehash*- of the generated partial is saved in a separate file.

3.4 Implementation

The proposed selective imaging model has been implemented using Java programming language on NetBeans IDE 6.9.1. Some additional java application programming interfaces (APIs) have also been used. The *Javacsv2.1* free java package is used for processing the .csv files. The Advance Forensic File Format (AFF4) java package – called *Truezip1.6.1*- is used for creating the AFF4 partial image and writing selected data items into it. The message digest 5 (MD5) is used for hashing the data items. Figure 4 shows a screenshot from the implemented prototype.

4 Result and Discussion

The performance of the proposed model is measured to evaluate its efficiency. Two types of computer storage are used namely hard disk and flash drive. To run the experimental study, Enron dataset is used. Due to the limited capacity of the flash drives, only 50,000 files from the Enron dataset are used. The exact size of the used Enron files is 124 MB. These files are grouped into 10 groups. Each group contains 5000 files and copied into the hard disk and flash drives with other random data until the hard disk and flash drives are occupied. Our experiment is executed on Microsoft Windows XP, Dell Computers with Intel Core Quad CPU (4 CPUs), 2.83GHz speed, and 4096MB memory.

Then, the performance of the imaging process is measured in ordered and normal cases. With the ordered case the *digital evidence ordering algorithm* (DEOA) is applied first to the relevant data items before imaging. These two cases are used to measure the implication of the relevant data offsets on the imaging process efficiency. Two imaging methods are considered. First when the data items are directly imaged to the destination storage file by file and second when the *Advance Forensic File Format* (AFF4) is used as destination storage. The cost of reading/writing and reading only the relevant data is measured. This is to clearly identify where the impact of the relevant data offset is. In both selective imaging methods the metadata and hash value of the imaged data are considered.

4.1 Ordering Algorithm Performance

Here, the negative impact of the ordering algorithm must be less than its positive impact on the imaging process. However, ten csv. files are used here where each file contains metadata of about 5000 data items from Enron dataset. So 50000 files are used and only about 820 of which are duplicated. In anyway, our experiment study shows that the cost of the ordering algorithm is only about 5.102 seconds.

Figure 2 shows the performance result of the normal and order imaging in a hard disk and flash devices when the data items are directly imaged (not to AFF4 image). It is clear from the result that the order imaging is more efficient than the normal imaging in both cases (reading/writing and reading only). For example, with the hard disk device the cost of reading/writing 50,000 files normally requires about 272.2243 seconds while with the ordering imaging requires only about 226.7867 seconds. As a result, the cost is reduced by about 17%. In term of the reading only the cost is also reduced by 17 % also and from 140.5423 seconds to 116.1765 seconds. The performance result of flash drive is illustrated in Figure 2. The relevant data inside the flash drive is imaged into hard drive storage. The cost of normal imaging is reduced by 29% (from 419.958 to 325.3977 seconds) when the ordering algorithm is used in read/write imaging. While with the reading only, the cost is reduced by about 4% only and from 188.374 to 181.279 seconds.

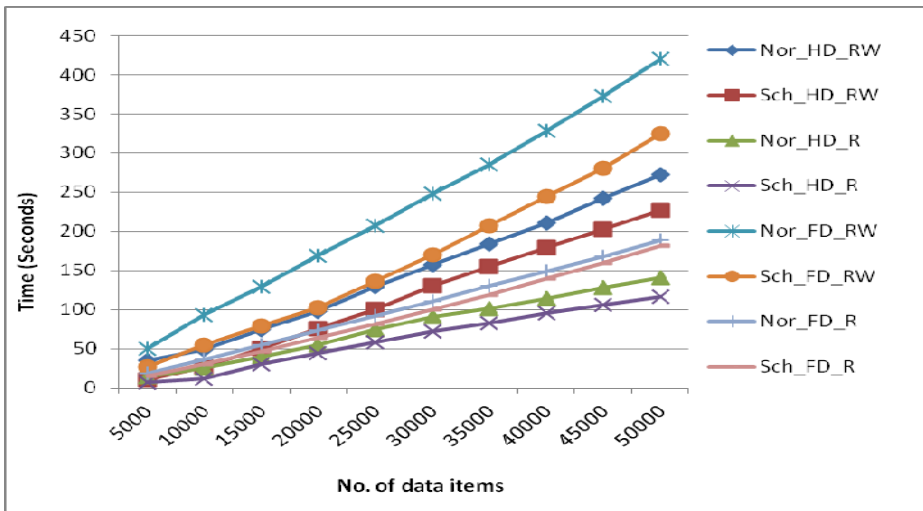


Fig. 2. The performance result of the hard disk and flash devices in a direct selective imaging

The performance result of imaging the relevant data into an AFF4 partial image from hard and flash drive is shown in Figure 3, respectively. The normal and ordered imaging cases are considered to study the impact of relevant data offset on the imaging efficiency. With the hard drive storage, the cost of imaging is reduced by 23% (from 189.349 to 144.66 seconds) when the ordering algorithm is used. With the flash drive, the cost is reduced by only 4% (from 218.3715 to 208.486 seconds).

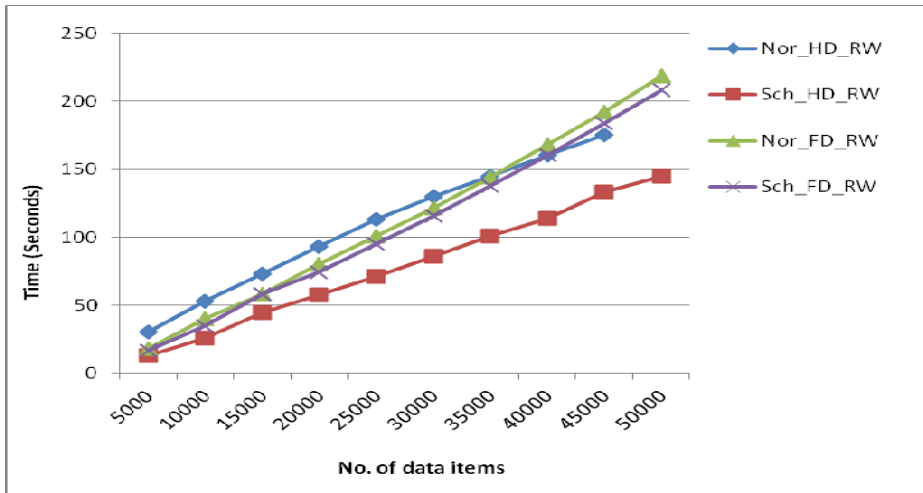


Fig. 3. The performance of the hard disk and flash drive in AFF4 partial image

5 Conclusion and Future Works

This paper proposes an efficient computer forensics selective imaging model that includes an ordering algorithm to study the impact of the relevant data offsets on the efficiency of the imaging process. The proposed model is implemented and its performance is measured. The result shows that the ordering algorithm has a small negative impact on the imaging process. This impact is the cost of merging and ordering the relevant data items. However, it has a good impact on the efficiency of the whole imaging process either when the relevant data are imaged directly (to another storage) or to a standard AFF4 forensics image format. Our future work is to present a detailed design and implementation of the proposed model which are not presented here because of the limited provided space. Also, investigating the efficiency of digital evidence analysis is our next target.

References

1. Kenneally, E.E., Brown, C.L.T.: Risk sensitive digital evidence collection. *Digital Investigation* 2(2), 101–119 (2005)
2. Turner, P.: Selective and intelligent imaging using digital evidence bags. *Digital Investigation* 3(1), 559–564 (2006)
3. Stüttgen, J.: *Selective Imaging: Creating Efficient Forensic Images by Selecting Content*. First. Mannheim University (2011)
4. Turner, P.: Digital provenance - interpretation, verification and corroboration. *Digital Investigation* 2(1), 45–49 (2005)
5. Turner, P.: Unification of digital evidence from disparate sources (Digital Evidence Bags). *Digital Investigation* 2(3), 223–228 (2005)

6. Richard, G., Roussev, V.: Breaking the performance wall: The case for distributed digital forensics. Paper presented at the Proceedings of the 2004 Digital Forensics Research Workshop (DFRWS 2004), Baltimore, Maryland (2004)
7. Turner, P.: Applying a forensic approach to incident response, network investigation and system administration using Digital Evidence Bags. *Digital Investigation* 4(1), 30–35 (2007)
8. Kloet, B., Metz, J., Mora, R.-J., Loveall, D., Schreiber, D.: libewf: project info. (2008), <http://www.uitwisselplatform.nl/projects/libewf/>
9. Garfinkel, S., Malan, D.J., Dubec, K.-A., Stevens, C.C., Pham, C.: Disk imaging with the advanced forensic format, library and tools. In: *Research Advances in Digital Forensics (Second Annual IFIP WG 11.9 International Conference on Digital Forensics)*. Springer (January 2006)
10. Cohen, M., Schatz, B.: Hash based disk imaging using AFF4. *Digital Investigation* 7, 121–128 (2010)
11. Beebe, N.: Digital Forensics Research: The Bad, The Good and the Unaddressed. In: *Advances in Digital Forensics V - IFIP International Conference on Digital Forensics*, Orlando, Florida, USA, pp. 17–36 (2009)
12. Beebe, N., Clark, J.: Dealing with Terabyte Data Sets in Digital Investigations. In: Pollitt, M., Sheno, S. (eds.) *Advances in Digital Forensics V. IFIP*, vol. 194, pp. 3–16. Springer, Heidelberg (2005)
13. Sanderson, P.: Mass image classification. *Digital Investigation* 3(4), 190–195 (2006)
14. Beebe, N.L., Clark, J.G.: Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results. *Digital Investigation* 4(1), 49–54 (2007)
15. Richard, G., Roussev, V.: File System Support for Digital Evidence Bags. In: Olivier, M., Sheno, S. (eds.) *International Federation for Information Processing. IFIP AICT*, vol. 222, pp. 29–40. Springer, Boston (2006)

Cloud Computing Risk Assessment: A Systematic Literature Review

Rabia Latif¹, Haider Abbas^{1,2}, Saïd Assar³, and Qasim Ali¹

¹ National University of Sciences & Technology, Islamabad, Pakistan
rabiya_128@yahoo.com

² Centre of Excellence in Information Assurance (COEIA),
King Saud University, Riyadh, Saudi Arabia
hsiddiqui@ksu.edu.sa, haidera@kth.se

³ Institut Mines-Télécom, Telecom Ecole de Management
Information System Departement, France
said.assar@it-sudparis.eu

Abstract. Cloud computing security is a broad research domain with a large number of concerns, ranging from protecting hardware and platform technologies to protecting clouds data and resource access (through different end- user devices). Although the advantages of cloud computing are tremendous, the security and privacy concerns of cloud computing have always been the focus of numerous cloud customers and impediment to its widespread adaptation by businesses and organizations. The paper presents a systematic literature review in the field of cloud computing with a focus on risk assessment. This would help future research and cloud users/business organizations to have an overview of the risk factors in a cloud environment. And to proactively map their indigenous needs with this technology.

1 Introduction

Cloud computing has gained considerable attention in the scientific community. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort [1]. This definition describes cloud computing as having five characteristics i.e., on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. Although there are many benefits to adopting cloud computing, there are also significant barriers to adoption. One of the most significant barrier to adoption is security [2]. As cloud computing represents a relatively new computing paradigm, therefore the most important concern is its security from both the perspective of cloud customer and Cloud Service Provider (CSP). Migrating critical applications and sensitive data to cloud environment is of great concern for organizations that are moving beyond their data centers. To mitigate these concerns, a CSP must ensure that customers will continue to have the same security and privacy controls over their applications and services and provide evidence to customers that their organization are secure and they can meet their service

level agreements [3]. Since the emergence of cloud computing in 2006, a lot of review papers based on cloud computing are available in the current literature but to date no systematic review of cloud computing risks has been published. Therefore, the primary goal of this research is to systematically select and review published research work and provide an overview of risk analysis, risk severity and impact of these risks on cloud users and providers.

The structure of the paper proceeds as follows: Section 2 presents the research methodology. Section 3 presents the overview of data concerning the reviewed studies. Section 4 gives the detailed description of the reviewed papers. Section 5 shows the analysis of the systematic review, and finally the conclusion is presented in section 6.

2 Research Methodology

The proposed research uses a systematic review methodology [4,5] to review existing literature concerning cloud computing security risks, vulnerabilities and threats.

2.1 Research Questions

In this phase, the review process is planned and research questions were identified. The following research questions are addressed in this study: **RQ1**. What are the risks associated with cloud computing from cloud customer's perspective? **RQ2**. What are the risks associated with cloud computing from cloud Provider's perspective?

2.2 Defining the Review Protocol

After the selection of the research question, a set of search terms called keywords was extracted. The keyword and relevant initiatives that make up research question and that were used during review protocol are: risk assessment, risk analysis, systematic literature review, cloud computing security, cloud vulnerabilities and cloud threats. To perform a systematic literature review, the primary research focus is from Springer Link, Elsevier, Science Direct, ACM, DBLP and IEEE digital library. The research was narrowed down through the set of *inclusion and exclusion criteria*. Only full papers in English from peer-reviewed articles, journals and conference proceeding on the specified topic, published from 2009 to 2013 were considered. A *quality assessment checklist* (QAC) is developed to assess the individual studies. The QAC is prepared based on kitchenham [4]. The Checklist includes the following questions: a) Does the research paper clearly specify the research methodology? (b) Is the research methodology appropriate for the problem under consideration? (c) Are the analysis of study properly done? If the study fulfills assessment criteria then it is filled with 'yes'.

2.3 Data Extraction and Synthesis

We identified a total of 100 studies. After filtering these studies according to inclusion/ exclusion criteria and QAC, 31 publications were identified as a primary study

for review [6-37]. These included journals (15), conference proceedings (11), white papers (3) and articles (2). The year of publication of the papers is shown in Fig. 1. Fig. 2 shows the distribution of papers in relation to the research questions (RQ).

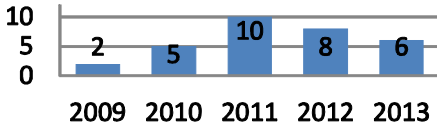


Fig. 1. Paper distribution over publication year

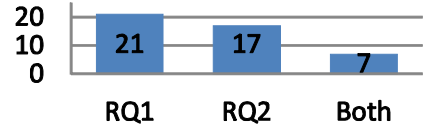


Fig. 2. Paper distribution in relation to RQ

3 Results

The detailed analysis of the selected studies was based on their similarities in terms of the risk analysis in cloud computing. According to the reviewed publication and to answer the research questions we identified five main categories of risks associated with cloud computing both from cloud provider and customer perspective. These categories include: Organizational, Technological, Data Security and Privacy, physical security and Compliance. Fig. 3 shows risk categories along with their sub-categories. We elaborate these categories in the subsequent section from both cloud providers and customer's perspective.

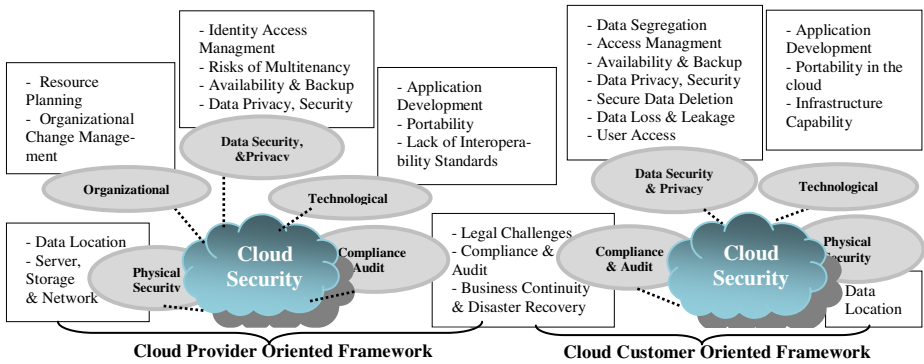


Fig. 3. Cloud Security Risk Categories and sub- Categories

The results of the systematic review are given in Table 1.

Table 1. Summary of the risks considered in each approach

Cloud Provider Risks	List of Studies
Data Security & Privacy	[7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [22]
Technological	[7], [8], [19], [21], [22]
Physical Security	[25], [26]
Organizational	[10], [22], [29]
Compliance and Audit	[20], [23], [24]

Cloud Customer Risks	List of Studies
Data Security & Privacy	[12], [16], [20], [24], [25], [27], [30], [31], [32], [37],
Technological	[10], [22], [33]
Physical Security	[22], [35]
Compliance and Audit	[28], [30], [34], [38]

3.1 RQ1: Risks from Cloud Providers Perspective

1. Data Security, Privacy & Control Risks

Data security and privacy risks are mitigated through data encryption and it is the CSP responsibility to handle these rudimentary risks [6]. To ensure data integrity, confidentiality and availability, the storage provider should offer encryption schema and scheduled data backups [7]. CSP is responsible to adopt added security measures to ensure data security. These security measures involve the use of strong encryption techniques for data security and fine-grained authorization to control user access to data [8]. Providers are more responsible for the privacy and security of data and application services in public than in private clouds [9]. The major problem with data encryption is the responsibility of key management. Ideally, it's the data owners. But due to the lack of user expertise to handle the keys, they usually hand over the key management to CSP. But again it will become more difficult for CSP to maintain keys for a large number of users [10, 11]. CSP is the one responsible for the security of the data while is being processed, transferred and stored [12]. CSP does not have permission for access to the physical security system of data centers rather they must depend on the infrastructure provider to get full data security. The CSP can only specify the security settings remotely, and don't know either they are fully implemented or not. It is major security risks for CSP if the security settings are not fully implemented [13].

— Identity and Access Management (IAM)

IAM improves operational efficiency, regulatory compliance by managing the major security concerns, automated provisioning, authentication and authorization services. Devki solves this issue by using various techniques such as single sign-on, federated identity, access control list, directory based service, access on the basis of attributes [21]. To avoid unauthorized access, the CSP should offer strict access control mechanism. In cloud computing administrative access is done through the internet and this increases the risk of unauthorized access to data and resources. Therefore, it is very important to control and monitor the administrative access to maintain protocols [7]. Data in the cloud is globally distributed which brings the issue of jurisdiction and privacy [11]. According to study only 37% of cloud providers were confident about security to authenticate users before granting access, whereas 50% of cloud users considered IAM as the cloud provider's responsibility. Therefore, achieving compliance requirements could be problematic [15]. According to [14], when the data is outsourced to a cloud, enforcing secure and reliable data access between several users is very critical. The user cannot even trust the server because the user's private data can be exposed in the event of server compromise. The solution is to encrypt data in differentiated manner and disclose the corresponding decryption keys only to authorized users. This approach has a drawback of performance loss and scalability [16]. Gartner list seven security issues from CSP perspective. The data security and privacy risks include privileged user access, which inquires about who has access to data [17].

— Multi-tenancy

Multi-tenancy is an essential attribute of cloud computing as it increases the use of underlying hardware resources and allowing for efficient resource provisioning.

Multi-tenancy security and privacy is one of the critical challenges for the public cloud [16]. It is the responsibility of CSP to ensure an isolated boundary for each user's data at both physical and application levels [8]. It is possible that the customers' personal and financial data are stored by the CSP. Therefore, CSP is responsible to secure the customers' data. Some providers use job scheduling and resource management, but most providers employ Virtualization to maximize the use of hardware [18]. These two methods allow attackers to have full access to the host and cross-VM side channel attacks to extract information from a target VM on the same machine. In multi-tenancy, data from multiple tenants is likely to be stored in the same database, the risk of data leakage between these tenants is high [12]. Data is placed in a shared environment with the data from other clients which poses a great risk of multi-tenancy for CSP. There is a need for some mechanism through which CSP must guarantee data isolation between clients and they also should be liable for ensuring this isolation[19].

— Data Availability and Backup

It is difficult for CSP to guarantee adequate availability and backup of data in the cloud because the data are hosted distantly in the cloud. Therefore it is not only difficult to backup the data but also to recover the data in case of failure [18]. In the cloud environment, there are several areas that will threaten the data availability including the availability of cloud computing services, whether the cloud providers would continue to operate in the future? whether the cloud storage services provide backup? [10]

2. Organizational Risks

Organizational risks are categorized as the risks that may impact the structure of the organization or the business as an entity. These risks include the loss of business reputation and any organizational change that can happen to the CSP and cause the provider's failure, termination of the acquisition [28].

— Organizational Change Management

Resistance to change consequent from organizational politics, changes to people work is a major organizational risk. To mitigate this, use insight from organizational change management and involve key stakeholders in the adoption procedure [21].

— Resource Planning

According to Hosseini et al. [21], the risk to resource planning is the loss of control over resources, which lead to ambiguous roles and responsibilities. To overcome this, it is essential to clarify roles and responsibilities before cloud adoption.

— Organizational Security Management

The existing security management models have considerably changed when enterprises adopt cloud. There is a need to reevaluate the existing security models and develop security standards to ensure the deployment and adoption of secure clouds [9].

3. Technical Risks

Technical risks are defined as the failures associated with the technologies and services provided by the CSP, including resource sharing isolation problems, malicious attacks on the CSP risks related to portability and interoperability [20]. Technical

risks are related to hardware including poor maintenance of hardware, unresponsive system, reduction in the availability and hardware failure [6].

— **Portability in the Cloud**

Interoperability between clouds are due to incompatibilities between CSP platforms. The solution is to use cloud middleware for the ease of cloud interoperability [21].

— **Application Development**

Risk of service interruption at providers side results in extensive outages and unavailability of services or loss of data. The solution suggested by the authors is to use multiple cloud providers and monitor applications from outside the cloud [21].

— **Lack of Interoperability Standards**

Cloud computing lacks interoperability standards. There is no standard of communication and data export format between and within CSP, which makes it difficult to establish appropriate security frameworks [18]. For CSP, adoption of universal standards is also recommended to ensure interoperability among CSP [7].

4. Compliance and Audit

Risks related to lack of jurisdiction information, changes in jurisdiction, illegal clauses in the contract and ongoing legal dispute. It is the responsibility of both CSP and customer to abide by the rules and regulation defined in the contract and audit SLAs regularly [22]. Traditional CSP is subjected to external audits and security certifications. If a CSP does not adhere to these security audits, then it leads to an obvious decrease in customer trust [23]. CSP should have security policies with recovery methods in case of disasters and the ability to restore data completely in a pre-established amount of time [19].

5. Physical Security

— **Data Location and Data Center**

CSP should guarantee secure operation of the cloud data center in order to provide a secure physical location for customers' data [24]. CSP manages the infrastructure including servers, networks, storage devices. CSP should implement and operate appropriate infrastructure controls including staff training, physical location security, network firewalls. To overcome these risks are of utmost importance because if the physical access control is weak, attackers can steal entire servers, even if they are protected by firewalls and encryption [24]. The cloud provider is not only responsible to store and process data in specific jurisdictions but should also responsible to obey the privacy regulations of those jurisdictions [25].

3.2 RQ2: Risks from Cloud Customer Perspective

1. Data Security, Privacy & Control Risks

— **User Access**

The customer is fully responsible for the management of all software security controls. These include application access control, IAM, software patching, viruses

protection [24]. One of the risks is how a customer face the privileged status of CSP and security issues such as fault elimination, data damage and data migration [26].

— Data Privacy and Security

It is an essential security concern for the end- users to know about the privacy and protection of their data from CSP in order to ensure that data privacy is not compromised. But eventually the customers are responsible for the security and integrity of their own data even it resides on providers premises [15]. The loss of encryption key or privileged access code will bring serious problem to cloud service users [36]. Accordingly, lack of cryptographic management information will heavily lead sensitive damages of data loss, and unexpected leakage of user data to the outside world. Customer data and commercial secrets should not be leaked while residing on CSP premises [24]. According to CSA group [30], the burden of avoiding data loss does not fall completely on the provider's shoulder. If a customer encrypts data before placing it to the cloud, and lost the encryption key, the data will be lost as well.

— Data Segregation

It is the responsibility of cloud customer to find out the techniques used by the provider to segregate the data and must ensure that the encryption schemes are deployed and are effective enough to provide security [29]. Encryption cannot be assumed as the single solution for data segregation problem. In some cases, customers may not want to encrypt data because encryption accident can destroy the data [23].

— Data Availability

When the client data is uploaded into the cloud, clients no longer possess any data on the cloud. Customers personal data and information on the Cloud in not available either lost or heck, it is difficult to retrieve the original data [31].

— Secure Data Deletion

Appropriate, error free and timely data deletion may be impossible and undesirable. One of the reasons is the extra copies of data reside at different locations and the other is that the disk to be destroyed also contain data from other clients [19]. Data is supposed to be destroyed completely, when it is no longer required. But due to the physical characteristics of storage medium, the data deleted may still exist and can be restored. This may cause a risk of sensitive data disclosure to the customer [11].

6. Technical Risks

— Infrastructure Capabilities

It is difficult to show CSP that their cloud performance is not in accordance with their agreed SLA because of the server's workload and variable nature of the network. This cause disputes and litigation. The solution is to evaluate the cloud performance under appropriate investigation before adopting. Another solution is to use third party monitoring tools for the verification of system performance [21].

— Application Development

The purpose is to allow developers to develop their applications over the provided platform. Therefore, the customers are mainly responsible for protecting their

developed applications and the platform. At the same time, the providers are responsible for isolating the customers applications and development environments [9].

— Portability

According to K. Popovic and Z. Hocenski [32], the risk of compatibility arises if the customer wants to move from one provider to the other because the storage services offered by one CSP may be incompatible with another provider's service.

7. Compliance and Audit

— Disaster Recovery

Cloud Customer should know what will happen to their data if a disaster occurs. Therefore, it is the customers primary security responsibility to ask whether the provider will be able to completely recover your data and how long it will take. [29]

— Legal Challenges

CSP is more susceptible to legal and regulatory concerns and commit to keep and process customers' data in specific jurisdictions that provides security and privacy of data as promised in their SLA's. Even then, the organizations are mainly responsible for the privacy of their data kept at the CSP site [33]. The computer processing power or storage one buys via a Cloud service may be based in another country or may be divided between multiple countries. Despite the advantages of cost and efficiency, it raises legal issues by exporting customer's data abroad [27,37].

8. Physical Security

— Data Location

As the data is stored redundantly in multiple physical locations by the CSP and that location information is not revealed to the customer. On the customer side, it is difficult to determine whether appropriate security measures are in place to secure customers' data [21]. The customer cannot avoid the downtime of a cloud computing environment, which is the time in which the CSP machines are not working properly. This situation brings immense discourage to the confidence of customers [34].

4 Discussion And Analysis

The aim of the proposed research is to thoroughly examine the selected papers and identify the security risks. Customer's thought that once the customer organization relinquish cloud computing responsibility to a CSP, all the security would now be the responsibility of the CSP. But it is equally important that how the customer data is moving outwards from the customer's organization. Before adopting cloud computing, risk management policies and mechanisms need to be developed and properly formulated [35]. In Table 2, we suggest the possible security measures that help in mitigating the identified risks to some extent. The risks related to cloud provider are represented as (CP) and the risks related to cloud customer are represented as (CC).

Table 2. Risks and Suggested Security Measures

	RISKS	SECURITY MEASURES
Data Security & Privacy	Ensure availability of customer's data in cloud (CP)	Specific security measures have been taken by CSP to prevent outages and attacks
	Risks related to data security and privacy (CP), (CC)	<ul style="list-style-type: none"> To mitigate these risks is using APIs to implement a robust access control, using encryption to protect data traffic. Analyze that data is protected during design time, as during run-time. Provide effective mechanisms for key generation, storage, and destruction of data
	Preventing unauthorized access to customer's data in the cloud (CP), (CC)	Can be resolved by implementing Management, authentication and authorization techniques on both customer and provider's sides
	Risks related to multi-tenancy (CP)	CSP should use effective encryption methods to guarantee data isolation between clients.
	Risks related to data deletion (CP)	The provider should define policies to establish procedures for the destruction of persistent media before throwing it out.
Technology	Lack of standardized technology in the cloud computing system (CC)	The customer should ensure if the provider uses standardized technology and it should be mentioned in its initial contract.
	Compatibility issue between cloud and IT systems in customer's organization (CC)	The solution is to use the hybrid cloud, which is capable of handling much of these compatibility issues
Organizational	Risks related to Resource Planning, Change Management (CC)	Involves stakeholders in cloud adoption procedures
	Risks related security management (CC)	Reevaluate existing security standards before cloud adoption.
Physical Security	The physical security of a cloud provider's data centers composed of servers, storage and network devices. (CP)	Cloud providers must have certain policies and procedures in place to prevent physical security breaches these includes physical location security like alarms, CCTV cameras etc.
Compliance	Enforce regulatory obligations in a cloud environment. (CP)	<ul style="list-style-type: none"> CSP must abide by all the regulations within a country, regarding cloud security. These regulations include HIPPA, FISMA CSP has to contend with the Legal Systems under different Jurisdictions with not so much of visibility as to where the Data resides and how it is routed by passing through different Legal Jurisdictions.
	Business Continuity and Disaster Recovery (CP)	Recommends replicating data across multiple infrastructures to avoid vulnerabilities in the event of a major failure

5 Conclusion

The main objective of the SLR presented in this paper is to categorize risks related to cloud computing paradigm. We have followed the SLR methodology depicted in [5] to identify 31 primary studies out of 100 papers. Our analysis is concerned with both service provider and consumer in cloud computing. The review and analysis of the selected studies identify a number of challenges and indicates that there is still enormous opportunities for researchers to contribute in this area. Some topics such as data security and privacy are widely investigated, while others, e.g. physical and organizational security, have received less attention. The framework we have proposed contributes to organize the available knowledge and indicates future research directions.

References

1. Rebollo, O., Mellado, D.: Systematic Review of Information Security Governance Frameworks in the Cloud Computing. *Journal of Universal Computer Sc.* 18(6), 798–815 (2012)
2. From hype to future: KPMG's 2010 Cloud Computing survey, <http://www.techrepublic.com/whitepapers/fromhype-to-futurekpmgs-2010-cloud-computing-survey/2384291>

3. Rittinghouse, J., Ransome, J.: Security in the Cloud: Cloud Computing. Implementation, Management, and Security, 1st edn. CRC Press (2009)
4. Hannay, J.E., Sjøberg, D.I.K.: A Systematic Review of Theory Use in Software Engineering Experiments. *Journal of IEEE Transaction on Software Engineering* 33(2), 87–107 (2007)
5. Kitchenham, B., Brereton, O.P.: Systematic literature reviews in software engineering –A systematic literature review. *Journal: Information and Software Technology*, 7-15 (2009)
6. Djemame, K., Armstrong, D.: Risk Assessment Framework and Software Toolkit for Cloud Service Ecosystems. In: *Int. Conference on Cloud Computing, GRIDs, and Virtualization* (2011)
7. Harauz, J., Kauffman, M., Potter, B.: Data Security in the world of cloud computing. *IEEE Security & Privacy* 7(4), 61–64 (2009)
8. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications* 34(1), 1–11 (2011)
9. Takabi, H., Joshi, J.B.D.: Security and Privacy Challenges in Cloud Computing Environments. Published. *IEEE Security and Privacy* 8(6), 24–31 (2010)
10. Chen, D., Zhao, H.: Data Security and Privacy Protection Issues in Cloud Computing. In: *Int. Conference on Computer Science and Electronics Engineering*, pp. 647–651 (2012)
11. Rahul, S.S., Rai, J.K.: Security & Privacy Issues In Cloud Computing. *International Journal of Engineering Research & Technology (IJERT)* 2(3) (March 2013)
12. Hashizume, K., Rosado, D.G., Medina, E.F., Fernandez, E.: An analysis of security issues for cloud computing. *Journal of Internet Services and Applications* 4(5) (2013)
13. Reddy, V.K., Thirumala, R.B., Reddy, L.S.S., Kiran, S.: Research Issues in Cloud Computing. *Global Journal of Computer Science and Technology* 11(11) (July 2011)
14. Pal, D., Krishna, R., Srivastava, P., Kumar, S.: A Novel Open Security Framework for Cloud Computing. *Int. Journal of Cloud Computing and Services Science* 1(2) (2012)
15. Argall, K.: Compliance in a Cloud Computing Environment. *HIPAA and PCI DSS* (2010)
16. Ren, K., Wang, C., Wang, Q.: Security Challenges for the Public Cloud. *Journal of Internet Computing IEEE* 16(1) (2012)
17. Lovell, R.: White Paper: Introduction to cloud computing (October 2009)
18. Pearson, S., Benameur, A.: Privacy, Security and Trust Issues Arising from Cloud Computing. In: *2nd Int Conference on Cloud Computing Technology and Science* (2010)
19. Ayala, L.C., Vega, M., Vargas, L.M.: Emerging Threats, Risk and Attacks in Distributed Systems: Cloud Computing. In: Elleithy, K., Sobh, T. (eds.) *Innovations and Advances in Computer, Information, Systems Sciences, and Engineering*. LNEE, vol. 152, pp. 37–52. Springer, Heidelberg (2013)
20. Rana, S., Joshi, P.K.: Risk Analysis in Web Applications by Using Cloud Computing. *International Journal of Multidisciplinary Research* 2 (January 2012)
21. Khajeh- Hosseini, A., Sommerville, I., Bogaerts, J., Teregowda, P.: Decision Support Tools for Cloud Migration in the Enterprise. In: *IEEE CLOUD 2011* (November 2011)
22. Chou, Y., Oetting, J.: Risk Assessment for Cloud-Based IT Systems. *International Journal of Grid and High Performance Computing*, 1–13 (April–June 2011)
23. Kumar, V., Swetha, M.S.: Cloud Computing: Towards Case Study of Data Security Mechanisms. *International Journal of Advanced Technology & Engineering Research* 2(4) (2012)
24. Julisch, K., Hall, M.: Security and Control in the Cloud. *Information Security Journal: A Global Perspective*, 299–309 (2010)
25. Kumar, A.: World of Cloud Computing & Security. *International Journal of Cloud Computing and Services Science (IJ-CLOSER)* 1(2) (June 2012)

26. Che, J., Duan, Y., Zhang, T.: Study on the Security Models and strategies of cloud Computing. In: Proc: Int Conference on Power Electronics and Engineering Application (2011)
27. Prasad, M., Naik, R., Bapuji, V.: Cloud Computing: Research Issues and Implications. International Journal of Cloud Computing and Services Science 2(2), 134–140 (2013)
28. Dahbur, K., Mohammad, B.: A Survey of Risks, Threats and Vulnerabilities in Cloud Computing. In: Int Conference on Intelligent Semantic Web-Services and Applications (2011)
29. Bisong, A., Rahman, S.M.: An Overview of the Security Concerns in Enterprise Cloud Computing. International Journal of Network Security & its Applications 3(1) (January 2011)
30. Cloud Security Alliance CSA: The Notorious Nine Cloud Computing Threats 2013 (2013)
31. Ahmad, T., Amanul, H.M., Al-Nafjan, M., Ansari, A.: Development of Cloud Computing and Security Issues. Information and Knowledge Management 3(1) (2013), <http://www.iiste.org>
32. Popović, K., Hocenski, Ž.: Cloud computing security issues and challenges. MIPRO (2010)
33. Jansen, W., Grance, T.: Guidelines on Security and Privacy in Cloud Computing. NIST (2011)
34. Peiyu, L., Dong, L.: Risk Assessment Model for Information System in Cloud Computing Environment. Advanced in Control Engineering and Information Science. V. 15 (2011)
35. Rosado, D.R., Gomez, R.D., Mellado, Medina, E.F.: Security Analysis in the Migration to Cloud Environment. Journal: Future Internet, 469–487 (May 2012)
36. Lee, K.: Security Threats in Cloud Computing Environments. International Journal of Security and Applications 6(4) (October 2012)
37. Sharma, M., Bansal, H., Sharma, A.K.: Cloud Computing: Different Approach & Security Challenge. International Journal of Soft Computing and Engineering 2(1) (March 2012)

Security Requirements Specification Framework for Cloud Users

Rida Naveed¹ and Haider Abbas^{1,2}

¹Department of Information Security, Military College of Signals,
National University of Sciences & Technology (NUST),
Islamabad, Pakistan

ridaanaveed@gmail.com

²Centre of Excellence in Information Assurance (COEIA),
King Saud University,
Riyadh, Saudi Arabia

hsiddiqui@ksu.edu.sa, haidera@kth.se

Abstract. Cloud computing has gained significance due to its accessibility and highly scalable computing resources in today's emerging IT technologies. These cloud resources are shared among all cloud entities at different levels of operation. And due to its complex architecture it is prone to a number of security threats. These security and privacy challenges must be taken into consideration by organizations when they have to outsource their data, infrastructure and applications into a cloud environment. The objective of this paper is twofold i.e. it highlights critical security challenges introduced in cloud environment, specific security requirements are analyzed for cloud users and a framework for engineering these security requirements is also presented. The paper proposes a cloud security assurance framework that helps users by providing a methodology for identifying security requirements of their assets at early stages of the cloud deployment process. It also provides mechanism to specify cloud system's deployment requirements.

Keywords: Cloud Security Requirements Engineering, Cloud Computing Security Requirements, Cloud Security Assurance Framework.

1 Introduction

Cloud computing has emerged as outsourced resource sharing computing technology that provide on-demand storage, software, computational power, infrastructure and network access to its users over the internet [6]. With the implementation of virtualization, service orientation and grid computing technologies, has increased trend of organizations and business entities towards its adaption. Because of various benefits like rapid resource sharing, location independence and elasticity, it has overcome a long awaited vision of separating users from their physical hardware needs thus providing them more flexible and scalable IT services [2][4][5].

Cloud computing, despite of its tremendous benefits, does not comes without its pitfalls [1]. The amalgamation of different computing technologies in cloud computing gives rises to various security and privacy concerns. These concerns, if not taken into consideration can become security threats for the organizations adapting cloud model. This paper intends to focus on several critical security challenges introduced in a cloud environment [5]. Specific security requirements with respect to cloud users for each of cloud's service model are investigated and a framework for engineering these security requirements is also documented. The paper proposes a cloud security assurance framework, which helps users from security burden [5], by providing them a methodology of identifying security requirements of their assets from the early stages of the cloud deployment process alongside cloud system's deployment requirements [3]. Cloud computing security requirements [10] and challenges for adopting it for US government [11] has been discussed in previous work related to this topic, but an idea of a comprehensive framework for cloud security assurance has not been considered so far. This work will help in the emergence of computer system with security enforcement mechanisms incorporated at the time system's functional requirements are been met thus eliminating security challenges within a system.

The structure of a paper proceeds as follows: Section 2 presents the cloud security challenges. Section 3 presents the security requirements of cloud service Models. Section 4 describes in detail the cloud security assurance framework model for governmental organization. In section 5 discussion and finally the conclusion is presented in section 6.

2 Security Challenges in Cloud Environment

Cloud computing has gained immense popularity among individual home users with small enterprises to local and foreign government businesses. Depending on an organization's need different deployment and service models with diverse technologies can be configured [12]. There are a number of ways in which cloud computing can be deployed either privately, public cloud, community cloud, or a combination of two or more public, private or community clouds i.e., hybrid cloud [12][13].

Moving towards cloud; need a paradigm shift in the way people think about security, users must understand that the Cloud Service Providers, CSPs are separate administrative entities and moving to cloud will deprive them of direct control over the systems that manage their data and applications [4]. Following are several critical security challenges that are faced to cloud users:

1. Due to clouds greater flexibility and cost-effectiveness, users tend to store more and more data onto it. And their confidentiality and integrity are at risk, as users no longer physically possess their data.
2. Cloud lack transparency of its operations from its users especially if their outsourced computational workloads contain sensitive information. Returning of incorrect results, software bugs, hardware failure, cross user data de-duplication [4], data deletion and attack on cloud servers can cause the cloud to behave deceitfully.

3. It is of critical importance for CSPs to have a trustworthy relationship with its customers on its service metering and usage charges. As cloud is a shared resourced network so its memory, network bandwidth, I/O and CPU cycles consumed per user cannot be isolated, nor can its charges per resource consumption is fairly computed.
4. Multi-tenancy and virtualization, increases risk of side channel attacks and privacy leaks, making reliable security difficult to achieve.
5. Data interception, impersonation, session hijacking, traffic flow analysis, infrastructure misuse, hardware theft, latency and natural disaster are possible security threats that can be introduced to a user when working in cloud environment.

3 Security Requirements of Cloud Service Models

The critical security challenges discussed above have gained significance and need to be addressed carefully. Cloud computing being an amalgamation of complex networked system is inherently affected by a great number of computer and network security issues. These security concerns arise by not considering the security requirements at the beginning of the system development process. As cloud is available to users in three service models, therefore, understanding and clearly documenting user specific security requirements is very critical in designing of vulnerability free computer systems. A detailed list of security requirements for cloud users in analyzed in Table 1 [5].

Table 1. Security Requirements for Cloud Users

Service Model	Users	Security Requirements	Security Goals
SaaS	End Users /Organizations who needs to access its application resources on rent	<ul style="list-style-type: none"> - True server authentication - Application software testing - Scalability - Maintenance of infrastructure - Service uptime and security - Security of sensitive data - Abstract interaction dynamics issues - Browser-based Risks - Network dependence issues - Efficiency vs. cost tradeoffs - Privacy in multitenant environment - Access control - Communication protection 	<ul style="list-style-type: none"> -Confidentiality - Integrity - Availability - Accountability
		----- -	
		<ul style="list-style-type: none"> - Service availability 	

Table 1. (continued)

PaaS	Developers /Moderators who needs to construct high quality dynamic applications and requires more application level logic to perform	<ul style="list-style-type: none"> - Browser-based Risks - Network dependence issues - Efficiency vs. cost tradeoffs - Compatibility issues between PaaS clouds - Processor scheduling concerns - Application reuse security issues - Access control - Application security - Data security - Cloud management control security - Secure images 	<ul style="list-style-type: none"> - Confidentiality - Integrity - Availability - Accountability
IaaS	System administrators who needs to access computational infrastructure available over the internet such as virtual computers, network, storage, infrastructure components such as firewalls, and configuration services.	<ul style="list-style-type: none"> - Abstract interaction dynamics issues - Browser-based Risks - Network dependence issues - Efficiency vs. cost tradeoffs - Compatibility with legacy software vulnerabilities and Data erase practices -Virtual machine updating, checking and maintenance, VM-level isolation - Verifying legitimacy of web sites 	<ul style="list-style-type: none"> - Confidentiality - Integrity - Availability - Accountability

4 Cloud Security Assurance Framework Model for Governmental Organization

This security requirements analysis must be done at the start of the cloud system development process so that the essential security enforcement mechanisms must be fitted in a system design process. This overcomes the common approach of including security within a system after the definition of a system [3]. This section presents a framework based on [14] for engineering security requirements for cloud environment. This cloud security assurance framework model, provides a methodology to

cloud users in general and government sector cloud users in specific, to identify security requirements of their assets at the time they decide to move their data towards cloud. It will help them to specify the level of security and privacy they required for their system that they would run on cloud infrastructure. Thus reducing security/functional requirements conflict by avoiding them from the very beginning of the development process.

As a case study, a law enforcement department has been taken (as a governmental organization) who has to make decision of moving their data to cloud service provider. A framework described shows how to move step by step for achieving secure cloud services for their organization. The first step is to identify functional requirements, secondly identify security goals, thirdly identify security requirements, and finally forming a cloud security assurance framework model.

4.1 STEP 1: Identify Functional Requirements

By identifying functional requirements means to draw all systems context for the law enforcement department under consideration. It is done by identifying the necessary task, action or activity that must be accomplished by each of its department. A brief overview of task or actions each wing has to perform is shown in Table 2.

Table 2. Functional Requirements of Law Enforcement Government Department

Law Enforcement Departments	Functional Requirements
Anti-Corruption Wing	Deals with organized crimes such as anti-corruption, spurious drugs, counterfeit currencies, PPC and other laws.
Economic Crime Wing	Responsible for investigation of cases related to government revenue thefts.
Technical Wing	Provides scientific assistance to various units of federal government departments.
Immigration Wing	Regulates flow of incoming and outgoing international passengers and prevents human smuggling via airports, land routes, sea ports and railway stations.
Anti-Trafficking Unit Legal Branch	Deals with the prevention and control of human trafficking. Provide legal guidance in all administrative and operational matters.
Intellectual Property Rights Branch	Ensures that every system provides different types of warranty to ensure peaceful possession of property, tangible and intangible.
Interpol	Coordinate efforts relating to international police corporation.
Counter Terrorism Wing	Responsible to identify, arrest and put to trial most wanted terrorists and to provide qualitative investigations for counter terrorism
Academy	Responsible to prepare and train the newly hired officers of federal government

4.2 STEP 2: Identify Security Goals

According to Haley's paper [14], security goals can be identified by three general steps i.e. *i) Identification of Assets* (Assets may include information/data asset, technology asset, human resource asset and service asset) *ii) Management Principles* (These principles may include separation of roles and duties, separation of function, data protection and no connections from outside, requirement of audit trails etc.) and *iii) Security Goals* (Identifying and conducting a harm/risk analysis for the assets, it need to be analyzed if these assets are covered by the organizations policies). A list of security goals is then determined by applying management principles to these assets and keeping in mind the confidentiality, integrity, availability and accountability.

4.3 STEP 3: Identify Security Requirements

The paper focused on identifying security requirements that are constraints on functional requirements of a system [14]. Generalized security requirement checklists for the government departments that need to be considered are given below:

- **Pricing** – The initial setup fee for cloud deployment, if the fees is charged according to bandwidth usage or the number of users and rate of increase in charges by the cloud must be taken into consideration by the cloud users. Infrastructure setup cost and cloud service usage cost comparison should also be made.
- **Service Provider's Size** – The actual size of the cloud service provider CSP, type of office it is housed in, size of its security team and if security a full-time job at the vender must be made sure by the cloud users.
- **Secure Area/ Physical Entry Control** – The location where cloud service provider hosts its facilities must be in secure area.
- **Power Supplies** – Protection of electronic equipment from power failures and other electrical anomalies must be ensured.
- **Cabling Security** – All power and telecommunications cabling must be protected from interception or damage.
- **Separation of Development and Operational Facilities** – Operational and development facilities must be separated to minimize the risk of unauthorized access or accidental changes to data or production software.
- **Environmental Monitoring** – Monitoring of host computer environments must be done, including temperature, humidity, and power-supply quality.
- **Capacity Planning and Acceptance** – Resource availability must be ensured by doing capacity planning and preparations in advance to meet customers growing demand.
- **Availability of provider** – Availability of provider must be ensured by knowing the uptime guaranteed by the provider, how it is calculated, compensation for not satisfying the guaranteed uptime, cost per minute of user's service downtime.
- **Sensitivity of Information Stored** – The cloud users must make a risk profile of the sensitive data they want to move towards cloud.
- **Data Storage** – Users must know the location of their data and data servers. Whether the data is hosted on dedicated, or shared, hardware.

- **Data Accessibility** – Cloud users must ensure that who has access to their data and applications and systems hosting them in cloud. What are the access controls in place according to roles and responsibilities of cloud users?
- **Authentication of Users** – It is CSP’s responsibility to authenticate the users to avoid unauthorized user sign in and uses of cloud services.
- **Data Encryption** – Cloud users must make sure that their data is properly encrypted, for example, using 256 bit AES or SSL for secure data transfer.
- **Controls** – User must look for the controls in place to ensure confidentiality and integrity of their data. Controls like roles, permissions to file access, virus detection and prevention and user awareness procedures must be implemented.
- **Data Security Responsibility** – Cloud users must know who is responsible for storing, processing and using their personal and sensitive information, personals responsible in case of any security breach, how they respond to it.
- **Network Monitoring** – Network monitoring must be done 24hrx365days to make sure that the cloud infrastructure, networks and resources are safe and protected.
- **Firewalls and Patches** – Cloud users should be concerned about the control of the influx and outflow of the traffic of their organization. Latest version of operating system and desktop applications should be used in conjunction with the cloud applications.
- **Data Backup** – User’s data should be backed up, preferably off-site.
- **Policies and Procedures** - Procedures and policies e.g. a clear desk policy, documented management authorization for the removal of property like equipment, data or software, management of all computer and networks, cloud provider’s security policies and standards, operating procedures documentation, incident management responsibilities and procedures, fault logs for recording reported faults by the users and for reporting and taking corrective action must exist.
- **Vulnerability Testing and Security Audits** – Users must make sure that the test for all categories of vulnerability and security audits must be done regularly.
- **Termination Clauses** – It must be known to the users that under what grounds cloud providers can dismiss their contract and how sooner will they get their data back from cloud after termination.

This aforementioned framework can be depicted as follows in Fig. 1.

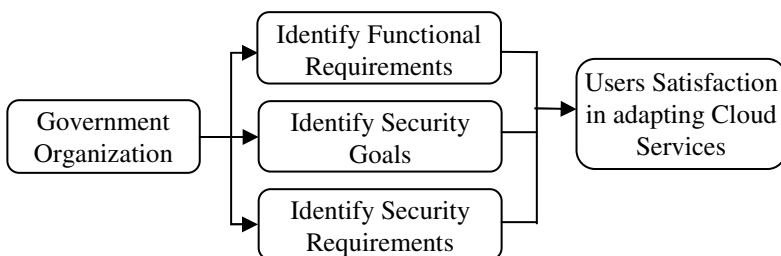


Fig. 1. Cloud Security Assurance Framework Model

5 Discussion

Several security requirements engineering methodology exist [14] but engineering security requirements of a system into cloud security assurance framework has not been addressed so far. Based on an extensive study; the framework is presented that would help users to identify security requirements of their assets and specify the level of security and privacy they required for their systems that they would run on cloud infrastructure. The paper explains the possible critical security challenges that the user might face in adapting cloud environment. Cloud service models, their use and security requirements for each is discussed in detail so that they must select an appropriate service that fits well to their specific environment.

6 Conclusion

The paper presents security assurance framework for cloud users to help in the security enforcement mechanisms incorporated in system's functional requirements and to counter security challenges. The framework emphasizes that security requirements must be considered at early stage before moving organization's sensitive data to cloud environment. Security requirement engineering needs to be devised to test and validate for the development of real time secure cloud systems.

References

1. Andrei, T.: Cloud Computing Challenges and Related Security Issues, <http://www.cse.wustl.edu/~jain/cse571-09/ftp/cloud/index.html> (last accessed on September 23, 2012)
2. Waqar, A., Raza, A., Abbas, H., Khan, M.K.: A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata. *Journal of Network and Computer Applications* 36(1), 235–248 (2013), <http://dx.doi.org/10.1016/j.jnca.2012.09.001>
3. Dubois, E., Mouratidis, H.: Guest editorial: security requirements engineering: past, present and future. *Requir. Eng.* 15(1), 1–5 (2010)
4. Ren, K., Wang, C., Wang, Q.: Security Challenges for the Public Cloud. *Internet Comput. IEEE* 16(1), 69–73 (2012)
5. Zissis, D., Lekkas, D.: Addressing cloud computing security issues. *Future Gener. Comput. Syst.* 28(3), 583–592 (2012)
6. Venters, W., Whitley, E.A.: A critical review of cloud computing: researching desires and realities. *J. Inf. Technol.* 27(3), 179–197 (2012)
7. Jensen, M., Schwenk, J., Gruschka, N., Iacono, L.L.: On Technical Security Issues in Cloud Computing. In: *IEEE Int. Conf. on Cloud Computing (CLOUD 2009)*, pp. 109–116 (2009)
8. Chen, Y., Paxson, V., Katz, R.H.: What's New About Cloud Computing Security? EECS Department, University of California, Berkeley, USA, UCB/EECS-2010-5 (2010)

9. Yu, H., Powell, N., Stembridge, D., Yuan, X.: Cloud computing and security challenges. In: Proceedings of the 50th Annual Southeast Regional Conference, New York, NY, USA, pp. 298–302 (2012)
10. Iankoulova, I., Daneva, M.: Cloud computing security requirements: A systematic review. In: RCIS 2012, pp. 1–7 (2012)
11. NIST, Cloud Computing Security Working Group, Challenging Security Requirements for US Government Cloud Computing Adoption (Draft) (November 2011)
12. Badger, L., Grance, T., Patt-Corner, R., Voas, J.: Draft-NIST-SP800-146-NIST Draft Cloud Computing Synopsis and Recommendations. Recommendations of the National Institute of Standards and Technology, <http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf> (accessed on September 12, 2012)
13. Huth, A., Cebula, J.: The Basics of Cloud Computing. USCERT <http://www.uscert.gov/sites/default/files/publications/CloudComputingHuthCebula.pdf> (accessed on December 12, 2012)
14. Haley, C.B., Laney, R., Moffett, J.D., Nuseibeh, B.: Security Requirements Engineering: A Framework for Representation and Analysis. *IEEE Transactions on Software Engineering* 34(1), 133–153 (2008)
15. Dubois, E., Mouratidis, H.: Guest editorial: security requirements engineering: past, present and future. *Requir. Eng.* 15(1), 1–5 (2010)
16. Cloud Security Alliance, Top Threats to Cloud Computing, V1.0 by Cloud Security Alliance (March 2010)
17. Rosado, D.G., Mellado, D.: Security Engineering for Cloud Computing: Approaches and Tools. IGI Global Snippet (2012)

Digital Evidence Bag Selection for P2P Network Investigation

Mark Scanlon and Tahar Kechadi

School of Computer Science and Informatics,
University College Dublin,
Belfield, Dublin 4, Ireland
{mark.scanlon,tahar.kechadi}@ucd.ie,
<http://csi.ucd.ie>

Abstract. The collection and handling of court admissible evidence is a fundamental component of any digital forensic investigation. While the procedures for handling digital evidence take much of their influence from the established policies for the collection of physical evidence, due to the obvious differences in dealing with non-physical evidence, a number of extra policies and procedures are required. This paper compares and contrasts some of the existing digital evidence formats or “bags” and analyses them for their compatibility with evidence gathered from a network source. A new digital extended evidence bag is proposed to specifically deal with evidence gathered from P2P networks, incorporating the network byte stream and on-the-fly metadata generation to aid in expedited identification and analysis.

Keywords: Peer-to-Peer, Network, Digital Forensics, Evidence Handling.

1 Introduction

Selecting an evidence storage format for any digital investigation can be an important decision with respect to the efficiency of the entire investigation. There are numerous choices of digital evidence bags (DEBs) available with varying levels of inter-compatibility and performance benefits. Most of the existing DEBs are focused on the storage of images from physical digital storage media, e.g., hard drives, memory cards, etc. However, with respect to network focused investigations, many of the existing procedures and tools merely offer a method for capturing the network traffic in a “raw” format. The analysis stage of the investigation is generally limited to start after the completion of the network recording/sniffing process. In a non-time-sensitive network investigation, this limitation might have a minor impact on the desired evidence collection. However, in a time-sensitive, high traffic network, this limitation could result in the analysis, and possible reaction/mitigation phase of the network investigation commencing too late, e.g., in a high packet frequency P2P network such as a P2P botnet. Due to the high churn rates typical of most P2P networks, the time

required from the recording of P2P network traffic to identifying the traffic and extracting any useful evidence is paramount.

When attempting to reverse engineer a newly discovered P2P network, the job is left in the hands of the digital investigator. It is his responsibility to attempt to match common packet patterns and deduce based on the gathered evidence the protocol of the network. This paper outlines the case for a P2P network specific DEB which is capable of storing the captured packets alongside some useful metadata to aid the investigator in this process.

2 Evidence Handling

When dealing with physical forensic evidence, the commonly used handling procedure is the “chain of custody” [1]. The chain of custody commences at the crime scene where the evidence is collected, when the investigating officer collects any evidence he finds and places it into an evidence bag. This evidence bag will be sealed to avoid any contamination from external sources and signed by the officer and will detail some facts about the evidence, e.g., description of evidence, location it was found, date and time found, etc. The chain of custody will then be updated again when the evidence is checked into the evidence store in the forensic laboratory. When it comes to analysing the evidence, it will be checked out to the forensic analyst’s custody and any modification to the evidence required to facilitate the investigation, e.g., taking a sample from a collected fibre to determine its origin or identification, etc., will be logged and documented.

The procedures outlined above for physical evidence need to be extended for digital evidence acquisition and analysis. Due to the fact that digital evidence is generally analysed on forensic workstations, most of the above sequences can be automated into concise logging of all interactions. During a digital investigation, there is no requirement to modify the existing evidence in any way. This is because all analysis is conducted on an image of the original source and any discovered evidence can be extracted from this image, documented and stored separately to both the original source and the copied image. It is imperative when dealing with all types of evidence that all procedures used are reliable, reproducible and verifiable. In order for evidence to be court admissible, it must pass the legal criteria for the locality that the court case is being heard, as outlined in greater detail in section 4 below.

2.1 What Does “Forensically Sound” Really Mean?

Many of the specifications for digital forensic acquisition tools, analysis tools, storage formats and hash functions state that the product in question is “forensically sound” or that the product works with the digital evidence in a “forensically sound manner”, without specifying exactly what the term means. In 2007, E. Casey published a paper in the Digital Investigation Journal entitled “What does “forensically sound” really mean ?” [2].

In this paper, Casey outlines some of the common views of forensic professionals regarding dealing with digital forensic evidence. Purists state that any digital forensic tools should not alter the original evidence in any way. Others point out that the act of preserving certain types of evidence necessarily alters the original, e.g., a live memory evidence acquisition tool must be loaded into memory (altering the state of the volatile memory and possibly overwriting some latent evidence) in order to run the tool and capture any evidence contained in the memory. Casey then goes on to explain how some traditional forensic processes require the altering of some of the evidence in order to collect the required information. For example, collecting DNA evidence requires taking a sample from some collected evidence, e.g., a hair. Subsequently, the forensic analysis of this evidentiary sample (DNA profiling) is destructive in its nature which further alters the original evidence.

2.2 Splitting Evidence

It is not always possible to store the entire image of a particular storage device or sequence of network packets in one large file. This could be for a number of reasons, such as the evidence being stored on a FAT32 formatted hard drive which is only capable of addressing a file less than 2^{32} bytes (4,294,967,296 bytes or 4 gigabytes) or if evidence needs to be backed up to external media, e.g., a data CD or DVD, capable of storing 700MB and 4.7GB respectively. If the acquired evidence is going to be transmitted over the Internet, it should be a capability of any digital evidence bag to split the evidence into smaller parts to minimise the cost of dropped connections. The CDESF working group conducted a survey in 2006 and found that each of the evidence storage formats they tested was capable of allowing split archiving and storage of evidence [3].

Should any tool split the evidence during acquisition, for transmission or storage purposes, this collected evidence should be recompilable into the original source for examination purposes. To ensure forensic integrity, the tools used for splitting and recompiling the evidence should be able to verify the recompiled image against the original untouched source using a sufficiently collision resistant hash values.

3 Evidence Storage Formats

There is currently no universal standard for the format that digital evidence and any case related information is stored. This is due to the fact that there are no state or international governmental policies to outline a universal format. Many of the vendors developing forensic tools have developed their own proprietary format. With such a relatively small target market, it sometimes makes business sense for them to try and lock their customers into buying only their software in the future. There have been a small number of attempts at creating open formats to store evidence and any related metadata. This section describes the most common of these formats below.

3.1 Common Digital Evidence Storage Format

The Common Digital Evidence Storage Format (CDESF) Working Group was created as part of the Digital Forensic Research Workshop (DRFWS) in 2006. The goal of this group was to create an open data format for storing digital forensic evidence and associated metadata from multiple sources, e.g., computer hard drives, mobile Internet devices, etc. [4]. The format which the CDESF working group were attempting to create would have specified metadata capable of storing case-specific information such as case number, digital photographs of any physical evidence collected and the name of the digital investigator conducting the investigation. In 2006, the working group produced a paper outlining the advantages and disadvantages of various evidence storage formats [3].

Ultimately due to resource restrictions, the CDESF working group was disbanded in 2007 before accomplishing their initial goal.

3.2 Raw Format

According to the CDESF Working Group, “the current de facto standard for storing information copied from a disk drive or memory stick is the so-called “raw” format: a sector-by-sector copy of the data on the device to a file” [5]. The raw format is so-called due to the fact that it is simply a file containing the exact sector-by-sector copy of the original evidence, e.g., files, hard disk/flash memory sectors, network packets, etc. Raw files are not compressed in any manner and as a result, any deleted or partially overwritten evidence that may lay in the slackspace of a hard disk is maintained. All of the commercial and open source digital evidence capturing tools available have the capability of creating raw files.

3.3 Advanced Forensic Format

The Advanced Forensic Format (AAF) is an open source, extensible format created by S. Garfinkel in Basis Technology in 2006 [6]. The AAF format has a major emphasis on efficiency and as a result is partitioned into two layers; the disk representation layer which defines segment name used for storing all data associated with an image and the data storage layer which defines how the image is stored, be it binary or XML[7]. The format specifies three variants; AFF, AFD and AFM. AFF stores all data and metadata in a single file, AFD stores the data and metadata in multiple small files, and AFM stores the data in a raw format and the metadata is stored in a separate file [7].

3.4 Generic Forensic Zip

Generic Forensic Zip (gfmzip) is an open source project to create a forensically sound compressed digital evidence format based on AAF 3.3 [8]. Due to the fact that it is based upon the AAF format, there is limited compatibility between the two in terms of segment based layout. One key advantage that gfmzip has over the AAF format is that gfmzip seeks to maintain compatibility with the raw format 3.2. It achieves this by allowing the raw data to be placed first in the compressed image [7].

3.5 Digital Evidence Bag (QinetiQ)

The method for traditional evidence acquisition involves a law enforcement officer collecting any relevant items at the crime scene and storing the evidence in bags and seals. These evidence bags may then be tagged with any relevant case specific information, such as [1]:

- Investigating Agency / Police Force
- Exhibit reference number
- Property reference number
- Case/Suspect name
- Brief description of the item
- Date and time the item was seized/produced
- Location of where the item was seized/produced
- Name of the person that is producing the item as evidence
- Signature of the person that is producing the item
- Incident/Crime reference number
- Laboratory reference number

Digital Evidence Bag (DEB) is a digital version of the traditional evidence bag, created by Philip Turner in 2005 [1]. DEB is based on an adaptation of existing storage formats, with potentially infinite capacity. The data stored in a DEB is stored in multiple files, along with metadata containing the information that would traditionally be written on the outside of an evidence bag. There are currently no tools released that are compatible with the DEB format.

3.6 Digital Evidence Bag (WetStone Technologies)

In 2006, C. Hosmer, from WetStone Technologies Inc., published a paper outlining the design of a Digital Evidence Bag (DEB) format for storing digital evidence [9]. This format for storing is independent from the Digital Evidence Bag outlined in 3.5. The format emerged from a research project funded by the U.S. Air Force Research Laboratory. The motivation for this format was similar to the motivation for that described in 3.5, i.e., to metaphorically mimic the plastic evidence bag used by crime scene investigators to collect physical evidence such as blood, fibres, hairs etc. This format will be released publicly when complete.

3.7 EnCase Format

The EnCase format for storing digital forensic is proprietary to the evidence analysis tool of the same name. It is by far the most common evidence storage option used by law enforcement and private digital investigation companies for the acquisition of digital evidence from physical storage media [7]. Because of the proprietary nature of the format, along with the lack of any formal specification from Guidance Software [10], much remains unknown about the format itself. Some competitors to Guidance Software have attempted to reverse engineer the

format to provide an element of cross-compatibility with their tools [6]. EnCase stores a disk image as a series of unique compressed pages. Each page can be individually retrieved and decompressed in the investigative computer's memory as needed, allowing a somewhat random access to the contents of the image file. The EnCase format also has the ability to store metadata such as a case number and an investigator [6].

4 Court Admissible Evidence

Since the United States leads the way with the implementation of many standards in relation to evidence handling and the court admissibility of evidence, many other countries look to the procedures outlined by the United States in this area when attempting to create their own legal procedures [11]. As a result of this, much of the information available regarding the admissibility of digital forensic evidence into court cases is specifically tailored to the United States, but will influence law makers across the globe. Carrier [12] states that in order for evidence to be admissible into a United States legal proceeding, the scientific evidence (a category which digital forensic evidence falls under in the U.S.) must pass the so-called "Daubert Test" (as outlined below). The reliability of the evidence is determined by the judge in a pre-trial "Daubert Hearing". The judge's responsibility in the Daubert Hearing is to determine whether the methodologies and techniques used to identify the evidence was sound, and as a result, whether the evidence is reliable.

4.1 Daubert Test

The "Daubert Test" stems from the United States Supreme Court's ruling in the case of *Daubert vs. Merrell Dow Pharmaceuticals* (1993) [13]. The Daubert process outlines four general categories that are used as guidelines by the judge when assessing the procedure(s) followed when handling the evidence during the acquisition, analysis and reporting phases of the investigation, [12] and [13]:

1. *Testing* – Can and has the procedure been tested? Testing of any procedure should include testing of the number of false negatives, e.g., if the tool displays filenames in a given directory, then all file names must be shown. It should also incorporate testing of the number of false positives, e.g. if the tool was designed to capture digital evidence, and it reports that it was successful, then all forensic evidence must be exactly copied to the destination. The U.S. National Institute of Standards and Technology (NIST) have a dedicated group working on Computer Forensic Tool Testing (CFTT) [14].
2. *Error Rate* – Is there a known error rate of the procedure? For example, accessing data on a disk formatted in a documented file format, e.g., FAT32 or ext2, should have a very low error rate, with the only errors involved being programming errors on behalf of the developer. Acquiring evidence from an officially undocumented file format, e.g., NTFS, may result in unknown file access errors occurring, in addition to the potential programming error rate.

3. *Publication* – Has the procedure been published and subject to peer review? The main condition for evidence admission under the predecessor to the Daubert Test, the Frye Test, was that the procedure was documented in a public place and undergone a peer review process. This condition has been maintained in the Daubert Test [12]. In the area of digital forensics, there is only one major peer-reviewed journal, the International Journal of Digital Evidence.
4. *Acceptance* – Is the procedure generally accepted in the relevant scientific community? For this guideline to be assessed, published guidelines are required. Closed source tools have claimed their acceptance by citing the large number of users they have. The developers of these tools do not cite how many of their users are from the scientific community, or how many have the ability to scientifically assess the tool. However, having a tool with a large user base can only prove acceptance of the tool; it cannot prove the acceptance of the undocumented procedure followed when using the tool.

5 Legal Considerations of Network Forensics

Collecting network traffic can pose legal issues. Deploying a packet sniffing or deep packet inspection device, such as those outlined above, can result in the (intentional or incidental) capture of information with privacy and security implications, such as passwords or e-mail content, etc. As privacy has become a greater concern for computer users and organisations, many have become less willing to cooperate or share any information. For example, most ISPs will now require a court order before providing any information related to suspicious activity on their network [15]. In Europe, continental legal systems operate on the principle of free introduction and free evaluation of evidence and provide that all means of evidence, irrespective of the form they assume, can be admitted into legal proceedings [16].

5.1 Jurisdiction

One aspect of the use of search and seizure warrants in an Internet environment concerns the geographical scope of the warrant issued by a judge or a court authorising the access to the digital data. In the past, the use of computer-generated evidence in court has posed legal difficulties in common law countries, and especially in Australia, Canada, the United Kingdom and the USA. The countries are characterised by an oral and adversarial procedure. Knowledge from secondary sources is regarded as “hearsay evidence”, such as other persons, books, records, etc., and in principle is inadmissible. However, digital evidence has become widely admissible due to several exceptions to this hearsay rule [16].

6 Conclusion

While a number of the digital evidence bags outlined above are capable of storing captured network streams, the DEB itself offers little additional useful features

with respect to the investigation of P2P networks. As a result, a new P2P DEB is proposed capable of aiding in the identification and investigation of these networks. This P2P DEB is capable of storing network traffic, alongside specific P2P network metadata, based on on-the-fly analysis during the network capturing process. Building upon existing network packet recording tools, such as libpcap [17], a additional packet analysis tool can identify known P2P packets as well as record the categorisation and frequency of each type of captured packet. This additional metadata will aid in the identification of new, undiscovered P2P networks.

References

1. Turner, P.: Unification of digital evidence from disparate sources (digital evidence bags). *Digital Investigation* 2(3), 223–228 (2005)
2. Casey, E.: What does for ensically sound really mean. *Digital Investigation* 4(2), 49–50 (2007)
3. Common Digital Evidence Storage Format (CDESF): Survey of existing disk image storage formats. In: *Proc. Digital Forensic Research Workshop 2006* (September 2006)
4. Group, D.F.R.W.D.C.D.E.S.F.C.W (September 2006), <http://www.dfrws.org/CDESF/index.shtml>
5. The Common Digital Evidence Storage Format Working Group: Standardizing digital evidence storage. *Communications of the ACM* 49(2), 67–68 (2006)
6. Garfinkel, S.: Aff: a new format for storing hard drive images (2006)
7. Richard, G., Roussev, V., Marziale, L.: Forensic discovery auditing of digital evidence containers. *Digital Investigation* 4(2), 88–97 (2007)
8. Zip, G.F.: (April 2009), <http://www.nongnu.org/gfzip/>
9. Hosmer, C.: Digital evidence bag. *Commun. ACM* 49(2), 69–70 (2006)
10. Features, E.F.: (August 2009), <http://www.guidancesoftware.com/WorkArea/DownloadAs-set.aspx?id=671.GuidanceSoftware>
11. Science and Technology Committee: Forensic Science on Trial, 75–76 (2005)
12. Carrier, B.: Open source digital forensics tools: Thelegal argument. @stakeResearch Report (2002)
13. Supreme Court of the United States, *Daubert v. Merrell Dow Pharmaceuticals*: (June 1993), <http://supct.law.cornell.edu/supct/html/92-102.ZS.html/>
14. Computer Forensic Tool Testingpro-gram, U.S.N.I.o.S., Technology (August 2009), <http://www.cftt.nist.gov/>
15. National Institute of Standards and Technology: NIST Special Publication 800-86 Guide to Integrating Forensic Techniques into Incident Response. Create Space, Paramount, CA (2012)
16. Karyda, M., Mitrou, L.: Internet forensics: Legal and technical issues. In: *IEEE Second International Workshop on Digital Forensics and Incident Analysis, WDFIA 2007*, pp. 3–12 (2007)
17. McCanne, S., Leres, C., Jacobson, V.: *Libpcap* (June 2012)

A Research for Partition Restoration Techniques

Jaegung Namgung¹, Il young Hong¹, Jungheum Park¹,
Changhoon Lee², and Sangjin Lee^{1,*}

¹ Center for Information Security Technologies
Korea University, South Korea
{mgw_leader,hiy2009,junghmi,sangjin}@korea.ac.kr

² Department of Computer Science and Engineering,
Seoul National University of Science and Technology, Seoul, Korea
chlee@seoultech.ac.kr

Abstract. As the capacity of storage is gradually larger, most of the users logically use the storage devices dividing into multiple partitions. Therefore recovering partition stably from artificial partition concealing or partition damage is a very important issue. In this paper, we suggest a partition recovery algorithm which can conduct reliable and efficient analysis using the characteristics of each file system, if the partition is secreted or isn't sectioned due to damage of partition area.

Keywords: digital forensics, master boot record, volume boot record, boot record, super block, fat file system, ntfs, ext file system.

1 Introduction

Recently thanks to the improved safety of storage devices, Damage of storage due to physical defects is gradually decreasing, but types of malignity code making MBR as a main target like Distrack, Dropper, Userinit etc. and Cases of infection with malignity code are gradually increasing. Malignity codes making MBR as a target let PC boot to perform functions as a zombie PC or not boot by deleting all of the data in the MBR and they perform malicious acts losing data not dividing partition through the partition table. Because MBR generally do not backup data in a different area of them, the system cannot recognize all partition and loses all of the data that exists in the partition.

In terms of digital forensics, the existence of a partition can have a great effect on analyzing the data. The existence of the partition that contains the file system is very important, because we can get time information to create, modify, and run each data from the file system that exists in each partition and confirm the meta-information like filename containing the data, the path to the file etc. through the file system. Thus, if the system does not recognize the partition Due to damage of the partition information, we should be able to conduct analysis recovering partition. Also, In

* Corresponding author.

terms of digital forensics, it is quite meaningful to provide an environment that can be analyzed recovering partition exists in previous system or VM, backup file etc. However, studies for partition recovery lack a lot yet, existing tools for partition recovery are open to the public but they have many limitations.

Therefore, this paper presents a strategy to conduct meaningful analysis in terms of forensic by recovering partition damage caused by damage of MBR area or partition of the previous system and partition (VM, backup image etc.) that exists in the file of the current system.

2 Background

In general, partitions are generated by the user for the convenience of storing and managing data and managed by the system. The system saves generated size and location information of the partition etc. to Partition Table of the MBR (Master Boot Record). Partition Table can save the partition information for a total of four and each partition information is 16 bytes, at boot time, the system classifies partition through stored in the Partition Table attributes of each partition and information of size and location. Depending on the each partition is formatted by the user with any file system, at the start of the partition, VBR (Volume Boot Record) or Super Block is created. Generally, FAT file system and NTFS start partition to VBR and EXT file system starts partition to Super Block[1]. So VBR and MBR, Super Block etc. are closely related to partition in structure and can be utilized important clues, when we need to recover deleted or secreted partitions for forensic analysis[2]. Thus, in order to recover partition stably, it is necessary to understand VBR and MBR, Super Block, each file system exactly[3][4].

3 The Limits of the Existing Partition Recovery Tools

Partition recovery tools can be classified as recovery tool used for the purpose to recover physically damaged the partition on the system and tool used for the purpose to conduct analysis recovering partition virtually from the image of the recovery target.

The tools used for the purpose to recover physically damaged the partition are the Partition Recovery of EaseUS and TestDisk provided in open source, the forensic tools used for the purpose to analyze file system recovering partition virtually from the image are EnCase Guidance company and FTK of Access Data etc.

Partition Recovery and TestDisk show the search results to the user searching the Super Block and VBR being in the storage device of active system[5][6]. The user can enter the partition they want to recover of search results. The entered partition is recovered completely by being stored. As these tools operate on the active state, Integrity cannot be guaranteed and it is necessary for them to modify the contents of the partition table of the MBR newly every time in order to recover the partition for the analysis of file system. Therefore, from the perspective of forensic, these tools have disadvantages that it is difficult to use for analytical purposes. Forensic

professional tools, such as Encase and FTK provide more appropriate functions for analysis than other general recovery tools[7][8]. Forensic professional tools can try to conduct analysis recovering the partition virtually, after reading the system of the partition recovery target the form of images. Therefore, if we use Forensic professional tools, it is possible for them to ensure integrity. Further, there is an advantage that can be analyzed recovering or cancelling the recovery freely, because they analyze the file system recovering the partitions virtually. However, the partition recovery functions forensic professional tools support have limits. First, they do not judge the validity of the file system that exists in each partition, they do not remove duplicates of duplicate VBR and Super Block that exists by means of the features of the file system. Also, as VM user-created or the VBR and Super Block present in the backup may not be detected by the forensic professional tools, the restoration work through the recovery of the partition existing to the backup file is difficult. Thus, research is needed that can provide a variety of functions and the convenience of analysis, while ensuring the integrity.

4 Proposed Algorithm for the Effective Partition Recovery

4.1 Generating Factors of VBR and Super Block

There can be a large number of VBR and Super Block within the storage device depending on how the system and the user use them. Because VBR is generated in order to distinguish a single partition within the storage device, when storage devices divided into multiple partitions, VBR is generated by the number of partitions. Also, because Super Block includes most of the functionality of VBR and the role for separating partition in EXT file system, if a number of partitions exist, there can be a number of Super Block. In general, VBR and Super Block are newly created, if the user format the system and there can remain VBR or Super Block existed in the previous system, as it does not completely delete. In the case of Windows system, it can generate the VBR of a file unit, according to need and VBR can be included in the backup image or virtual image that the user generated for a particular purpose.

4.2 Search and Verification of the Partition

For the recovery of all partitions that exist, the search for VBR and Super Block having the information of each partition is required. Because file systems such as FAT, NTFS and EXT begin one partition by starting VBR, Super Block, they can find the beginning location of each partition through search of the VBR and the Super Block. But the VBR and the Super Block found by a simple search on may not be perfect as they are the trail used long before. Also, duplicates may occur due to the characteristics of each file system and data with a format similar to VBR or Super Block may be browsed because of false search. Therefore, in order to reduce duplicate and search VBR and Super Block that exist as a complete file system, we must seize the internal characteristics of VBR and Super Block and they should be used as signatures, and the procedure is necessary to verify the normalcy of the file system contained in VBR and Super Block that we search seizing the characteristics of each file system.

4.2.1 Search and Verification Method of the FAT Partition

Structure of the VBR that make up the partition on the system is started through the start of the sector. Generally since VBR is composed of 512-byte of sector size, we can search for the VBR on a sector-by-sector basis utilizing the characteristics of VBR. VBR starts to 0xEB, because it has Jump Code at offset 0 and go through the process of moving to the boot code. Offset of 3-10 provide a string of several formats we being able to verify intuitively the kind of file system that constitutes the VBR. The VBR, In general, to use the FAT32 file system, is the OEM Name and it has the string "MSWIN4.1" or "MSDOS5.0", in the case of FAT16, it has string "MSWIN4.1" or "MSDOS5.0. Therefore, it is possible to judge the file system used by the VBR and whether the VBR is normal through the string stored in this area. However, OEM Name territory does not play any role in addition to providing file system information by using the string. So, in the OEM Name area, even if what changes will be applied, there is no problem on the system. Thus, it is an area that can be easily forged and cannot be fully trusted. However, it is certainly useful string to search for the VBR, as it has left the string to formal way. Finally, offset 510 to 511 have a fixed value of 0xAA55 to signature items. In the case of value that exist in that location is not a signature, it can be used as a basis that can be excluded from the VBR. All of FAT file systems have an area of Root Directory basically, the sub Directory and all files are managed on the basis of the Root Directory.

VBR items needed to calculate the position of the Root Directory and it can be calculated through the item of Number of FATs and Reserved Sector Count of VBR, in Root Directory Cluster. It is possible to verify whether the VBR is normal, by judging Root Directory is actually at a calculated location.

Attribute area of Root Directory is item to manage the property of volume, Directory, file and in there, by normal value, hexadecimal value of 0x01, 0x02, 0x04, 0x08, 0x10, 0x20, 0x0F are used as a valid value. Therefore, if an invalid value is present, it is not Root Directory and the VBR is not normal VBR.

FAT32 indicates where FSInfo is stored, through item of FSInfo (File System Information). FSInfo stores simple information in the Reserved Sector area in order to make efficient use of FAT32. FSInfo has three signature such as Lead, Struct, and Trail and each value always has hexadecimal 0x41615252, 0x61417272, 0xAA550000 value. Therefore, we can verify whether the VBR is normal from Signature of FSInfo, if it does not have a normal value, we can judge that file system is not normal.

FAT32 saves copies at the back of sector by the number set to Boot Record Backup Sector item creating copy of VBR. That is, Same VBR exist, to the intervals of sectors by the set number. Therefore, if same VBR appear to the interval of the size set in the Boot Record Backup Sector, as they use same file system, VBR found in this location offer the convenience of the analyst by being displayed as a backup.

4.2.2 NTFS Partition Search and Verification Techniques

Like the FAT file system, NTFS constitutes the partition through VBR. VBR of NTFS and VBR of FAT file system use the same Jump Code from the offset 0 and utilizes same Signature from the offset 510-511. But, in the OEM Name, the string

"NTFS" and 0X20 value as the value of reserved area is used. Thus, it is possible to distinguish the file system used by the VBR through these differences. By the method described above, when detected VBR is configured with NTFS, it is possible to verify the partition, by comparing the contents of the sector from MFTMirr location and start position of the MFT of NTFS. NTFS manages the Directory and all of the files that exist in the volume through the MFT (Master File Table). MFT is an essential metadata of NTFS file and the location of it can be checked through the Start of MFT item of VBR. And MFT stores a copy for backup in the MFTMirr area. MFTMirr is able to confirm its position via the Start of MFTMirr of VBR. The MFT has size information of MFT Entry and Signature, in general, the Signature is the string of "FILE" and MFT Entry size is 1024 bytes. Thus, after the system using the NTFS moves the Start of MFT item to MFT in order to judge whether VBR is normal, it is possible to validate whether VBR is normal by confirming whether the correct values are to Signature items and Allocated Size of MFT Entry items. And as MFTMirr area is copied by same contents of the MFT's, normalcy of VBR can be verified.

4.2.3 Search and Verification Method of EXT3 / 4 Partition

Unlike VBR FAT and NTFS file system, EXT file system is capable of verification and search of partition through the Super Block that EXT includes internally. The layout of the EXT4 file system and there is a Super Block in first Block of each Block Group, Super Block occupies a space of 1 Block size. Block-size can be checked through the items of Block Size of Super Block and Block Size defines the size of the Size with a value of one of 0x00, 0x01 and 0x02. 0x00, value of Block Size means the 1KB and 0x02 means the 4KB.

The rest of the values cannot exist except these values. Therefore, if there is a value other than 0x00, 0x01, and 0x02, it can be excluded from the search for Super Block. In addition, Super Block has the signatures of two bytes like VBR. Unlike the signature 0xAA55 of VBR, Super Block has a 0xEF53 value from the offset of 56-57.

If the result of the partition retrieved via the same method as above is the Super Block, we can verify the partition through features of the EXT file system such as location of the Super Block value, the value of the location of the journal log block. EXT file system includes the Super Block continuously to Odd-Block Group including Block Group 0. The first block of odd all groups including Group 0 exists Super Block and normalcy of the file system can be verified through the confirmation of Magic Signature included in the each Super Block. Super Block of EXT file system is included first to the block group 0 and later, in the odd group, the Super Block repeatedly appears. Thus, it is possible to check the Magic Signature repeatedly in Super Block that exists at the odd block group and if the Signature are different, it can be judged as wrong Super Block. Also, since the Super Block repeatedly exists, it is possible that the same Super Block is multiple. But duplicates can occur by the number of Super Block, because professional tools forensics find the Super Block without considering the characteristics of the Super Block. Thus, same Super Block should not be duplicated by comparing the Super Block of each block group that contains the Super Block. EXT4 file system and EXT3 file system, include journal

log block in the block group specified through the Super Block. Journal log has 0xC03B3998 value as the signature and normalcy of file system is judged through this signature.

4.2.4 Search and Verification Techniques of VM and Backup Image Partition

VM and backup image have a file system via an inside virtual machine and include VBR or Super Block depending on the kind of file system. Therefore, VBR and Super Block can be searched and verified through search and verification techniques described above. However, VM and backup image exist in the form of the file in a specific partition and their VBR or Super Block may start in the middle of sector according to the feature of the application controlling the file. That is, VBR and Super Block of file form cannot be searched to general search of the unit of sector.

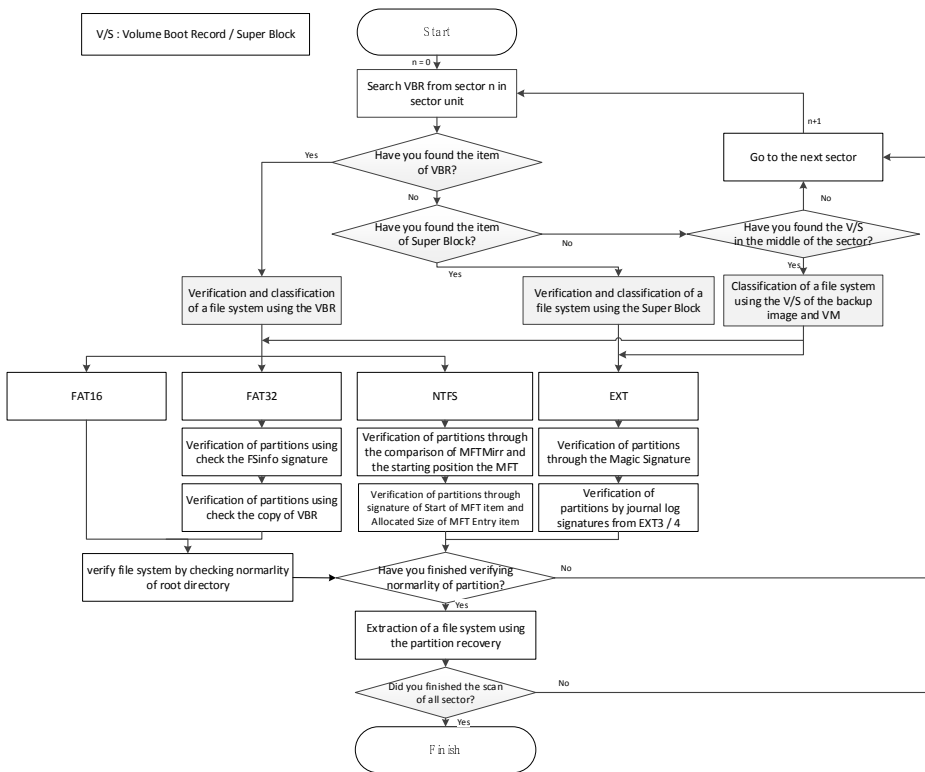


Fig. 1. Proposed algorithm for the recovery of the effective partition

Therefore, Signature search of the unit of byte should be conducted in order to search, verify and recover all partitions including VM and backup image. Also, VM and backup image should be recovered, the position that VBR of inside file starts considered as the 0th sector logically, because they exist as a virtual machine within a file.

4.3 Algorithm to Propose

Figure 1 shows the final algorithm for the recovery of partition proposed in this paper. Partition recovery algorithm in this paper can reduce absence and mistake of research by including the detection and validation techniques for each file system partition previously proposed. Also, it has enabled the effective analysis and recovery of partition without repetitive recovery of same partition by analysts checking for duplicate of partition to include same information. I made partition of backup image and the VM searched through search of the unit of byte.

5 Results and Experiment of the Proposed Algorithm

The proposed algorithm in this paper can ensure the integrity unlike the existing partition recovery tool and preserve the original state of Partition Table of the MBR furthermore, it can check the duplicate of search results unlike forensic special tools, verify file system and search partition existing in VM. Table 1 shows the experiment results with the target image of analysis based on the proposed algorithm. Experimental results show some summarized among a number of results and I ordered them as the form to indicate search and verification of partition for each file system. As shown in Table 1, forensic tools and recovery of existing provide function for search and recovery of VBR and Super Block, simple analysis. But, they do not provide other functions like check of duplicate for efficient analysis, verification of file system, search of partition included inside of VM, etc.

Table 1. Comparison of experimental results of the proposed algorithm with existing tools

VBR and SuperBlock in storage		Existing partition recovery tools			Forensics tools			The proposed algorithm		
Classification	Sector location	S	D	V	S	D	V	S	D	V
NTFS	343,728	O	X	X	O	X	X	O	O	O
	2,237,600	O	X	X	O	X	X	O	O	O
FAT	26,606,232	O	X	X	O	X	X	O	O	O
	26,606,238	O	X	X	O	X	X	O	O	O
EXT	52,430,848	△	X	X	O	X	X	O	O	O
	61,107,072	△	X	X	O	X	X	O	O	O
VM	28,539,838	X	X	X	X	X	X	O	O	O

※ S – Search whether; D – Duplicate check; V – Verification of file system.

6 Conclusions and Future Works

If users hide artificially partition or the system cannot recognize the partition because of deletion of partition table due to malicious code, we cannot obtain meta-information and all data from file system that exist in the partition. Extraction of some

data is possible by carving technique but when a file is broken, it is hard to fully extract the file and analyze meaningfully as we can obtain meta-information. Therefore, it is very important to interpret the file system in the partition by restoring the partition from the point of view of forensic. Existing forensic tools provide the ability to be able to interpret the file system at a position obtained by searching the signature of the Super Block and VBR in the storage device. However, its function is very limited and if there are many the Super Block or VBR, they have difficulties to recover one by one in order of analysis. Further, the Super Block and VBR in the VM and the backup file are excluded from the search target if the starting position is not the start area of a sector. Thus, in order to analyze the file system in the particular file, they have disadvantage that should analyze the file by reading directly. This paper proposed the algorithm for meaningful analysis to files in the recovered partition by recovering all recovery possible partitions that exist within the area assigned and non-assigned and it can reduce the hassles of the analysis by reducing duplicate of recovery target partition through analysis of the features of VBR, Super Block and each file system. In future, if we develop automated tools based on the algorithm proposed and utilize them for analysis, it will be possible to reduce the effort of recovery through existing tools and it will give much help to the digital forensic analysis.

Acknowledgments. This work was supported by the IT R&D program of MOTIE/KEIT. [10035157, Development of Digital Forensic Technologies for Real-Time Analysis]

References

1. Carrier, B.: File System Forensic Analysis. Addison Wesley Professional (2005)
2. Fairbanks, K.D.: An analysis of EXT4 for digital forensics. *Digital Investigation* 9, 118–130 (2012)
3. Master boot record, http://en.wikipedia.org/wiki/Master_boot_record/
4. Volume boot record, http://en.wikipedia.org/wiki/Volume_Boot_Record/
5. EaseUS, <http://www.easeus.com/partition-recovery/>
6. TestDisk, <http://www.cgsecurity.org/wiki/TestDisk/>
7. Guidance Software homepage, <http://www.guidancesoftware.com>
8. FTK, <http://www.accessdata.com/products/digital-forensics/ftk/>

Definition of Evaluation Index Model for Network Management System

Fei Xu¹, Jinqiao Shi¹, K.P. Chow², Xiaojun Chen¹, and Peipeng Liu¹

¹Institute of Information Engineering, Chinese Academy of Sciences,
100093 Beijing, China

²Department of Computer Science, University of Hong Kong
Hong Kong, China

{xufei, shijinqiao, chenxiaojun, liupeipeng}@iie.ac.cn,
chow@cs.hku.hk

Abstract. The growing of network applications and the advances in network services demand for more complex and flexible network management techniques and systems. We proposed a three-dimensional evaluation theory model for network management system in our previous work [6]. In this paper, we formally define a specific network and security management system. Based on the definition, we give the detail and formal definitions of several critical and important evaluation indexes. We believe that the evaluation index definition for network management evaluation can reflect more accurately the status of the network in today's network environment.

Keywords: network management, evaluation index, three-dimensional model.

1 Introduction

The widespread network applications and network services bring a new challenge to network management along with the convenience they provided. Network users start to request healthier network environment and timely responses while prefer to keep the network performances uncompromised. It requires the network management system to provide a variety of management functions without visibly degrade any network performance. However, the workload of the network management system becomes heavy in such a complex network environment because of the need to execute many management tasks or manage a large number of network equipment [1,2]. On the other hand, the application of network management system may lead to unintended outcome of the network [3,4]. The pros and cons of a network management system should both be considered during an evaluation. The future network management system evaluation models are required to be more flexible, expansible and systematic than offered by the existing evaluation methods [5].

We have proposed a three-dimensional evaluation model for network management system which takes cost, impact, as well as operation evaluation into consideration simultaneously in our previous work [6]. As a detailed elaboration for the proposed

model, we define an evaluation index which is based on the proposed evaluation index architecture. The rest of the paper is structured as follows: Section 2 gives the formal definition of specific network; Section 3 presents the definition of security management system. The definition of evaluation indexes are given in Section 4 and Section 5 summarizes the contribution of the paper.

2 Formal Definition of Specific Network

We define a specific network (N) as follow:

$$N=(U, S, M, E)$$

where U is the user set. For a user u_i , the property is expressed as

$$A^{u_i}(t)=(\alpha_1^{u_i}, \alpha_2^{u_i}, \dots, \alpha_n^{u_i})$$

Hereinto, $\alpha_1^{u_i}, \alpha_2^{u_i}, \dots, \alpha_n^{u_i}$ are used to represent various attributes of the user u_i . User label generation function processes the user properties and gets the user tags.

S is the service set. For a service S_i , and its property is expressed as

$$A^{S_i}(t)=(\alpha_1^{S_i}, \alpha_2^{S_i}, \dots, \alpha_n^{S_i})$$

Hereinto, $\alpha_1^{S_i}, \alpha_2^{S_i}, \dots, \alpha_n^{S_i}$ are used for representing different attributes of the service S_i . Service label generation function processes these properties and gets the service tags.

M is the intermediate network equipment, which includes the deployed network equipment and installed software in order to achieve specific network management objectives. For every equipment m_i , its property is expressed as:

$$A^{m_i}(t)=(\alpha_{cost}^{m_i}, \alpha_{delay}^{m_i})$$

Hereinto, $\alpha_{cost}^{m_i}$ is used to represent the cost of deployment of the equipment and $\alpha_{delay}^{m_i}$ is used to represent the network delay due to the deployment of the equipment.

E is the situation of the network connection. For a user u_i , it represents the number of connections the user requests, irrespective of whether the connection is successful and fail. Given a service S_i , it indicates the success rate of the child services of the service.

3 Formal Definition of Security Management System

The formal definition of the security management system is as follow:

$$M = (C, P, S, O)$$

where C is the cost of software and hardware of the security management system;

P is the management strategy set, which consists of a series of management strategies;

S is the subject of access, the user label generated by the user label generating function;

O is the object of access, the service label generated by the service label generating function.

We can define the request management action to the user as follow:

$$Action = match(P, < S, O >)$$

Hereinto, the *match* function is used to judge whether the service requested by the user matches the administration strategy configured in the operating environment, and returns either false or true.

If the *Action* is of value **true**, it means that the request action should be under control, otherwise it should be released.

4 Formal Definition of Evaluation Indexes

In this section, we give the formal definitions of several critical and important index factors in network management system. The proposed index factor calculations can comprehensively and effectively reflect the network environment situation.

4.1 Mismatch Ratio

Mismatch ratio is the rate of unmatched data packets over all processed data packets in user or service classification, i.e., in a given period of time, the percentage of unmatched or neglected data packets of a network monitoring device. This ratio reflects the ability of a monitoring device. The mismatch ratio of a network monitoring device is measured by the following formula:

$$P = \sum_{i=1}^k \frac{N_{drop}(t_i, t_{i+1})}{N(t_i, t_{i+1})} / k$$

where P represents the average mismatch rate of the network monitoring device, $N_{drop}(t_i, t_{i+1})$ represents in the time segment (t_i, t_{i+1}) , the unclassified or neglected number of data packets, k represents all the packets monitored in the time segment (t_i, t_{i+1}) , and $N(t_i, t_{i+1})$ is the interval time between the two sample time segments.

4.2 Stability

Stability is the ability of continue functioning of a network monitoring device in an abnormal network environment change (for example, an abrupt flush of network flows). Stability is a numerical measure, the larger the stability is, the stronger the monitoring device's ability to adapt the abrupt change of the networking environment. It can be measured by the following formula:

$$W_t = 1 - \frac{\sum_{t=1}^k N_d(t_i, t_{i+1})}{N}$$

where W_t represents the stability in time segment T , $N_d(t_i, t_{i+1})$ is the count that the device cannot function properly in the time segment, N is the count of the abrupt environment change in the time segment, and the formula $t = \sum_{t_{i+1}-t_i}^k$ represents in the given time segment t , the ratio that the network device can function properly without being disrupted by any sudden network change.

4.3 Accuracy Rate

The percentage of false positives P_{error} usually represented by the ratio of the number of forecast fail times N_{fail} and the number of total forecast times N_{sum} , i.e. $P_{error} = N_{fail} / N_{sum}$.

The percentage of accuracy positives is represented as:

$$P_{accuracy} = 1 - P_{error} = 1 - N_{fail} / N_{sum}$$

The false positives for a particular user service can be represented as

$$P_{error}' = P_{e_user} + P_{e_serve} + P_{e_proces}$$

where P_{e_user} represents the error rates to a user, P_{e_serve} represents the error rates to service resource, $P_{e_process}$ represents the rates of error process. Total false positives can be represented by:

$$P_{e_sum} = E(P_{error}) = \sum P_{ei} \times P_i$$

P_{ei} represents the false positive of a particular user service, and P_i represents the rate of the requests produced, P_{e_sum} is the sum of the product of the false positive of a particular user service that may be appeared in network and the rates it appears. The accuracy rate can be represented by the average false positive accessed by analysis of measurement as follow:

$$P_{accuracy} = 1 - \overline{P_{e_sum}} = 1 - \overline{E(P_{error})}$$

4.4 Other Index Definition

Serviceability is the characteristic of the network information that can be visited by authorized entity and be used in accordance with the demand. When the network information service is requested to access the authorized entity, it can provide the service independent of whether the network part is damaged or degrade.

The percentage serviceability P_{us} is usually measured by the ratio of the normal use time and the total run time as follow:

$$P_{us} = (T_{run} - T_{stop}) / T_{run}$$

Where T_{run} is the total network run time, T_{stop} is the time of the network failure, and the network serviced work time is:

$$T_{work} = T_{run} - T_{stop}$$

This is a simple but efficient way to value the serviceability.

Delay of the network equipment is the processing delay when the equipment classifies the network translation data, i.e. the time that the network monitoring equipment needed to classify the network flow according to the rules of different user classification and service classification plus the network monitoring equipment's processing to the total network translation rate.

The delay of the network monitoring equipments is usually represented as follows:

$$T_q = \begin{cases} 0, v \geq u \\ \frac{u}{v^2 - vu}, v < u \end{cases}$$

T_q is the average classification delay of the network monitoring equipments, v is the average classification rate of the network monitoring equipment, and u is the average translate rate of the network.

Maintenance can be measured by error maintenance and update in demand. Error maintenance is the works required to repair and then recover the equipment to normal function of the system when the system appears malfunction. Update in demand maintenance is the workload required to maintain the proper function when the equipment appears updating to demand or being removed from the system.

The workload is usually measured by the amount of code (lines of code) and is represented as follow:

$$N = p_{err} \cdot N_{err} + p_{update} \cdot N_{update}$$

p_{err} is the probability that an error appears, N_{err} is the amount of code to change the error, p_{update} is the probability that an update appears; N_{update} is the amount of code for this update.

Ratio of the system function completed is the amount of coverage of system's network management functions to the standard network management functions and system planning functions, i.e. which functions the system has achieved, and which functions are not available.

The indicator defines the amount of system functions that has been achieved. Ratio of the system function completed can be represented as follows:

$$P_{cov} = N_{suc} / N_{sum}$$

N_{suc} is the number of functions have achieved, N_{sun} is the number of functions the system should have. The indicator reflects the completeness of the system.

Response time is the time from the system receives the network alarm to the time when filter control strategy is developed, until the strategy is executed correctly. The indicator shows the extent of the system to meet the specific area of network management needs.

Response time includes warning time, strategy development time and the policy deployment time. It can be expressed by the following formula:

$$T = t_{warning} + t_{decide} + t_{deploy}$$

$t_{warning}$ is the warning time, t_{decide} is the decision time, t_{deploy} is the deployment time. This indicator reflects the timeliness of the system when an alarm occurs.

5 Conclusion

This paper proposes the formal definition of a specific network, the formal definition of security management system, and the formal definition of several evaluation index factors for network management system evaluation model. Combined with the concept of cost evaluation and impact evaluation, the index system gives a collective evaluation of the network management system, which provides different opinions of whether such system should be deployed and how. The proposed indexes system can evaluate the network management system more comprehensively and effectively.

References

1. Sugauchi, K., Miyazaki, S., Covaci, S., Zhang, T.: Efficiency Evaluation of a Mobile Agent Based Network Management System. In: Zuidweg, H., Campolargo, M., Delgado, J., Mullery, A. (eds.) IS&N 1999. LNCS, vol. 1597, pp. 527–535. Springer, Heidelberg (1999)
2. Que, W.K., Zhange, G.Q., Wei, Z.H.: Model for IP Network Synthetically Performance Evaluation. *Computer Engineering* 34(8) (2008)
3. Cappelli, R., Maio, D., Maltoni, D., Wayman, J.L., Jain, A.K.: Performance Evaluation of Fingerprint Verification Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(1) (2006)
4. Fang, B.X., Guo, Y.C., Zhou, Y.: Information content security on the Internet: the control model and its evaluation. *Information Sciences Science China* 53(1), 30–49 (2010)
5. Luo, F., Song, M., Ren, Z.J., Pan, Y.F., Song, J.D.: An Improved Evaluation Method in Mobile Network Management System. *Journal of Beijing University of Posts and Telecommunications* 27(4) (2004)
6. Xu, F., Shi, J.Q., Chen, X.J., Chow, K.P.: A three-dimensional evaluation model for network management system. In: *Information Technology and Quantitative Management, ITQM 2013* (2013)

Investigating and Measuring Capabilities of the Forensics File Carving Techniques

Khawla Alghafli, Andrew Jones, and Thomas Martin

Khalifa University of Science, Technology and Research
Khawla.alghafli@kustar.ac.ae

Abstract. File carving is a type of digital forensics recovery technique which focuses on recovering files from digital media without using file system metadata. This technique can be used in several situations such as recovering deleted files or recovering files from storage media with corrupted or unknown file systems. This paper explores and discusses the existing theory of file carving techniques. We conduct experimental testing for some of the current state of the art carving tools. These experiments will measure various criteria such as precision, recall and overall system performance.

Keywords: Digital forensics, data recovery, file carving, fragmentation.

1 Introduction

In the last few decades, digital devices are involved in most aspects of our life. Consequently, many crimes have moved to take advantage of these devices. The aim of digital forensics is to find the evidence of computer crimes that would be accepted in court.

Operating systems allow users of the digital devices to create, modify and delete files. The important tasks of the file system management module in the operating system are storing of the files in the system and locating them when they are needed. The storing of the file can be done in two possible manners depending on the available free space. These manners are either in contiguous memory locations or non-contiguous memory locations. The file system keeps a record of the memory location that each file in the system occupies.

Data recovery is a primary element of digital forensics. There are cases where we need to recover files from a system which has a corrupted or re-formatted file system. File carving is a data recovery technique that recovers files from storage media based on the file content and structure without using file system metadata [1].

The purpose of this paper is to provide a discussion of the current state of the art of file carving techniques. We will begin by providing a background review of available carving techniques. The positive and negative aspects of each technique will be highlighted. Moreover, we have designed experiments to evaluate the performance of the current state of the art of tools for file carving, namely Scalpel [2], Photorec [3] and Access Data Forensics Tool Kit (FTK) [4].

2 Background Review of File Carving Techniques

2.1 Header-Footer or Header-Maximum File Size Carving Technique

The first generation of carving techniques included header-footer and header-maximum file size carving techniques. The header contains the first sequence of bytes that identify the starting point of a file. For example, the header of JPEG files is `\xFFD8`. The footer contains a sequence of bytes that identify the end of file. For example, the footer of JPEG files is `\xFFD9`.

In this technique, the carver has a database of headers and footers for specific file types [2]. It retrieves files by searching for the pattern of a header and marks it as the starting point of a file. It then searches for the pattern of the corresponding footer. All the sequence of bytes between the header and the footer are carved as a file. In the cases where some file types do not contain footer, it uses the file size to determine the end of file. An example of a file type that has no footer pattern is the BMP file. The size of BMP file follows the header pattern [5]. The carver will identify the header as starting point and the end point is the address of the header plus the size of the image.

This technique has several problems and limitations. The first one is that it assumes that the file is not fragmented and the file bytes exist in sequential order between the header and the footer. The actual file may not exist in sequential order, in which case the carved file will not be completely correct. The second problem is that when the header or footer pattern is small, the probability that false positives occur will be high. This is because the short pattern can be found in the data set many times by coincidence where it is not a header/footer.

There are many carving tools that are based on this technique. For example, the open source software Foremost has been used widely for many years. It was developed by the US Air Force Office of Special Investigations and the Center for Information Systems Security Studies and Research [6].

2.2 File Structure Based Carving Technique

The second generation of carving techniques is file structure based carving. This technique does not depend only on the header and footer but rather it uses the elements of the internal structure of the file. Fig. 3 shows an example of the internal layout of the PDF file. In this example the carver uses header, footer, string identifiers and the size of each segment that follows the string identifier to carve PDF files.

This technique produces better results than header – footer carver. By this we mean it will produce less false positive results. This is because this technique does not depend only on a small pattern of a header or footer which can appear many times in the data set where it is actually a part of file data not a header or a footer, but also it uses extra information to validate the carved file.

There are known carving tools that use the file structure to carve files such as Foremost [6] and ReviveIT (revit) [7].

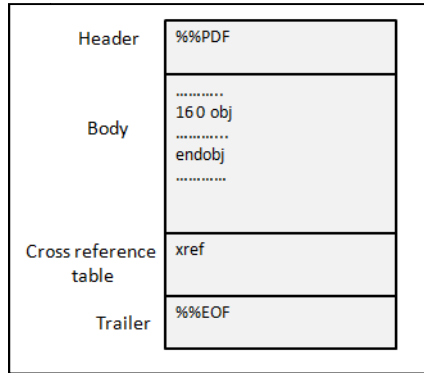


Fig. 1. PDF file structure

2.3 Smart Carving Techniques

Smart carving technique is based on the idea of the file is stored in the physical media in blocks. The block size varies depending on the type of storage media and file system. The file will be stored in a number of blocks determined by dividing the file size by the block size. Thus, if any fragmentations occur, it will be according to this number of blocks. This technique consists of investigating and verifying the blocks in the storage media which are related to the file. It determines that the block is related to this file or not by looking for the data inside the block and performing some statistical calculations. In this way, it attempts to recover fragmented files.

One approach that is based on smart carving technique is the Bifragment gap carving [1]. This approach works on the scenario where the files are fragmented into two fragments and contain known headers and footers. To carve a file by correctly removing data that was placed between the first fragment and the second fragment, S. L. Garfinkel proposed to place a gap between the header and the footer as shown in Fig. 6 and as follows:

- Let s_1 be the header of the file and the start of the first fragment and let e_1 the end of the first fragment.
- Let s_2 be the start of the second fragment and let e_2 be the end of the second fragment and the file footer.
- Let g be the length of the gap between e_1 and s_2 .
- Start with $g = 1$ and then try all values of s_2 and e_1 that produce $g=1$. For each value of s_2 and e_1 validate result if it is a valid file or not using various validation techniques proposed in [1].
- Increment g by 1 each time and validate all combinations of e_2 and s_1 that produce g .
- Stop when $g = e_2 - s_1$.

The limitation of this approach that cannot carve nonlinear fragmentation because it assumes the sequence of the fragments is the same in the original file. Another limitation is that it will not carve files that are fragmented into more than two fragments.

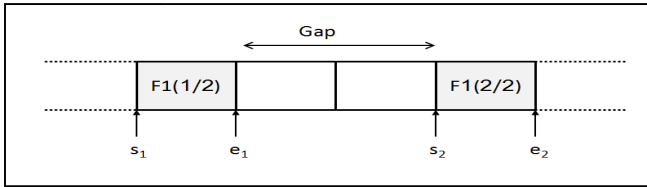


Fig. 2. Bifragment Gap Carving technique

3 Evaluating the Current State of the Art of the Carving Tools

The tools that will be tested are Scalpel, FTK and Photorec.

3.1 Evaluation Criteria

Any carving tool can produce three types of results when carving any digital media, which are:

1. True Positive (tp): A file which is present in the digital media and correctly recovered.
2. False positive (fp): A file which is not correctly carved and indicated as a positive result.
3. False negative (fn): A file which is available in the digital media but the carver was not able to recover it.

The evaluation criteria that will be to measure are the following:

1. Precision: indicates that the recovered files are correct and relevant [8].
2. Recall: the percentage of target files that were recovered [8].
3. Reliability: how successful the carver is in retrieving files types that it claims to support [9].

Precision is defined as the ratio of the number the recovered files that the system extracted correctly over the total number of recovered files [10] :

$$Precision = tp / (tp + fp) \tag{1}$$

Recall is defined as the as the ratio of the number the recovered files that the system extracted correctly over the total number of files available in the digital media [10]:

$$Recall = tp / (tp + fn) \tag{2}$$

The precision and recall will be combined to measure the overall performance of a file carver. This is called the F measure and it is calculated as follows [10]:

$$F_{measure} = \frac{1}{a \frac{1}{P} + (1-a) \frac{1}{R}} \tag{3}$$

where P is precision, R is the recall and α is the factor which determines the weighting of the precision and recall in the overall performance.

In this evaluation we want the same weight for the precision and recall. Thus, the value of α is 0.5. Therefore, the equation of F measure can be simplified as follows:

$$F_{measure} = 2PR/(R + P) \quad (4)$$

The false negative results can be classified into two types, which are supported false negative (sfn) and unsupported false negative (usfn) [9]. The supported false negative is the file of a type that the carver claims that it can recover, and yet the carver was not able to. The unsupported false negative is a file with a type that the carver does not claim that it has ability to recover. The reliability of the file carver will be calculated using the following equation [9]:

$$Reliability = \frac{\text{no.of supported files in the dataset} - \text{sfn}}{\text{no.of supported files in the dataset}} \quad (5)$$

3.2 Forensics Images to be Used in the Testing Process

We used disk images with a known layout of the memory locations of each file. This is because we want to compare the results of carving tools against what are the available files in the image. Luckily, there are a number of forensic images with a known layout.

The first image was created by Nick Mikus [11]. It is based on the FAT32 file system but the file system metadata was corrupted. It contains contiguous allocated and deleted files and one JPEG file with a corrupted header. We will call this image FAT32. The second image was also created by Nick Mikus [12]. It is based on the EXT2 file system but the file system metadata was corrupted. It contains 10 files and 9 out of 10 are fragmented. We will call this image EXT2. The third image is DFRWS 2006 carving challenge image [13]. The aim of this challenge is to develop file carving algorithms that have the ability to recover more files without incurring false positive results. This image has no file system and it contain 32 different file types and most of them are fragmented. We will call this image DFRWS.

3.3 Testing Procedures

Each tool was tested by executing it on each forensic image. The results were classified by true positive and false positive. The classification was done as follows:

1. Calculate the Message Digest Algorithm (MD5) of all recovered files. If any equals the MD5 of a file in the image and both of them are identical this result will be counted as true positive.
2. J. Metz, B. Kloet, and R. J. Mora in [14] said that determine the file is true positive or not just by matching the MD5 is flawed. This is because some of carved files may contain addition zeros after the end of files mark compared to the original files and when the MD5 is calculated the MD5 digest is different. Thus, we will compare the hex dump of the carved file to the original file. If there are additional zeros or less after the end of the carved file, we will count this result as true positive.

3. The remaining results are counted as false positive.
4. The number of false negative will be calculated using the following equation:

$$fn = \text{total No. of files in the image} - tp \quad (6)$$

3.4 Testing Results

The results of testing Scalpel, FTK and Photorec are shown in Table 1, Table 2 and Table 3 respectively. The overall averages of testing three carving tools are shown in Table 4.

Table 1. Scalpel testing results

Image name	TP	FP	FN	Recall	Precision	$F_{measure}$	No. of Supported Files	SFN	Reliability
FAT 32	5	221	10	0.33	0.02	0.04	13	8	0.38
EXT2	1	38	9	0.10	0.03	0.04	7	6	0.14
DFRWS	4	1185	28	0.13	0.00	0.01	28	23	0.18

Table 2. FTK testing results

Image name	TP	FP	FN	Recall	Precision	$F_{measure}$	No. of Supported Files	SFN	Reliability
FAT 32	10	20	5	0.67	0.33	0.44	10	0	1
EXT2	0	0	10	0.00	0.00	0	10	10	0
DFRWS	0	0	32	0.00	0.00	0	26	26	0

Table 3. Photorec testing results

Image name	TP	FP	FN	Recall	Precision	$F_{measure}$	No. of Supported Files	SFN	Reliability
FAT 32	13	3	2	0.87	0.81	0.84	10	2	0.80
EXT2	10	1	0	1.00	0.91	0.95	10	0	1.00
DFRWS	16	19	16	0.50	0.46	0.48	32	16	0.50

Table 4. Overall average of testing carving tools

Tool name	Recall	Precision	$F_{measure}$	Reliability
Scalpel	0.19	0.02	0.03	0.24
FTK	0.22	0.11	0.15	0.33
Photorec	0.79	0.73	0.76	0.77

3.5 Discussion of the Results

The results of the previous section are quantitative. We can say the precision of tool A is higher than tool B but we cannot state if the precision of tool A is perfect, good or bad. Consequently, there is a need to map quantitative results into qualitative results. There is no standard way of how to map quantitative results of file carver into qualitative results. S. J. J. Clad in his thesis defined a way to map descriptive score of carving results to four normative scores [9]. Our proposed method to map recall, precision, reliability and overall performance score is illustrated in Table 6:

Table 5. Mapping quantitative results of testing to qualitative results

Quantitative result range	Qualitative result
0.00 - 0.20	Bad
0.21 - 0.40	Mediocre
0.41 - 0.60	Good
0.61 - 0.80	Very good
0.81 - 1	Perfect

Scalpel scored a rating of bad with regard to recall and precision. Therefore, the $F_{measure}$ was also bad. The reliability of Scalpel was mediocre. FTK produced better results than Scalpel. Testing FTK on FAT image with no fragmented file produce a good $F_{measure}$ with a value 0.44. However, this result was dropped off when testing images with a fragment file. FTK cannot recover any fragmented files. The average recall was mediocre and the precision was bad. This leads the FTK to have bad measure. The reliability of FTK was mediocre. Photorec had the best results among these three tools. Both of the recall and precision have a score of very good. Consequently, $F_{measure}$ is also very good. Also, the reliability of Photorec is very good.

4 Conclusion

File carving is an essential part of forensic data recovery in any cases involving deleted data, corrupted file system, re-formatted file systems or unknown file systems. This paper presents the existing theory of file carving techniques. We conducted experimental testing to measure capabilities three of the current state of the art carving

tools, namely Scalpel, FTK and Photorec. The results were that the Photorec achieved the best results and its overall performance was considered to be very good.

Finally, there is a need for a research in improving validators to reduce the number of false positive results which will lead to improved carving precision. Also, there is a need for solutions to handle the problem of carving fragmented files.

References

1. Garfinkel, S.L.: Carving contiguous and fragmented files with fast object validation. *Digital investigation* 4(1), 2–12 (2007)
2. Richard III, G.G.: Scalpel: A Frugal, High Performance File Carver. In: *Digital Forensics Research Workshop*, New Orleans, LA, pp. 1–10 (2005)
3. PhotoRec, Digital Picture and File Recovery (January 2013), <http://www.cgsecurity.org/wiki/PhotoRec>
4. Forensic Toolkit (FTK) 3.4.1 Download Page (December 2012), <http://www.accessdata.com/ftk-3-4>
5. Wouters, W.: BMP Format. Clean Coding Company, Tech. Rep. v1.1 (1997)
6. Foremost (2012), <http://foremost.sourceforge.net/>
7. reviveit. Online (March 2013), <https://code.google.com/p/reviveit/>
8. Grossman, D.A., Frieder, O.: Introduction. In: *Information Retrieval Algorithms and Heuristics*, ch.1, pp. 1–8. Springer, Netherland (2004)
9. Kloet, S.J.J.: Measuring and Improving the Quality of File Carving Methods. MSc thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, The Netherlands (2007)
10. Manning, C.D., Schütze, H.: Lexical Acquisition. In: *Foundations of Statistical Natural Language Processing*, ch. 8, pp. 265–314. MIT Press, Cambridge (1999)
11. Mikus, N.: Basic Data Carving Test #1 (March 2005), <http://dftt.sourceforge.net/test11/index.html>
12. Mikus, N.: Basic Data Carving Test #2 (March 2005), <http://dftt.sourceforge.net/test12/index.html>
13. DFRWS 2006 Forensics Challenge File Image Details (2006), <http://www.dfrws.org/2006/challenge/>
14. Metz, J., Kloet, B., Mora, R.J.: Analysis of 2007 DFRWS Forensic carving challenge, Hoffmann Investigations. Tech. Rep., Netherlands (2007)

Application for Reversible Information Hiding in Multiple Secret Images Sharing Based on Shamir's Scheme

Chyuan-Huei Thomas Yang, Che-Lun Chung, Yu Min Lin, and Song Yong Fan

Dept. of Information Management, Hsuan Chuang University, Hsinchu, Taiwan
{chyang, MD1004017, MD1014007, MD1014016}@hcu.edu.tw

Abstract. We proposed an application for reversible information hiding in multiple secret images sharing based on Shamir's method. Many researchers develop secret image sharing method with Shamir's scheme in one image. We mixed the multiple secret images with one host image and employ the modified Shamir's method to share the participants. Lagrange method is applied with enough numbers of the shadow images in the retrieval step. The experiment results demonstrate our method can retrieve secret images lossless.

Keywords: reversible information hiding, image sharing, Shamir's scheme.

1 Introduction

During these few years we face the digital media resources spread broadly and wildly. It becomes more and more important to find the techniques to protect copyright and authorized material. The goal of secret image sharing is a vital method to protect digital images. The secret image sharing methods generate shadow images from one secret image then distribute to participants [10]. If we have enough numbers of shared shadow images we may retrieve the original secret image. Several authors have modified or extended the original Shamir's method for the secret image sharing. Some authors proposed their methods create with Shamir's method [1, 3, 10, 11], with Galois Field [2, 4, 5~7], or without Shamir's method [8, 9].

Firstly we are going to describe the secret image sharing method with Shamir's method. Basically, it takes parts of information from a secret image to generate n shadow images then distribute to people whom assigned to be owned. The t owners may recover the original secret image, where $t \leq n$. It is called (t, n) threshold scheme for secret image sharing. Thien and Lin [10] proposed a method (t, n) threshold scheme. Each shadow image is smaller than the secret image in their method. It gives the benefit to process the shadow images when stores, transmits, or processes image hiding. They also gave a reversible version to overcome the losses of 251~255 grayscales when applied Shamir's algorithm. Since their method is simple and lossless, Yang et al [11] the discrete Haar wavelet transform to reduce the secret image to its quarter size firstly, i.e. the level one LL subband. Then only apply the modified Shamir's algorithm to this

LL subband to generate the shadow images. Anbarasi and Kannan [1] transform the secret data to m -ary notational system. They use the polynomial function generated using $(t-1)$ digits of secret color image pixels. Chai [3] proposed a new scheme utilized (t, n) -threshold to share the secret image and embedded the produced n shadows into different cover images, any t out of n shadows can reveal the original secret image, and the cover images could be reconstructed with limited distortion. Angelina et al [2] apply the interpolation-based (k, n) -threshold secret image sharing scheme, where a secret data payload is optimized using Lagrange Interpolation operated in $GF(2^8)$. The proposed scheme provides an authentication mechanism by parity-bit checking and lossless recovery of the secret data using $GF(2^8)$ operation. Chien and Hwang [4] propose a novel method for secret image sharing based on a power-of-two Galois field rather than primes. Kim et al [5] proposed scheme solves the problems include the number of participants is limited because of modulus prime number m . Li et al [6] scheme generates shares from the cover image and secret image, and then embeds the shares into the cover image obtaining stego-shadow images based on 24-ary notational system. Lin and Wang [7] invented an improved method of invertible secret image sharing scheme. The scheme divides the input secret image into multiple sections with each section fits in an α -bit space, and all arithmetic operations are performed in power-of-two Galois Field $GF(2^\alpha)$. Some authors did not follow the Shamir's method. Naskar et al [8] Apart from the Shamir's secret sharing technique, they suggested a scheme which deploys simple graphical masking, done by simple ANDing for share generation and reconstruction can be done by simple ORing the qualified set of shares. Tang and Huang [9] proposed a new secret image sharing scheme based on bivariate polynomial, besides has most advantages of Shamir scheme, access structure in the new scheme has been greatly enriched. They claimed their scheme is much more suitable for image processing.

The rest of this paper is organized as follows. Next section we are going to review the Shamir's scheme and Lagrange algorithms then describe our proposed method. The section 3 demonstrates the experiment results. The conclusions are given in the final section.

2 Proposed Method

We proposed a multiple secret image sharing method. It applies the modified Shamir's algorithm [7] to one host image with several images to generate the shadow images. The retrieval step we exploit Lagrange method with enough numbers of the shadow images to obtain the original secret image. Here we describe Shamir's base secret image sharing method that we used.

2.1 Reviewed the Shamir's Scheme

We apply Shamir's base method that is similar to the lossless version of Thien and Lin [10] method (t, n) threshold scheme. If the secret image S will be shared n participants, i.e. n shadow images (S_1, S_2, \dots, S_n) will be created, and the secret image S

may not be retrieved at least t shadow images. This is called the (t, n) threshold Shamir's scheme. It generates the $t-1$ degree polynomial, by letting the t coefficients be the grayscale of t pixels. The difference between [10] method and the original Shamir's scheme is that they do not use random coefficient. Since the grayscale of a pixel is between 0 and 255, the prime number p is 251 that is the greatest prime number less than 255. Thus the lossless version must separate 0 to 255 grayscales into two parts, the first part is the range from 0 to 250, and the second part is 251 to 255. We do not use the 251 to 255 grayscale pixels. A location map will be marked those grayscales. We select t consecutive grayscales from the secret image in raster scan way until exhausted. Therefore, we define the following $t-1$ degree polynomial

$$q(x_i) = a_{t-1}x_i^{t-1} + \dots + a_1x_i + a_0 \pmod{P} \quad (1)$$

where a_0, \dots, a_{t-1} are the t grayscales of the secret image, and then evaluate $q(x_1), q(x_2), \dots, q(x_n)$. x_1, x_2, \dots, x_n are identification value of the participants. A set of the simple identification values is $\{1, 2, \dots, n\}$. The n output value pixels $q(x_1) \sim q(x_n)$ are sequentially assigned to the n shadow images. For each t pixels of the secret image, each shadow image receives one of the generated pixels; the size of each shadow image is $1/t$ of the secret image. The following steps are the lossless secret image sharing method.

The **sharing** steps are

1. Convert the 2-D secret image into the 1-D array by raster scan, then use pseudo-random generator (PRNG) to shuffle the pixels of the 1-D array, and use a location map to record the positions of those grayscales are 251~255, and replace them by 250 ~ 246.
2. Sequentially take t not-shared-yet pixels of the permuted image (1-D array).
3. Use the t pixels in Step 2 and Eqs. (3) then generate n pixels $q(x_1), q(x_2), \dots, q(x_n)$ for the n shadow images.
4. Repeat Steps 2 and 3 until all pixels of the permuted image are processed.

The **retrieval** steps are

1. Collect the shadow images and the identification numbers (x_i) of t or larger than t participants.
2. Convert each of those 2-D shadow images into 1-D arrays by raster scan, and calculate the original secret by the following Lagrange interpolation to get the original secret image $q(q(x_1), q(x_2), \dots, q(x_n))$,

$$q(x_i) = \sum_{b=1}^t y_{ib} \prod_{j=1, j \neq i}^t \frac{x_i - x_{ij}}{x_{ib} - x_{ij}} \pmod{P} \quad (2)$$

where x_i is the identification number and y_{ib} is the current pixel's grayscale of the shadow image which x_i owns.

3. Collect $q(x_i)$ then use the location map to change those grayscales 250 ~ 246 back to 251~255, and use pseudo-random generator (PRNG) to shuffle back to original order.
4. Convert the 1-D array into the 2-D original secret image.

2.2 Proposed Method

Our proposed method applies one host image and several secret images to the modified (t, n) threshold Shamir's scheme which is discussed in the sharing part of above subsection 2.1 to generate the shadow images. We exploit Lagrange method with enough numbers of the shadow images to obtain the original secret image that was described in the retrieval part in above subsection 2.1.

Sharing Steps

- Step1: Convert each secret image to bit streams, then combine bit streams into one bit stream S .
- Step2: Use PRNG to shuffle the bit stream S with the PRNG (Pseudo Random Number Generator).
- Step3: Convert the bit stream S to a 1-D array with n -digit number where $2^n < P$ (251)
- Step4: Record the positions of those grayscales of host image are 251~255 to a location map, and replace them by 250 ~ 246.
- Step5: Combine the 1-D array of host image and multiple secret images sequentially into one 1-D array S .
- Step6: Sequentially take t not-shared-yet pixels of the permuted 1-D array and apply Eqs. (2.3) then generate n pixels $q(x_1)$, $q(x_2)$, ..., and $q(x_n)$ for the n shadow images.
- Step7: Repeat above Steps 5 until all pixels of the permuted bit stream S are exhausted.
- Step8: Convert each 1-D arrays into 2-D shadow images
- Step9: Distribute all image sizes, the location map of host image, one identification numbers (x_i) , and one shadow image to each participant i ($1 \leq i \leq n$).

Retrieval Steps

- Step1: Collect the shadow images and the identification numbers (x_i) of t or larger than t participants.
- Step2: Convert each of those 2-D shadow images into 1-D arrays by raster scan, and calculate the corresponding pixels by the following Lagrange interpolation to get the original host and multiple secret images $q(x_i)$,

$$q(x_i) = \sum_{b=1}^t y_{ib} \prod_{j=1, j \neq b}^t \frac{x_i - x_{ij}}{x_{ib} - x_{ij}} \pmod{P} \quad (3)$$

where x_i is the identification number and y_{ib} is the current pixel's grayscale of the shadow image which x_i owns.

- Step3: Distribute the $q(x_i)$ to host image and other multiple secret images sequentially ending with each image size.
- Step4: Use the location map to change those grayscales 250 ~ 246 back to 251~255, and use pseudo-random generator (PRNG) to shuffle back to original order.
- Step5: According to each image size to convert each 1-D array into host and secret images.

3 Experiment Results

Due to our goal is to apply one host image with multiple images to image sharing method and still achieve to recover the secret image reversible. We examine several frequently host images such as Lena, Baboon, Airplane and Peppers, shown in Figure 1 to do experiment. Figure 2 shows two secret images which are our school's marks. An example is used Lena and Fig. 2 secret images to present the (3, 5) threshold condition is shown in Fig. 3.

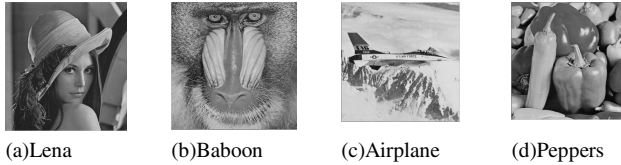


Fig. 1. The sizes of host images are all 512 by 512

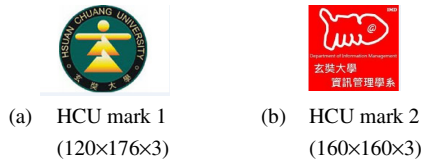


Fig. 2. Secret images

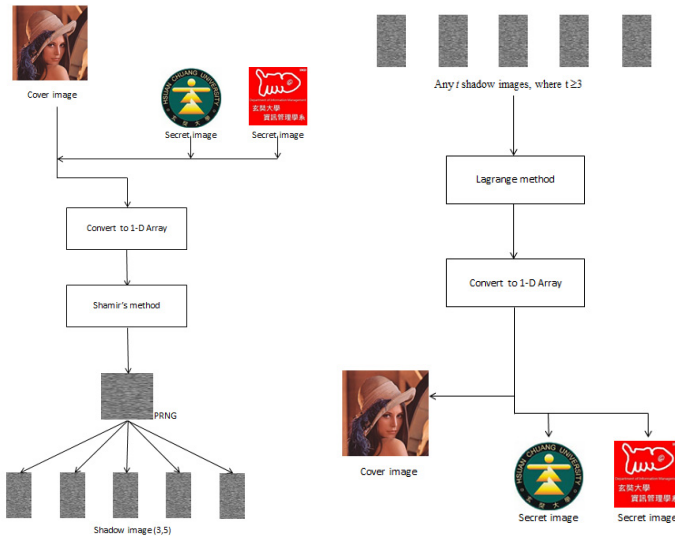


Fig. 3. Left figure is the sharing steps, and right figure is retrieval steps

Above figure presents the (3, 5) threshold condition used our proposed method. Any three to five shadow images of participants may retrieve the secret image lossless.

4 Conclusions

In this paper, we apply secret image sharing based on Shamir's scheme for multiple secret images. The traditional image sharing scheme based on the (t, n) threshold scheme divides one secret image into n noise-like shadow images that are distributed and stored in different owners. The secret image can be rebuilt completely with at least t participants of these n shadow images. We may combine the data hiding or watermarking methods to each secret image before sharing steps to accomplish sharing and preserve the copyright protection more safely.

References

1. Anbarasi, L.J., Kannan, S.: Secured Secret Color Image Sharing With Steg-anography. In: International Conference on Recent Trends in Information Technology (ICRTIT), pp. 44–48 (2012)
2. Angelina, E.T., Ivan, C.C., Mariko, N.M., Gina, G.G., Hector, P.M.: Improved Secret Image Sharing Scheme with Payload Optimization. In: IEEE 55th International Midwest Symposium on Circuits and Systems, pp. 1132–1135 (2012)
3. Chai, Y.P.: An improved secret image sharing scheme with steganography. In: International Conference on Mechatronic Science, Electric Engineering and Computer, pp. 1335–1338 (2011)
4. Chien, M.C., Hwang, J.G.: Secret image sharing using (t,n) threshold scheme with lossless recovery. In: 5th International Congress on Image and Signal Processing, pp. 1325–1329 (2012)
5. Kim, D.H., Lee, G.H., Park, M.H., Yoo, K.Y.: A Reversible Secret Sharing Scheme Based on $GF(2^8)$. In: Ninth International Conference on Information Technology: New Generations, pp. 425–430 (2012)
6. Li, L., Abd El-Latif, A., Yan, X., Wang, S., Niu, X.: A Lossless Secret Image Sharing Scheme Based on Steganography. In: Second International Conference on Instrumentation, Measurement, Computer, Communication and Control, pp. 1247–1250 (2012)
7. Lin, Y.Y., Wang, R.Z.: Improved Invertible Secret Image Sharing with Steganography. In: Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 93–96 (2011)
8. Naskar, P.K., Chaudhuri, A., Basu, D., Chaudhuri, A.: Second International Conference on Emerging Applications of Information Technology, pp. 177–180 (2011)
9. Tang, D., Huang, J.: Secret Image Sharing Scheme Based on Bivariate Polynomial. In: 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering, pp. 193–195 (2012)
10. Thien, C.C., Lin, J.C.: Secret image sharing. *Computers & Graphics* 26(5), 765–770 (2002)
11. Yang, C.-H.T., Huang, Y.-H., Syue, J.H.: Reversible Secret Image Sharing Based on Shamir's Scheme with Discrete Harr Wavelet Transform. In: International Conference on Electrical and Control Engineering, pp. 1250–1253 (2011)

The Arm Strength Training Machine with Biofeedback

Tze-Yee Ho¹, Mu-Song Chen², Yuan-Joan Chen³, and Hung-Yi Chen¹

¹ Dept. of Electrical Engineering, Feng Chia University,
100 WenHwa Road, Seatwen, Taichung, Taiwan, R.O.C.

² Dept. of Electrical Engineering, Da-Yeh University,
No.168 University Rd., Dacun, Changhua, Taiwan, R.O.C.

³ Dept. of Info. Management, Ling Tung University,
1 Ling Tung Road, Taichung, Taiwan, R.O.C.

tyho@mail.fcu.edu.tw, chenms@mail.dyu.edu.tw,
honjoan@gmail.com, i0936847262@hotmail.com

Abstract. The aim of this paper is to design and implement arm strength training machine with electromyographic (EMG) biofeedback based on a microcontroller system. The hardware design based on a microcontroller is analyzed and discussed. The software programming is developed in MPLAB integrated development environment from the Microchip Technology Inc. and the friendly user interface is created as well. Finally, an arm strength training machine with electromyographic biofeedback is realized and demonstrated. The experimental results show the feasibility and fidelity of the complete designed system.

Keywords: electromyographic biofeedback, microcontroller, arm strength training machine.

1 Introduction

A study to investigate neuromuscular electrical stimulation initiated by a surface electromyographic biofeedback threshold on knee extension active range of motion (AROM), function, and torque in patients with post-operative arthroscopic knee surgery has been addressed in [1]. It concludes that the usage of surface EMG-triggered for neuromuscular electrical stimulation can improve the extension AROM. Another study addressed in [2] indicates that chronic stroke patients who gained maximal functional benefits from the biofeedback intervention initially had greater active range of motion at all major upper extremity joints. Consequently, the proper utilization of electromyographic biofeedback can lead to substantial improvements among select chronic stroke patients and can be of considerable functional benefit to others. Therefore, the usage of EMG not only can help the physical therapy but also achieve the more effective rehabilitation. Conventional exercise apparatus typically couple a stack of iron weights through a series of pulleys and levels to hand grips [3]. To vary the force opposing the user, the user is required to change the position of mechanical locking pin and physically add or remove weights from the stack.

The time consuming and inconvenient to change the exercise force level between lifts, are some drawbacks for such an exercise apparatus [4]. To solve this problem, a closed-loop motor control system that can generate an user opposition force and more particularly simulate a weight stack, is presented in [5]. Therefore, in order to obtain more effective rehabilitation when manipulate the arm strength training machine (ASTM), an EMG system incorporated with the original functions is designed and implemented in this paper.

The hardware circuits of the PMSM (Permanent Magnet Synchronous Motor) drive, such as AC/DC rectifier, DC link, DC/AC inverter, EMG sensors, Hall-effect position sensors and speed encoder, are well designed, simulated and implemented. The software programs are written in C language and programmed based on the MPLAB integrated development environment (IDE) tool by Microchip technology incorporate [6]. The PMSM motor drive is used to simulate the weight stack which is usually employed to the conventional exercise machines. Thus, the principle of PMSM motor drive is firstly derived and described in this paper. Later, the system hardware and software are designed and realized in subsequence. Finally, an arm strength training machine with electromyographic biofeedback is realized and demonstrated. The experimental results show the feasibility and fidelity of the complete designed system.

2 The System Hardware Structure of ASTM

The system hardware of an arm strength machine based on PMSM motor drive is shown in Fig. 1. It consists of a dsPIC30F4011 microcontroller, protection circuit, optical coupling isolation, inverter, current sensor, encoder, and communication interface. The PMSM motor drive is used to simulate the weight stack which is usually employed to the conventional exercise machines.

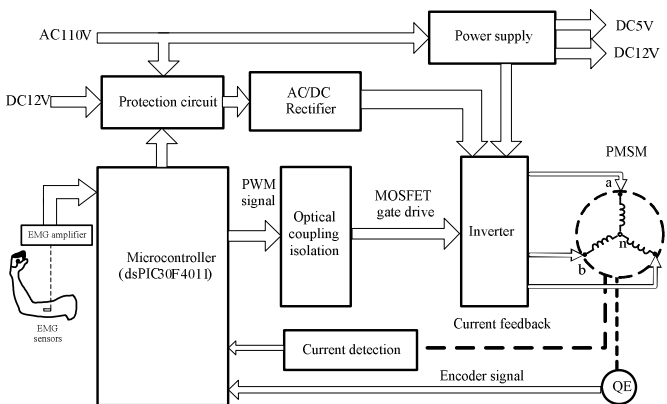


Fig. 1. The system hardware of ASTM with EMG

The microcontroller dsPIC30F4011 manufactured by Microchip technology incorporate is the core controller of the ASTM. It is a 16-bit CPU with the capability of digital signal processing. Moreover, it supports many powerful modules such as built-in PWM module, addressable encoder interface module, and input capture module, these make the design more friendly and thus shorten the development schedule. The independent power source is employed to supply the gate of MOSFETs. The photocoupler TLP250 is used for electrical isolation between the microcontroller system and the high DC voltage bus voltage as well as the independent power source. The EMG biofeedback mechanism consists of the EMG electrodes, the EMG amplification circuits and the bandpass filter, as shown in Fig. 2. Two electrodes of EMG sensor are attached to the surface skin of an arm. A third electrode is attached to the common point for voltage reference. The potential difference is generated when the muscle group contracts and then fed into the instrumentation amplifier for amplification. Since the input resistance of an instrumentation amplifier is very high, it is suitable to pick up the EMG signal with high output resistance. In order to obtain a “clean” signal that is no dc offset and high frequency noise, the output signal from the instrumentation amplifier is filtered by a bandpass filter. The clean signal is then rectified to dc value by using the rectifier circuit. The rectified signal is input into the microcontroller to display the contraction of muscle group when the user manipulates the ASTM.

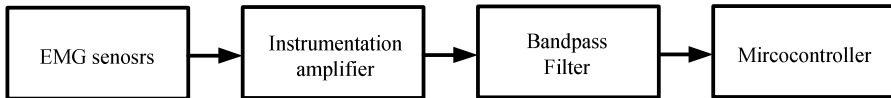


Fig. 2. The function block diagram of EMG system

The equivalent circuit of a PMSM motor is shown in Fig. 3. The stator phase voltage equations (V_{an} , V_{bn} , V_{cn}) related to the stator phase currents (i_a , i_b , i_c) and back electromotive force (e_a , e_b , e_c) for a PMSM motor, can be expressed by (1) [7].

$$\begin{aligned}
 V_{an} &= R_a i_a + L_{aa} \frac{di_a}{dt} + L_{ab} \frac{di_b}{dt} + L_{ac} \frac{di_c}{dt} + e_a \\
 V_{bn} &= R_b i_b + L_{ba} \frac{di_a}{dt} + L_{bb} \frac{di_b}{dt} + L_{bc} \frac{di_c}{dt} + e_b \quad . \\
 V_{cn} &= R_c i_c + L_{ca} \frac{di_a}{dt} + L_{cb} \frac{di_b}{dt} + L_{cc} \frac{di_c}{dt} + e_c
 \end{aligned} \tag{1}$$

Where R_a , R_b , R_c , represent the phase resistance for each phase, L_{aa} , L_{bb} , L_{cc} represent the self inductance for each phase and L_{ab} , L_{bc} , L_{ca} represent the mutual inductance between either of two phases, e_a , e_b , e_c , represent the back EMF for each phase.

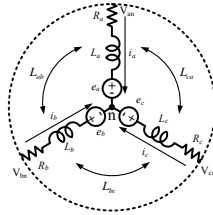


Fig. 3. The equivalent circuit of a PMSM motor

Hence the electromagnetic torque can be represented as

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega_m} \tag{2}$$

The load model can be expressed in terms of a moment of inertia, J , in $\text{kg}\cdot\text{m}^2/\text{sec}^2$ with a viscous friction B , in $\text{N}\cdot\text{m}/\text{rad}/\text{sec}$. The electromagnetic torque, T_e , in $\text{N}\cdot\text{m}$ then drives the load torque, T_L , in $\text{N}\cdot\text{m}$ as represented in (3) [6-7].

$$J \frac{d\omega_m}{dt} + B\omega_m + T_L = T_e \tag{3}$$

3 The System Software Development

The system software program is developed under MPLAB IDE software platform and written in C language. Most of the functions of electric bicycle are programmed in the microcontroller firmware which includes the circuit protection mechanism, the ADC converter for EMG system, PWM generation, motor currents calculation, rotor position and speed calculation and rotor pole position [7-8]. The flowchart of main program for microcontroller firmware is shown in Fig. 4. The initializations for I/O configuration, Timer 1, Timer 2, ADC and PWM settings are firstly processed in the main program.

Most of the ASTM functions are programmed in the microcontroller firmware which includes the circuit protection mechanism, the analog to digital converter for EMG system, PWM generation, motor currents calculation, rotor position and speed calculation, rotor pole position as well as the transmit and receive of communication interface as shown in Fig. 4. The flowchart of main program for microcontroller firmware is shown in Fig. 5. Since the resolution of encoder is 2500 pulses per revolution. The value of the counter in the microcontroller will be 5000 counts. Because the sinusoidal waveform is symmetric for 0° to 180° , only the sine values of 0° to 90° are created which covers the 312 counts of encoder for an 8-pole rotor. The motor speed is obtained from the difference between current counter value and the last counter value which both are acquired from the Timer2 in capture interrupt service routine. The EMG analog signal input to the microcontroller is first converted to the

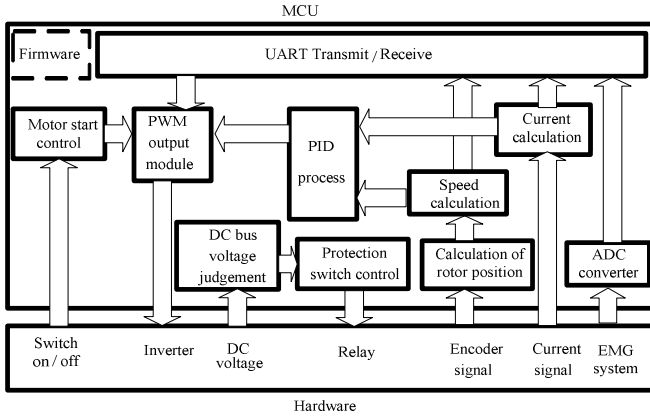


Fig. 4. The flowchart of main program

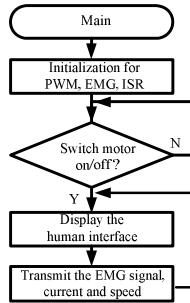


Fig. 5. The Flowchart of Main Program

digital signal via the ADC module embedded in the microcontroller. The corresponding digital signal is then divided into several segments for the output LEDs display. Accordingly, the human interface can display the same converted EMG signal as well.

4 The Experimental Results

The prototype of arm strength training machine is tested under different load conditions in which are fulfilled with the dynamometer. The designed human interface is also tested as well. Fig. 6 shows the waveforms of expected speed and actual speed under the SPWM current conduction mode when motor load of 7 kg-cm is applied and released. The current response while the ASTM being manipulated by 7 kg force is shown in Fig. 7.

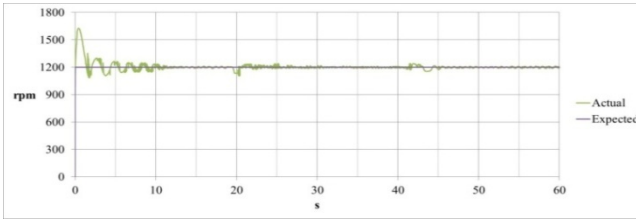


Fig. 6. The speed response for 7 kg-cm being applied and released by SPWM control

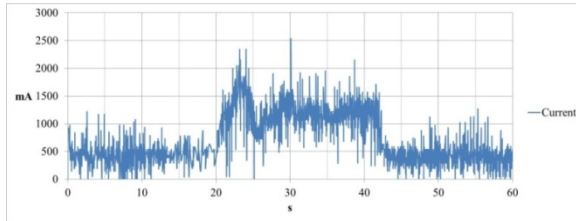


Fig. 7. The current response for 7 kg-cm being applied and released by SPWM control

The experiment is cyclically set by pulling the handle bar of ASTM for 20 seconds and then releasing for next 20 seconds. The experiment is repeated by the same cycle. Observing the waveform of Fig. 7, it can be seen that the motor draws about the 1.5 A current to counter the force exerted by the user. This verifies the system design feasibility. The data displayed in Fig. 6 and Fig. 7 are firstly saved in the memory and then sketched by using the Microsoft EXCEL software. The waveform of EMG signal for 7 kg force applied is shown in Fig. 8. This proves that the EMG signal biofeedback can reflect the muscle contraction and be displayed on the human interface when the user operates the ASTM.

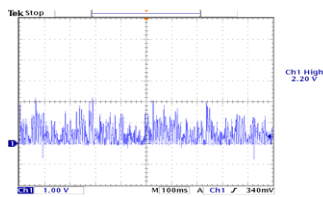


Fig. 8. The EMG signal for 7 kg force applied.

5 Conclusions

The system hardware structure including a microcontroller, protection circuit, optical coupling isolation, three-phase inverter, current sensor, EMG sensors, encoder, and communication interface is well designed. Further, both of the microcontroller firmware and user interface are program and described in detail. Finally, an arm

strength training machine with electromyographic sensors for biofeedback is realized and demonstrated in this paper. The experimental results show the feasibility and fidelity of the complete designed system.

References

1. Boucher, T., Wang, S., Trudelle-Jackson, E., Olson, S.: Effectiveness of surface electromyographic biofeedback-triggered neuromuscular electrical stimulation on knee rehabilitation. *North American Journal of Sports Physical Therapy* 4(3), 100–109 (2009)
2. Wolf, S.L., Binder-Macleod, S.A.: Electromyographic Biofeedback Applications to the Hemiplegic Patient. *Physical Therapy, Journal of American Physical Therapy* 63(9), 1393–1403 (1983)
3. Bugallo, F.: Weight lifting apparatus having increased force on return stroke. U.S Patent No. 4563003 (1983)
4. Hon, P.: Apparatus for training, investigation and re-education in particular for the neuromuscular function. U.S Patent No. 4979733 (1990)
5. Ho, T.-Y., Chen, Y.-J., Chen, P.-H., Chiang, C.-H.: The Design and Implementation of Windowing Interface Arm Strength Training Machine. In: *The 1st IEEE Global Conference on Consumer Electronics 2012 (GCCE 2012)*, Makuhari Messe, Tokyo, Japan, October 2-5, pp. 205–206 (2012) 978-1-4673-1501-2/12/ 2012 IEEE
6. Ho, T.-Y., Chen, Y.-J., Chung, C.-T., Hsiao, M.-H.: The design and implementation of a windowing interface pinch force measurement system. In: *SPIE Symposium on BIO-S*, pp. 7555–7756 (January 2010)
7. Pillay, P., Krishnan, R.: Modeling, simulation, and analysis of permanent-magnet motor drives. *IEEE Transactions on Industry Applications* 25, 265–273 (1989)

Privacy Breach Investigations of Incident Response to Personal Information Protection Act

Da-Yu Kao^{1,*}, Cheng-Yu Peng², Frank Fu-Yuan Huang³, and Shih-Jeng Wang¹

¹Department of Information Management, Central Police University, Taoyuan, Taiwan

²Graduate Institute of National Development, National Taiwan University, Taipei, Taiwan

³Directorate-General of Personnel Administration, Executive Yuan, Taipei, Taiwan

camel@mail.cpu.edu.tw

Abstract. The proliferation of big data is developing with substantial advancement to enter into glorious future, but the vulnerability of personal data has always been a disaster to this dream. In order to foster confidence in information systems, a novel examination of incident response approach is evaluated from a Taiwan hacking ring case. With the guidance of this case study, we can discriminate normal information sharing from internet privacy violation. Enhancing data privacy is a baffling task because of its newness and technological furtherance. It is believed that this study will clarify the obscure technological and social aspects of data privacy enhancement. The proposed security measures can prohibit individuals or organizations from the risk of getting hurts, facilitate to cut down its roots and remove its foundation.

Keywords: Information Security, Internet Privacy Violation, Hacker Case, Incident Response.

1 Introduction

As times goes by, few public issues attract more attention than the protection of privacy. The rapid growth of the internet has raised far-reaching questions about the future of privacy. Identity, financial information, education, and work performance data are commonly regarded as private, despite many are commonly accessible through credit-reporting organizations. The distinction between public and private behaviors is often ambiguous [1]. Technology gives with the one hand and takes with the other. While Personal Information Protection Act (PIPA) in Taiwan helps to safeguard data subject's privacy, the safeguard serves a large range of other values and interests. Little evidence exists to indicate that the architects of PIPA share a deep-seated hostility to computers and other forms of IT. The development and existence of PIPA have inspired in legal fields.

* Corresponding author.

The literature reviews of regulating data processing principles are discussed in Section 2. Section 3 describes the case scenario and case analyses. Discussions and analyses to improve the security measures on internet privacy violation are presented in Section 4. The conclusion is drawn in Section 5.

2 Literature Review

There are principles which regulate the manner of data processing. These apply a variety of criteria (eg: fair, lawful, legitimate, objectively justifiable, or necessary) to steer the processing of personal data along with certain avenues. We select a core set of privacy principles that are frequently addressed in privacy laws and regulations, and add other principles/properties that are also desirable for privacy enhancement. This section provides an overview of the basic principles applied by PIPA to the processing of personal data. In Fig. 1, these principles can be summed up from the following viewpoints: Regulate the Processing and Monitor the Set of Principles.

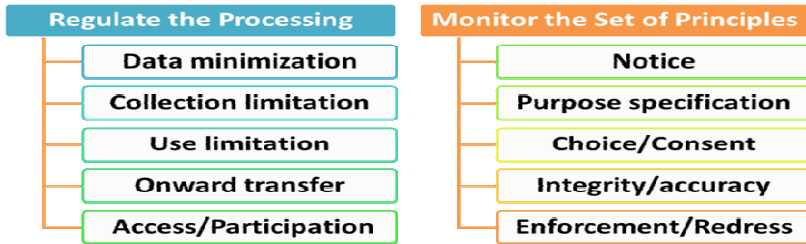


Fig. 1. Two Categories of Privacy Principles

2.1 Regulate the Processing

To regulate the processing on the internet, some privacy principles are discussed below [3].

(1) Data Minimization

The amount of collected personal data should be limited to what is necessary to achieve the purposes for which the data are gathered and further processed.

(2) Collection Limitation

There should be limited to the collection of personal data and any such data should be obtained by lawful and fair means.

(3) Use Limitation

Personal data should not be disclosed, made available or otherwise used for purposes other than those specified. Moreover data controllers' disclosure of personal data to third parties shall be restricted.

(4) Onward Transfer

Personal data should not be transferred to a third country/party if it does not ensure an adequate level of protection.

(5) Access/Participation

Persons should be able to participate in, and have a measure of influence over the processing of data on them by other individuals or organizations.

2.2 Monitor the Set of Principles

To monitor the set of privacy principles on the internet, some principles are discussed below [6].

(1) Notice

Develop comprehensive and consistent policies. Then make privacy policy statements clear, concise, and conspicuous to those responsible for deciding whether and how to provide the data.

(2) Purpose Specification

Personal data shall be collected for specified, lawful or legitimate purposes.

(3) Choice/Consent

Where appropriate, individuals should be provided with clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice in relation to the collection, use and disclosure of their personal information.

(4) Integrity/Accuracy

A data controller should ensure the collected personal data is sufficiently accurate and up-to-date for the intended purposes and all corrections are propagated in a timely manner to all parties that have received or supplied the inaccurate data.

(5) Enforcement/Redress

Effective privacy protection must include mechanisms for enforcing the core privacy principles.

3 Case Study of Internet Privacy Violation

3.1 Scenario

On Aug. 28, 2008 Taiwan's Criminal Investigation Bureau (CIB) has successfully arrested six people suspected of cracking an identity theft ring that victimized millions of people from state organizations, government organizations, state-run firms, telecoms companies and even a television shopping network. The story is summarized as follows in Fig. 2 [5].

(1) Taiwanese Hackers Stole 50 Million Records of Personal Data

More than 50 million records of personal data are believed to have been lifted by the six criminals, including information about Taiwan incumbent President, ex-President, and the head of police agency. It was both business and consumers' responsibility to take care that they are operating securely on-line and everyone should be using firewalls, anti-virus software, and strong passwords.

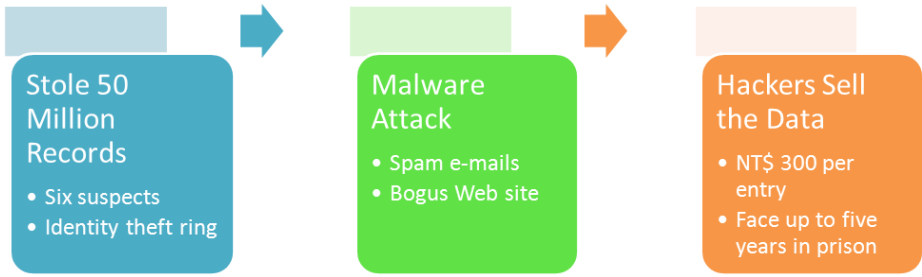


Fig. 2. Taiwan Nails Major Hacking Ring Case

(2) Malware Attack in the End users

At the center of the scam were spam e-mails masquerading as legitimate requests from banks and asking customers to submit personal information on-line. In phishing, people answering an e-mail are directed to a bogus Web site where they are asked to update personal information, such as passwords and account numbers.

(3) Hackers Sell the Personal Data through Taiwanese Underworld

Money appears to have been the motive for this hacking ring. They then offered to sell the personal data for NT\$ 300 dollars per entry through the Taiwanese underworld. If they are convicted, they will face up to five years in prison on charges of hacking and fraud charges.

3.2 Case Analyses

Information is increasingly being regarded as a valuable resource in itself. There exists a rapidly growing market in information service. A market can be bought and sold for significant financial sums. This section takes up the question why internet privacy violation exists. In Fig. 3, the catalysts for emergence and continued existence of internet privacy violation fall into three broad categories: (1) IT developments; (2) Public fears; (3) Legal factors [2].

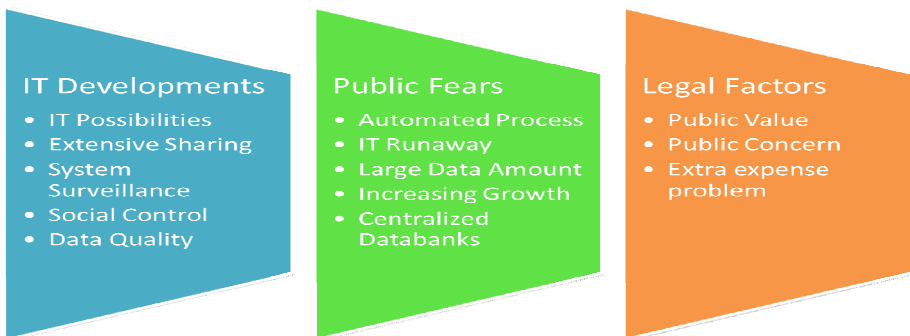


Fig. 3. Catalysts for PIPA Emergence

(1) IT Developments

New computer networks will increase the difficulties experienced by totalitarian regimes. Technological and organizational developments can be analyzed from the following aspects [3].

■IT Possibilities

The emergence of internet privacy enhancement cannot properly be explained without taking account of developments in information technology. These developments have brought vastly expanded possibilities for amassing, linking and accessing personal data.

■Extensive Sharing

Internet privacy enhancement has been characterized through the increasingly extensive sharing of personal data across the traditional organizational boundaries.

■System Surveillance

The growth in social scale and various systems have contributed to the development of system surveillance.

■Social Control

New technologies help to shift the parameters of social interaction, creating new opportunities of activity, and magnifying existing opportunities.

■Data Quality

The emergence and continued existence of PIPA is an accumulating body of evidence indicating that the quality of data utilized by numerous organizations is deficient.

(2) Public Fears

The fears below are primarily economic in nature and shared by governments and business. The public fears are shown below.

■Automated Process

We face with information systems of growing complexity and diminishing transparency. Data are being handled by many persons and organizations. The encroachment of automated process pretends a future in which we know little or nothing.

■IT Runaway

People fear that the environment resulting from this complexity will elude full human comprehension. They warn of a future in which humans will increasingly come under the sway of IT runaway that cannot be effectively steered.

■Large Data Amount

The increasing transparency of data subjects revolves the effects of the data amount. A general anxiety will result in an unprecedented aggregation of power in large organizations.

■Increasing Growth

The developments centrally figure growth in the amount of personal data held by various types of organizations. While a great deal of data protection disclosure has been

preoccupied with technological threats to privacy and related values, it nevertheless appears to be infused by growing awareness of the double-sided character of technology.

■Centralized Databanks

The developments centrally figure integration of these data holding into centralized databanks.

(3) Legal Factors

The legal factors are illustrated below: Public Value, Public Concern, and Extra expense problem.

■Public Value

The overarching public values can be summed up in terms of ensuring data validity and information utility, together with information system’s manageability, robustness, accessibility, reliability and comprehensibility.

■Public Concern

Public concern over the IT development has concentrated partly on the potential of data sharing or re-usage. Public concern on their potential can significantly roll back the privacy and autonomy of citizens and undermine in turn the foundations for democratic society.

■Extra Expense Problem

Protecting legal person data will increase the workload of data controllers. This will make data-processing more expensive and cumbersome. The extra expense could be most problematic for small organizations.

4 Discussions and Analyses

The case study is an intensive analysis of an internet privacy violation, stressing developmental issues in relation to context. The Fig. 4 illustrates how an organization might assign issues to the specific instance of personal information [6].

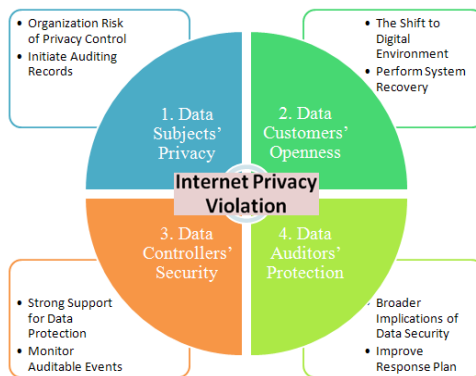


Fig. 4. Proposed Incident Response Approach on Internet Privacy Violation

4.1 Four Concerns on Internet Privacy Violation

Due to the particular risks of internet data harm, organizations should develop additional concerns, such as data subjects' privacy, data customers' openness, data controllers' security and data auditors' protection. Organizations should integrate these additional concerns into their existing incident handling response approaches [1].

(1) Data Subjects' Privacy

PIPA gives data subjects control over personal information by requiring organizations to obtain the consent to collect, use or disclose personal information about any person. There are some limited exceptions to the consent requirement. It gives us certain rights and imposes specific obligations on organizations.

(2) Data Customers' Openness

Destroy, erase or make anonymous personal information about us that it no longer needs for the purpose for which it was collected or for a related business or legal reason.

(3) Data Controllers' Security

PIPA is enacted to govern the collection, processing and use of personal information so as to prevent harm on personality rights, and to facilitate the proper use of personal information.

(4) Data Auditors' Protection

Central to these debates is the role of law. To what extent can the law safeguard the right of privacy in an era of rapidly evolving technology? What competing interests must be considered? What is the appropriate role the courts and the legislatures? These questions are not new, but they have acquired greater urgency as the law is asked to evaluate an increasingly complex array of privacy matters.

4.2 Four Problems on Internet Privacy Violation

Certain circumstances within any organization or specific system, such as the context of use or obligation to protect, may cause different outcomes. Obligation to protect is a particularly important factor that should be determined early in the categorization process [3].

(1) Organization Risk of Privacy Control

Once our personal information is stolen from hackers, we no longer have control over its privacy. To operate effectively organizations need complete and accurate information about the individual. However, if individuals do not trust the organization to protect the confidentiality of their personal information, they will likely withhold or ask the organization not to record sensitive information.

(2) The Shift to Digital Environment

The shift to digital environment alters our understanding of privacy protection. The expansion of the internet has greatly increased our ability to collect, process, and use all kinds of information.

(3) Broader Implications of Data Security

Use of the internet has resulted in recognition that information technology security is of major importance to our society. Security principles target organizational governance and executive management [2].

(4) Strong Support for Data Protection

Evidence exists to indicate that extension of protection to data on legal persons can be occurred in order to provide more complete data protection for both legal and natural persons. There is also evidence suggesting that an established business group has expressed strong support for giving data protection rights to legal persons.

4.3 The Proposed Security Measures on Internet Privacy Violation

Management of incidents involving personal information often requires close coordination among personnel from across the organization, such as data subjects, data customers, data controllers and data auditors. When organizations deal with incident response for personal information breaches, some proposed solutions are listed below [6].

(1) Initiate Auditing Records

Organizations can regularly review audit records for the indications of the inappropriate or unusual activity affecting personal information. Then the organizations investigate suspicious activity or suspected violations, report findings to appropriate officials, and take necessary actions.

(2) Monitor Auditable Events

Organizations can monitor events that affect the confidentiality of personal information, such as unauthorized access to personal information. Evidence is also important in privacy breach investigation, where computer activity often leave a tell-tale mark [4]. Trace evidential information is the clue that occurs when different computers contact one another.

(3) Perform System Recovery

Existing technologies and techniques for containment, eradication, and recovery may be used for breaches involving personal information. It is important to determine whether personal information was accessed and how many records or individuals were affected.

(4) Improve Response Plan

The incident response plan should be continually improved during each incident. Lessons learned might also indicate the need for additional training, security controls, or procedures to protect against future incidents.

5 Conclusion

Handling the internet privacy violation may require additional actions by an organization. In the light of reportedly rising concerns about privacy and their

inhibition to electronic business growth, organizations could market privacy as a win-win situation. Our understanding into consumers' privacy attitudes into their privacy-related decision-making remains limited. We still have a hard time recommending concrete measures for improving data protection on the internet. The proposed approach can prohibit individuals or organizations from the risk of getting hurts, facilitate to cut down its roots and remove its foundation.

References

1. Dennis, T.C.: Taiwan Proposes Amendments to its 1995 Data Protection Act: Scope Expanded but no Supervisory Authority. *Privacy Laws & Business International Newsletter* 97, 19–20 (2009)
2. Jewkes, Y., Devon, G.: *Handbook of Internet Crime*. Willan Publishing (2009)
3. Jonathan, C.: *Principles of Cybercrime*. Cambridge University Press, Cambridge (2010)
4. Marcella, A.J.: *Cyber Forensics: A Field Manual for Collecting, Examining, and Preserving Evidence of Computer Crimes*. Auerbach Publisher, Florida (2008)
5. Criminal Investigation Bureau, <http://www.cib.gov.tw/>
6. Wang, Y., Kobsa, A.: Privacy-enhancing technologies. In: Gupta, M., Sharman, R. (eds.) *Social and Organizational Liabilities in Information Security*, pp. 203–227. IGI Global, Pennsylvania (2009)

Measuring Digital Crime Investigation Capacity to Guide International Crime Prevention Strategies

Joshua I. James¹ and Yunsik Jake Jang²

¹Digital Forensic Investigation Research Group,
University College Dublin, Belfield, Dublin 4, IE
Joshua@cybercrimem.com

²International Cybercrime Research Center,
Korean National Police University, Yongin-si, South Korea
ccimem@gmail.com

Abstract. This work proposes a method for the measurement of a country's digital investigation capacity and saturation for the assessment of future capacity expansion. The focus is on external, or international, partners being a factor that could negatively affect the return on investment when attempting to expand investigation capacity nationally. This work concludes with the argument that when dealing with digital crime, target international partners should be a consideration in expansion, and could potentially be a bottleneck of investigation requests.

Keywords: Digital Forensic Investigation, Digital Investigation Capacity Measurement, Cybercrime Investigation Capacity Measurement, International Crime, Strategic Policing.

1 Introduction

Continuous development of information communication technology and large scale proliferation of digital devices is leading to an increase in victims of digital crime, as well as tools and evidence in criminal investigations [1]. Many crimes, even traditionally non-digital crimes such as murder, now normally have some sort of digital component [2,3]. A recent U.N. report also showed a global need for expansion of digital investigation capability [4], especially with cross-border technical assistance.

When expansion of investigation capacity is made, an organization should be able to quantify how this expansion affects the organization to ensure a maximized return on investment. Some works have previously examined investigation capacity in law enforcement [5], but have not specifically focused on digital investigation services or the interplay between national expansion and the effect on partner countries. When an organization has to interact with other organizations nationally and internationally – especially in some sort of throughput or dependent relation – they should also consider how their expansion will affect these other organizations, and how the other organizations may affect return on investment.

1.1 Contribution and Structure

This work examines the needs of digital investigation capacity expansion, and how such capacity expansion can be measured. This work argues that measurement of capacity and its effect on external partner organizations can lead to more strategic investment strategies to reduce global digital crime.

First, an overview of digital crime investigation capacity expansion in a global context is discussed. Section 3 proposes a method to model digital crime investigation capacity within a country. Next, the concept of investigation capacity saturation is discussed, and a method for choosing expansion investment based on strategic partner organization capacity saturation is given. Section 4 then gives final thoughts and areas for future work.

2 Digital Crime Investigation Capacity Expansion in a Global Context

Digital crime investigation units are oftentimes looking to expand investigation capacity. Expansion of investigation capacity could allow an organization to increase their scope of service, investigate all national case requests in a timely manner (reduce or eliminate a backlog), open more cases for international investigations, or even just allow an organization to better meet the needs of their own citizens. Expansion could come in the form of more funding for training, equipment, personnel, etc. [6,7]. For example, Irish digital forensic investigators estimated that up to 40% of all exhibits receiving a full digital forensic analysis are determined to not be relevant to the case, and time and storage reductions could be made by slightly changing the investigation work flow [8].

Many organizations attempt to expand their investigation capacity when their budget allows. However, in the case of digital crime, many crimes have an international component. However, when one country needs assistance from another, the the investigation capacity of both countries should be considered. For example, if one country increases it's investigation capacity, this increase in capacity may result in more international requests. These requests will have a direct impact on the requested country's ability to handle other investigation requests. Consider Figure 1 and Figure 2. In Figure 1, if the current investigation capacity of Requestor 1 allows for 3 international requests, other countries may be able to handle these extra requests. However, after expansion of the investigation capacity of Requestor 1, international requests may now be doubled (Figure 2).

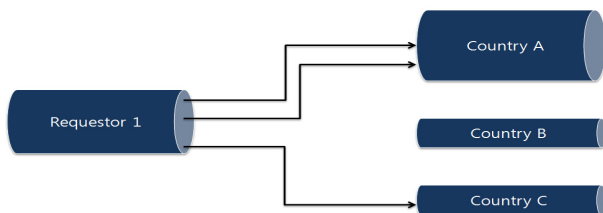


Fig. 1. Requestor 1 making international requests with capacity before expansion

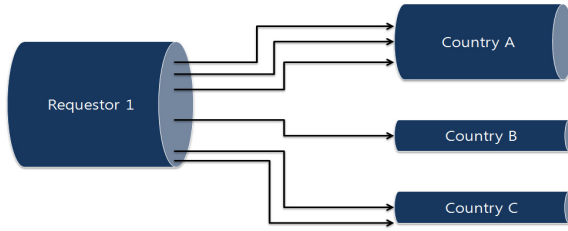


Fig. 2. Requestor 1 making international requests with a national capacity after capacity expansion

While expansion may seem to be a benefit for Requestor 1, consider Figure 3, where Country C is receiving international requests. Country C's investigation capacity is significantly lower. So much so that Country C can only keep up with their national investigation requests. If Requestor 1 is now making more international investigation requests to Country C, Requestor 1 may have to wait longer for a response, meaning that their case throughput capacity is throttled by the investigation capacity of international partners.

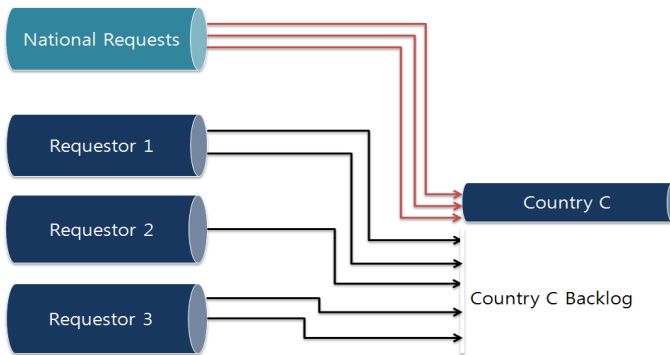


Fig. 3. Investigations requested to Country C where the national requests take up all current capacity, and international requests are forced to a backlog

Countries considering expansion of investigation capacity should factor in the effect such a national-level expansion may have on other countries. And specifically determine at what point there are reduced benefit to expanding investigation capacity.

3 Modeling Digital Crime Investigation Capacity

Digital crime investigation capacity can effectively be defined as *the maximum amount of cases categorized as digital crimes that can be investigated by a group over a given period of time.*

A case backlog is defined as *a queue of cases that are not being actively investigated, and are waiting to be started or finished by a group*. It may be common to have periods with a higher number of case requests than others. During peak periods, a backlog of cases may be created.

With these definitions, if total investigation requests over a given period of time are below investigation capacity then the cases will be started and cleared with no – or minimal – impact on a unit’s backlog. If total investigation requests over a given period of time are above capacity, then the cases will be started and cleared in a longer period of time, and have a measurable impact on the case backlog.

Essentially, capacity in investigations cannot be thought of as applicable to each singular case. For example, if one investigator does not currently have work assigned, that ‘unused’ capacity may not be able to be allocated to other ongoing tasks.

Capacity should be measured over time, and averaged for the group. This means that along with calculating the average number of cases completed per investigator, per year, the number of investigators also must be averaged per year. This allows units with high turn-over, or units with only part-time investigators assigned to still be able to attempt to measure capacity over the long term.

3.1 An Equation for Investigation Capacity

In this work investigation capacity is defined as an equation over time where the cases completed at the investigator, unit or national level is divided over the average number of available investigators for the same time period.

Let T be the time-span of interest

Let Ia be the time an investigator is available to work on cases

$$\text{For } T: (\text{cases closed})/(\text{average investigators}) \quad \dagger$$

where:

- average investigators = $(Ia_1/T) + (Ia_2/T) + (Ia_3/T) \dots + (Ia_n/T)$

This formula gives the average cases completed per available investigator. For example, if there were 4 cases closed over a 6 day period, one investigator was available full time, and one investigator was available part time (50%), then the calculation would be as so: $4/((6/6) + ((6/2)/6)) = 4/(1 + (3/6)) = 4/(1+0.5) = 4/1.5 = 2.7$

This means the throughput per investigator over this 6 day period is approximately 2.7 cases. Since there are essentially 1.5 investigators available, the overall potential throughput is $2.7 * 1.5 = 4$. Throughput, however, is not necessarily the same as investigation capacity.

For units or countries that currently *have* a case backlog, the average number of cases completed per time-span, per investigator multiplied by the current number of available investigators is an indication of the current unit or national investigation capacity.

For units or countries that do not have a case backlog, the additional available investigation capacity may have to be estimated at least in two ways. Either by

sampling maximum case throughput at a specific point in time where a backlog temporarily existed due to a surge in requests, or by using qualitative methods to ascertain at what capacity the investigators and their managers feel they performing at, or both.

For example, if the average cases completed annually in a country with 10 full-time investigators is 500, this country has no backlog, and each investigator estimates an average of 20% of 'down time', then the estimated national capacity would be 600 cases per year. $((500/10)+((500/10)*0.2)) * 10 = 600$

With the consideration of downtime, the equation † should be updated as follows:

$$\text{For } T: ((\text{cases closed}/\text{average investigators})+(\text{cases closed}/\text{average investigators}) * \text{down time}) * (\text{average investigators}) \quad \ddagger$$

This formula will calculate, at least, an overall capacity estimate for the group over a period of time. Capacity measurements can be averaged for all groups/units in a country.

3.2 Investigation Capacity Saturation

Investigation capacity can be compared with investigation requests to determine the capacity saturation in a particular unit or country. Capacity saturation can be thought of as a country's ability to handle more case requests without the request being backlogged or jumping the queue. To calculate capacity saturation, the number of incoming case requests can be divided by the investigation capacity. Using this calculation, if capacity saturation is above 1 (100%), then this indicates a backlog. The higher the capacity saturation, the more likely an organization will take a longer time to respond to a request, or potentially not respond at all.

If the investigation capacity of Country A continually increases, and Country A increases the number of international requests made, the capacity of other countries may become over-saturated. At this saturation point, additional investment in Country A's investigation capacity could result in a *reduction* of throughput until other countries' capacity increases. In this case, it is more beneficial for Country A to invest in investigation capacity of countries to whom investigation requests are commonly made, but that have a lower investigation capacity.

4 Conclusions

Countries develop investigation capabilities and capacity at different rates, depending on budgets, focus, law, etc. This work gave an overview of digital crime investigation capacity expansion in a global context, and especially how expansion of investigation capacity at local organizations could have an effect internationally that may be negative. After, a method to model digital crime investigation capacity was proposed. Investigation capacity saturation was discussed, and a method for choosing expansion investment based on strategic partner organization capacity saturation was demonstrated.

References

1. Casey, E.: *Digital evidence and computer crime: forensic science, computers and the Internet*, 2nd edn., p. xviii, 690p. Elsevier Academic, Amsterdam (2004)
2. RTE. Text message evidence in murder trial. RTE News (2010), <http://www.rte.ie/news/2010/0423/drimnagh.html> (retrieved)
3. Nguyen, L.: Tori Stafford trial: Cellphone record shows gap during abduction, murder. Postmedia News (2012), <http://www.canada.com/life/Tori+Stafford+trial+Cellphone+record+shows+during+abduction+murder/6486178/story.html> (retrieved)
4. Comprehensive Study on Cybercrime. United Nations Office on Drugs and Crime (2013), http://www.unodc.org/documents/commissions/CCPCJ_session22/13-80699_Ebook_2013_study_CRP5.pdf
5. Hekim, H., Gul, S., Akcam, B.: Police use of information technologies in criminal investigations. *European Scientific Journal* 9(4), 221–240 (2013), <http://eujournal.org/index.php/esj/article/view/778> (retrieved)
6. Gogolin, G.: The Digital Crime Tsunami. *Digital Investigation* 7(1-2), 3–8 (2010), doi:10.1016/j.diin.2010.07.001
7. Casey, E., Ferraro, M., Nguyen, L.: Investigation Delayed Is Justice Denied: Proposals for Expediting Forensic Examinations of Digital Evidence. *Journal of Forensic Sciences* 54(6), 1353–1364 (2009), doi:10.1111/j.1556-4029.2009.01150.x
8. James, J.I., Gladyshev, P.: A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digital Investigation*, 1–10 (2013), doi:10.1016/j.diin.2013.04.005

Study on the Growth Strategy to Become a Global Logistics Company, through the Expansion of the Domestic Logistics Companies in China

Hyunwoo Kim¹ and Chulung Lee^{2,*}

¹ Graduate School of Industrial Management Engineering, Korea University, Korea
blue108@korea.ac.kr

² Division of Industrial Management Engineering and Graduate School of Management
of Technology, Korea University, Seongbuk-gu, Seoul 136-701, Korea
leecu@korea.ac.kr

Abstract. Opening the era of competition between the very large logistics companies, and domestic logistics industry has seen many changes. It also has the competitiveness and technical know-how is second comparison of global logistics companies in the world and consists of the large-scale M & A activities between logistics companies, logistics companies larger. However, the reality is that the logistics of such a change being made in the domestic and overseas expansion and the size of the domestic logistics companies are extremely minor. The purpose of this paper is to analyze the current situation and the overseas expansion of the domestic logistics companies to offer the region through new global consumer market changes and future outlook of the Chinese market, the domestic logistics companies that can grow into a global logistics company aimed at domestic enterprises strategies and a network of global logistics companies to derive the comparative analysis the strategies, critical success factors.

Keywords: Global logistics, Global network, Global logistics strategy, Logistics strategy.

1 Introduction

Globalization the parts takes place in the existing domestic raw materials procurement, production, assembly, sales and customer service across the region, and the formation of several local economies in the world, and the need to respond to the diverse needs of customers, a comprehensive change management.

Consisting of the world's major economies, including the FTA and BRICS recent reorganization of the global market, global logistics environment is constantly changing and its importance is growing. Has seen many changes by opening the era of competition between the very large logistics companies in Korea. In addition, even if just a few years ago based on a high-quality logistics services, Logistics Company,

* Corresponding author.

the sales of more than 1 trillion was limited. But now sales of more than 3 trillion were four companies or firms. And has the world's global logistics companies such as competitiveness and technology, and know-how. However, the domestic logistics enterprises in overseas markets is very little attempt

Table 1. Comparison of scale logistics enterprises overseas M & A

Unit: Dollar				
Division	2008	2009	2010	2011
GLOBAL	995 billion	649 billion	1,107 billion	527 billion
KOREA	780 million	272 million	148 million	96million

Source: Chamber of Commerce and Industry for the logistics industry, M & A and implications of the report ".

M & A in the global logistics market size is U.S. \$ 527 billion. Given that only 0.18% \$ 9,600 million, the scale of the domestic logistics enterprises overseas M & A domestic logistics companies soon Globalization and global logistics company's growth strategy should proceed. Shall establish a global logistics company's growth strategy, especially China, which is planning a national growth strategy within the next 10 years, through the promotion of domestic consumption.

2 Analysis of the Growth Strategy of the Domestic and International Logistics Companies

2.1 Logistics Company's Growth Strategy, The Theoretical Background

In a variety of ways to promote growth in the the Ralph Scheuss research general corporate growth strategy 1) Business (original core of the business strategy, long tail), 2) Market strategy (Ansoff Matrix theory), replication strategy (Franchising, Licensing, etc.), and as built integration strategy (M & A, strategic alliance, joint venture).

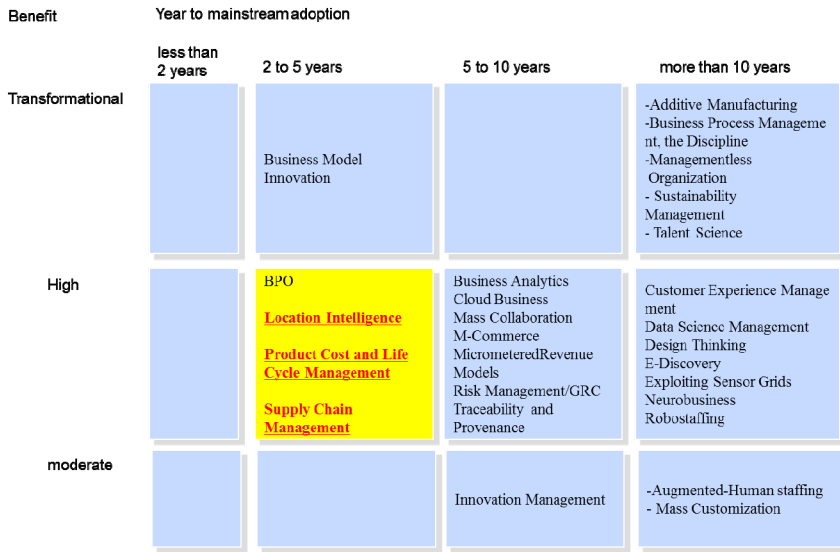
2.2 Logistics Research the Company's Growth Strategy

Existing research the the simple logistics strategy and network building, a lot of research has been carried out based on the manufacturing. Thus, previous research on logistics services to the company's growth strategy has been very limited.

However, in recent years, previous research was conducted on the strategy of growth factors and global logistics companies. Logistics Companies in cumulative abnormal returns on the effectiveness of increased M & A verification Park, Young - Tae (2008), Lee. Geun - Su (2009) use the AHP method to derive the key success factors of a global logistics company, Kim, HS (2008) The theory of economies of scale through the One-Stop SCM infrastructure vision to have a competitive edge in the domestic global logistics company's growth strategy was remarkable.

2.3 Recent Trends Emerging Technologies

According to Gartner (2012) reported data is expected to be transformed into the mainstream of the Trends of the Emerging Technologies in the time of the nearest



Source: Gartner

Fig. 1. Emerging Technologies Priority Matrix 2012

2-5 years Product Cost and Life Cycle management and Supply Chain Management technology.

2.4 Global Logistics Company's Growth Strategy Case Studies

Overseas major global logistics companies new logistics services provided through M & A, focused logistics competitiveness of a variety of shoes, and the growth was In addition to the non-logistics sector as complex logistics, e-commerce and financing services and to expand the service area vertical and horizontal integration established a system.

Table 2. Global logistics company's growth strategy

Enterprise	Growth strategy
Kuehne Nagel	Start forwarding company gradually expand the current range of services, but the current contract logistics capability based on supply chain management services.
KWE	Expand the area of business through advanced techniques and expertise of international air freight.
UPS	Since its establishment in 1907, UPS has grown into a group of the world's Package Delivery and Logistics / Finance Business aggressive entry in the SCM market.4PL company, founded by UPS Supply Chain Solutions provides comprehensive logistics services, express to escape the center of the image changes to the Comprehensive Supply Chain Solutions Group.

Table 2. (continued)

LI&Fung	Global sourcing company in product development, production management, customs clearance to delivery in one-stop service and SCM competency development growth.
DPWN	Company's strategy consists of NO.1 Worldwide Strategy(2003-2005) Globalization Strategic(1997-1999) Value Integration(2000-2003). The DPWN focus only on postal delivery was in Germany the past has grown into a global integrated logistics provider to a wide range of services with a portfolio strategy.
TOLL	Network expansion strategy to acquire the business locations (countries) of the Global Platform, the secure the cargo landing area customers, logistics companies, to strengthen the global SCM competence and sales of select.

Source: theorem

2.5 Case Study of the Growth Strategy of the Domestic Logistics Companies

In domestic, overseas markets, but any attempt to strengthen the third-party logistics and shippers with recent specific or did not visualize.

Table 3. The growth strategy of the domestic logistics companies

Enterprise	Growth strategy
CJ Korea Express	It plans to achieve logistics technology research and advanced technology to ensure the logistics efficiency, eco-friendly, energy-saving type. 5 trillion won to expand the global M & A and infrastructure investment. Indochina and South America, Africa, the Middle East, Eastern Europe, and subsequently secure a new base in the U.S region, expansion of the area.
Hanjin Express	Integration strategy based 3PL logistics infrastructure expansion and seaports, cargo terminals, etc. Build new core competencies of the logistics IT network and expanding.
Pantos	Linked to domestic demand in the local distribution center operations and inland transportation, import and export logistics competitiveness based on recent Logistics (W & D) focused on the expansion based on the LG Group's overseas expansion.
Hyundai Logistics	Overseas business portfolio, logistics consulting, customs clearance, quarantine extended to sector / global SCM system.

Source: theorem.

3 Foreign Companies through the Analysis of the Domestic Logistics Business Strategy Proposed

The results of the present analysis, the cases at home and abroad for the growth of the domestic logistics companies for global logistics companies, foreign companies and domestic companies and overseas markets situation is very pessimistic. Therefore, this study and propose domestic companies in overseas markets.

3.1 Issues Proposed Frame-Work for the Derivation of the Success Factors of Logistics Enterprises

Analysis of the major global companies’ overseas growth strategy in order to achieve economies of scale and increase sales growth was the way for the expansion of the customer, improve logistics service quality, diversified logistics services.

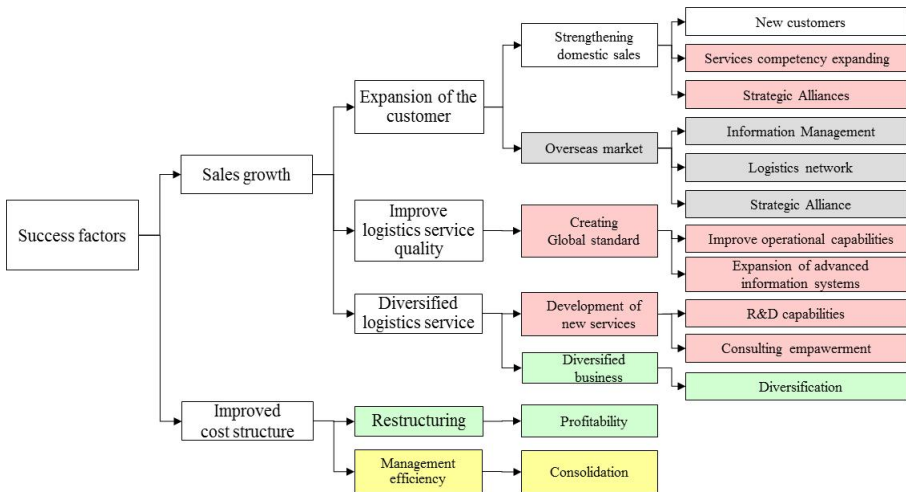


Fig. 2. Frame-work for the derivation of the key success factors

3.2 Modeling Foreign Global / Domestic Logistics Company, Logistics Company's Growth Strategy

Shin CW (2012) in a study on the growth strategy, the company's global logistics. Analyzed the change in injection growth model of a global network of integrated logistics services, specialized services, regional centralization, and global logistics company overseas companies as a model of early growth strategy to strengthen the global network and integrated logistics services. Strategies such applications in the domestic logistics companies, specialized professional services and strategic a centralized regional center of the domestic logistics companies, global network and integrated logistics services to the company's growth strategy should be modeled.

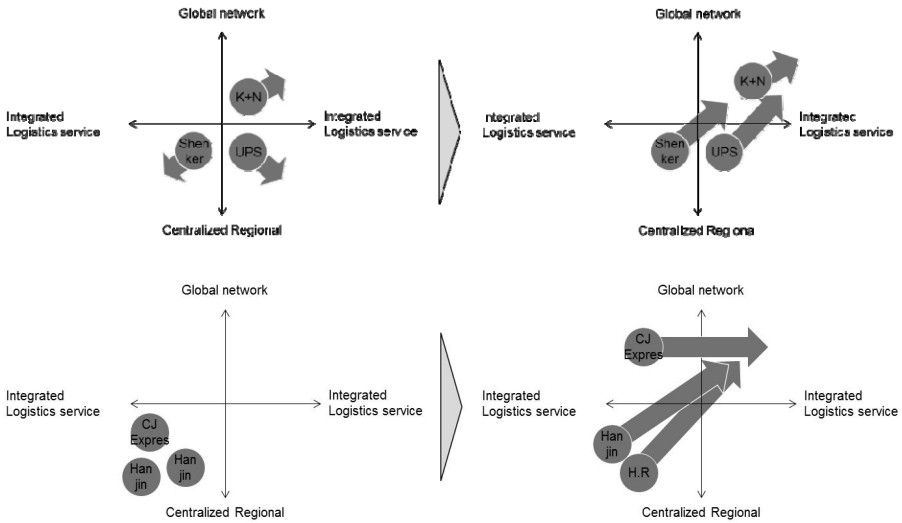


Fig. 3. Modeling foreign global / domestic logistics company, Logistics Company’s growth strategy

3.3 Positioning Strategic Areas and the Entry of the Domestic Logistics Companies and Change of Strategy

- 1) Should be done in a way to advance to the overseas expansion and rapid economic growth in China, India, Southeast Asia, the Americas, Europe, as previously mentioned analysis. In addition, it shall enter of mot businesses such as parcel, W & D, Sea forwarding, Air forwarding, Land transport

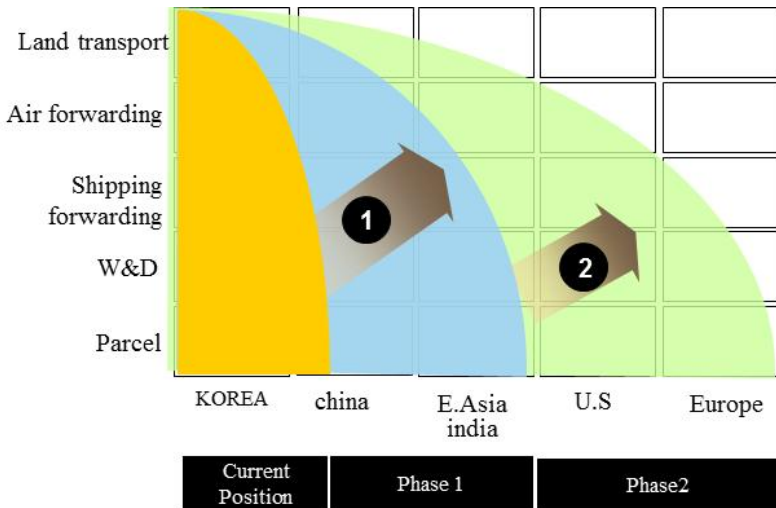


Fig. 4. Positioning strategic areas and the entry of the domestic logistics companies

2) Primarily to advance China's domestic market, overseas expansion strategy can make the modeling of the integration-localized matrix model. Domestic logistics companies are run by the simple strategy of internationalization. Transnational strategy to assimilate into the local market and optimize the value chain in terms of production, in terms of marketing strategy for the writing, so that all of the logistics industry in the future globally integrated pressure and localized pressure increase should be changing.

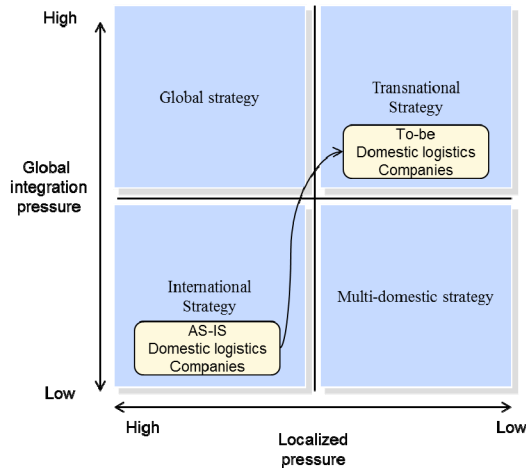


Fig. 5. Model of Integration-localized matrix

4 Conclusions

Quickly in order to respond to changes in the international logistics market, the domestic logistics companies to promote overseas expansion. Manufacturing and overseas markets with expansion of the domestic logistics companies, which is expected to be very difficult to grow as a global logistics company,

Thus, the domestic logistics enterprises should continue efforts to add value and are summarized as follows:

- ① Increase the Value Chain and SCM integration with the ability to service the ability to provide
- ② Secure IT-based and value-added services
- ③ In order to secure a network with local Chinese companies to promote JV & M & A
- ④ China, located in the core business area, secure network.
- ⑤ Expand the company's professional services focused area of competence.

Modeling and the framework of this paper, the current domestic logistics companies expect to be able to have the opportunity to grow as a global logistics company in the fastest time.

References

1. Shin, C.W.: A study on the Growth Strategies of Global Logistics company (2013)
2. DBR: Decision-making Tool it 100 34P-35P (2013)
3. J.F & Gartner Fellow Emeritus: Emerging Technologies: What's Hot for 2012 to 2013 (2012)
4. Park, Y.T., Jeon, H.J.: An Empirical Study on the Effects of the M&A Announcement on the values of Acquirer Firms- Focused on Global logistics Firms (2008)
5. Lee, G.S.: Study on the type of growth strategy of the domestic logistics companies for global logistics companies (2009)

A Dynamic Model of Technological Innovation in 3D TV Industry: Case of LG Electronics

Jun-oh Hwang

Graduate School of Management of Technology, Korea University,
Seongbuk-gu, Seoul 136-701, Korea
nicepeter2011@gmail.com

Abstract. This paper focuses on the existing matured PDP/LCD TV markets and the growing 3DTV market. In recent five years, LG Electronics(LGE) has faced with limitations to keep its leading position continuously in the matured PDP/LCD TV markets, and been chased right under its nose by technology alliances between Japan and Taiwan, the establishment of integrated companies, and the development of new technology in its competitors. Accordingly, this paper has its own research objective to find strategic implications to retake its previous position once known as a renowned TV manufacturer by getting product leadership back through the 3DTV technology innovation and the development of new products. In this paper, we reinterpret success factors for 3DTVs with a dynamic model of technological innovation by applying it to product and process innovations of 3DTVs, and suggest marketing strategies to reduce gaps between technological innovation on the high-tech life cycle of technology and market innovation. The performance and technological innovation of products are important to survive in the high-tech industry. However, convenience, accessibility, and emotional aspect are more important three factors from the consumer's properties to maximize the consumer's utility, which should be considered in the stage of product development. The maximized consumer's utility enables LGE to hold the product leadership consistently by reducing the chasm between the early adapter and the early majority.

Keywords: Dynamic model, Technology Innovation, 3D TV, High-Tech, Chasm.

1 Introduction

1.1 History of 3DTV and Its Market Situation

3D technology has had three opportunities to boost 3DTV industry during the past 170 years, each for every 60 years. The first one came in the last 1890s when 3D movie technology was invented first, but it failed to go the market. The second chance in the 1950s brought a boom for making 3D movies. However, the boom could not last long since the technology failed to clear the problem of fatigue from watching 3D movies and the advent of black-and-white TV took interests of the public away from

the 3D movies. Currently, we are in the period of 3DTVs popularized to people due to the third opportunity, moving the 3D industry to home. This has been possible because of both the popularity of 3D movies ignited by Avatar (2009) and improvements in the 3D technology [3].

(1) Definition of 3D technique

Of various 3D techniques, it is common to use the binocular disparity method to help viewers to feel the perception of 3D depth by presenting two offset images from contents containing each left and right movie images separately to the left and right eye of the viewer. The 3D displays on markets at present are for toys, arcade games, television broadcasts, and 3D movies, etc [7].

(2) Types of 3D display techniques

Display methods to display 3D images are categorized into Stereoscopic displays, Head-mounted displays, and Holography. Most stereoscopic displays comprise of CRT, LCD, OLED, or PDP, and can display perceptual depth in space onto the front and the rear of the screen by dividing each 2D image to two separate images of the binocular disparity and by presenting them to the left and right eye of the viewer. The stereoscopic display method employs two different types of displaying 3D images: stereoscopy requiring wearing special glasses and autostereoscopy without the need of glasses. There are two different methods in the stereoscopy: Patterned Retarder and Shutter Glasses. In South Korea, LGE adopts the patterned retarder method, but Samsung employs the shutter glasses method, once resulted in many disputes over those in both companies in early market [7].

2 Literature Review of Technology Innovation

2.1 Technology Innovation

Taton[16] & Schumpeter[15] propose that the development and progress of a technology is a change led by a minority or a natural phenomenon. Schmookler[14] & Merton[10] assert that technological progress results from economic demands and growth. Morison[11] & Sahal[13] insist the need of new integrated approaches because theories and analyses on the past technological development make errors that they cannot explain the complexity of a change in technology by a single theory, and technology advances through the continuous response by the mutual interaction among history, individuals, and market demand. Dosi et al[4] explain change through innovation by a meaning which includes an action leading to a new discovery as well as manufacturing technology and product attributes and its economic usage. Porter[12] insists that innovation is a new process for commercialization, and this process cannot be evaluated by separating the strategic aspect of a firm and its competitiveness. Also, he notes that in terms of technology and demand, new knowledge through innovation is closely related to a market. Afuah[2] notes that in a demand-side within supply and demand, innovation takes advantage of new knowledge to provide new products or services, and this can mean a combination of creation and commercialization.

2.2 Dynamic Model

(1) Dynamic Model of Production and Process

Fig.1 explains how a production innovation and process innovation change as time goes using three steps in a paper. In business aspects, this model can details a change arising in the period of technological evolution starting from a fluid phase representing the flow of dynamical changes to a transition phase to a specific phase[1][17]. Also, it can be applied well to the TV industry, an area of interest in this paper.

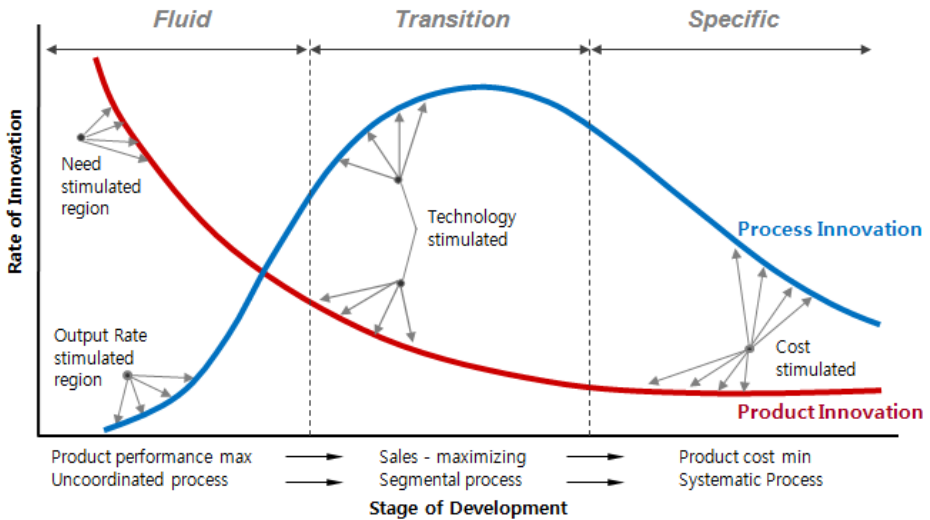


Fig. 1. Innovation and stage of development [17]

(2) Technology Life Cycle Model

Afuah[2] notes on the model that the degree of technological uncertainty can vary depending on both the complexity of a technology and its changing status in his paper.

(3) S Curve

M. Foster[8] suggests a new theory which provides a method to predict the end of existing technologies and the advent of technological discontinuity and predicts technologies of discontinuity known unmanageable to some degree.

(4) Interaction Model

The model interprets continuous interaction and feedback as important characteristics in the process of an innovation. Especially, it suggests that design plays an important role in an innovation, and emphasizes both interactions: one between downstream and upstream and the other between technologies in each stage of an innovation and innovative activities [8].

3 Reinterpretation and Expansion of Dynamic Model

3.1 Reinterpretation

For decades, a dynamic model has proven successful in many applications and improved continuously. To apply the model to 3DTV, a high-tech product, the state of changes in technology by the model is divided into three stages in product innovation, and in process innovation, the model is modified and reinterpreted in the perspective of Quality, Delivery, and Cost to achieve the level of productivity for mass production needed for the product innovation as shown in Fig.2 & Table 1.

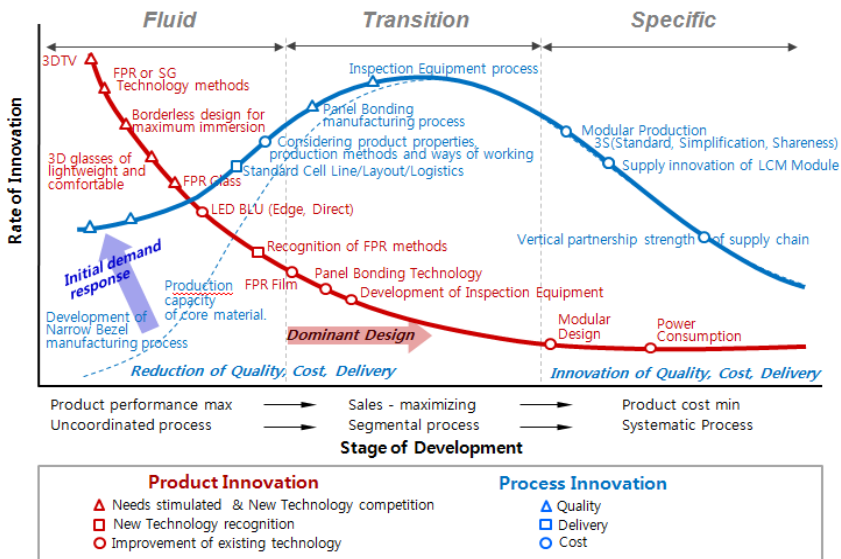


Fig. 2. Dynamic Model of product and process innovation in 3DTV

Table 1. Innovative activities of 3DTV development

	Fluid (Product Innovation)	Transition (Process Innovation)	Specific (Specific Innovation)
Innovation Characteristics	<ul style="list-style-type: none"> Product User Products change frequently Original Technology, Product Performance, Product Release Date 	<ul style="list-style-type: none"> Set Maker, End User Severe changes in the production process due to the increase in demand Variety of products 	<ul style="list-style-type: none"> Part Maker Incremental product improvement, continuous improvement in quality and productivity Cost competitiveness
Production Innovation	<ul style="list-style-type: none"> FPR vs SG Technology Borderless design for maximum immersion 3D glasses of lightweight and comfortability FPR Glass Technology LED BLU (Edge, Direct) 	<ul style="list-style-type: none"> Recognition of FPR methods FPR Film development to save cost Panel Bonding Technology Development of Inspection Equipment Emergence of Dominant Design 	<ul style="list-style-type: none"> Modular Design Power Consumption
Process Innovation	<ul style="list-style-type: none"> Development of Narrow Bezel manufacturing process Production capacity of core material 	<ul style="list-style-type: none"> Inspection Equipment process Panel Bonding manufacturing process Consideration product properties production methods and ways of working Standard Cell Line/Layout/Logistics Mass-production Stabilization of Supplier 	<ul style="list-style-type: none"> Modular Production 3S (Standard/Simplification/Share) Diversification of LCM Module suppliers Vertical partnership strength of Affiliates

In the fluid stage, the characteristic of a product changes frequently, mostly driven by users of the product. In the product innovation perspective, two main types of 3D technology are both the SG (Shutter Glass) and the PR (Patterned Retarder) type. The Shutter Glass type can keep relatively high level of resolution in a 3DTV by presenting a left or right image alternatively to a viewer, called time division method while it can result in dizziness, and requires high cost glasses, disadvantages to its popularization. The Patterned Retarder type, also called space division switching, splits the screen into many sections, and sends left and right images at the same time, and then separates the left/right signals through a patterned retarder film that is attached to the TV panel. Despite its strengths such as no harm to eye health, high brightness, and cheap glasses, both reduced resolution and high cost glass filter limit its popularity. However, market demand on 3DTV has become steadily increasing due to 3D innovation which changes our perspective of TV from just viewing to experiencing existing flat panel display (Flat TV). Accordingly, for initial demand response to the market demand, the process innovation for 3DTV has appeared to secure sales through manufacturing process development for Narrow bezel and securing production capacity for core materials, resulting in a modified model where the process innovation curve in the previous dynamic model has risen upwardly.

In the transition stage, substantial change in the process innovation appears, resulting from production volume increase. Also, 3D technology is recognized to consumers, and dominant design appears in the stage. To have competitive prices for increase in the demand, standard work procedure reflecting characteristics of a product, the installation of assembly lines of a production method, and automation of panel bonding and inspection equipment should be developed. Also, layout for optimal utilization of those and logistics system design should be made, and partners supplying core materials stably should be secured.

In the specific stage, a product enters in the mature phase, and two types of cost innovation are applied intensively to manufacture a standard product. First, customer's desire for product diversity is satisfied by a modular design, and cost reduction follows the minimization of manufacturing complexity in an internal firm. Second, cost leadership is achieved by increasing LCD module supply efficiency through vertical partnership among subsidiaries. Here, the vertical partnership among subsidiaries means the strategic supply relationship for TV manufacturing business maintained among panel maker, parts maker, and set maker in the positions of internal subsidiary companies.

3.2 Expansion

Based on the reinterpretation of the 3DTV's product and process innovation phased into three stages, respectively, to which the dynamic model proposed previously was applied, this chapter suggests a marketing strategy for entering mainstream market from initial market, and applies the model after expanding its scope of interpretation. The technology adoption life cycle in Fig.3 models five adopter categories presented orderly according to time sequence, and a gap exists between every two neighboring categories. A gap between Early Adopters (❶) full of new technologies and Early Majority (❷~❹), on the contrary, incapable of accepting new technologies is known as 'Chasm', and both categories require different marketing strategies [9].

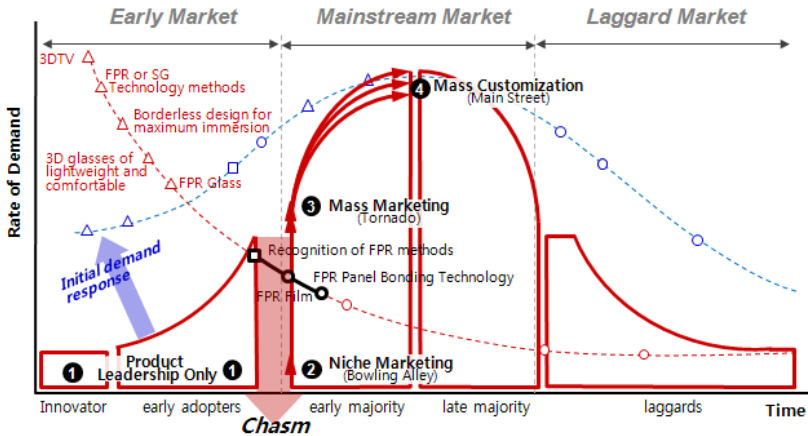


Fig. 3. Technology adoption life cycle of 3DTV

① Product Leadership Only

Samsung Electronics, the existing leading TV set manufacturer, launched the 3DTV of the Shutter Glass (SG) type by focusing its brand power and technological prowess as its strengths. A full-scale competition between the global first and second TV set makers has begun as soon as LG Electronics, the first follower, entered the 3DTV market with the Film Patterned Retarder (FPR) type. The 3DTV market became unstable, and the conditions lasted long because of cut-throat marketing competition among 3DTV set makers in South Korea and confused early adopters over a question of which company they would purchase a 3DTV from. The situation eventually resulted in an experiment in ‘Consumer Report’ to compare and evaluate the SG type 3DTV and the FPR type 3DTV, a program which reports consumer complaints or related issues broadcasted via Korean Broadcasting System (KBS). Total forty consumers participated in the experiment where the following properties were compared between both types such as optical test, subjective test, resolution, three-dimensional effect, color sense and definition. As the result of the experiment, LG Electronics was awarded five ‘Excellence’ at resolution, three-dimensional effect, definition, and fatigue, out of eight evaluation categories while Samsung Electronics received only three ‘Excellence’ as its result[6]. Consequently, LG Electronics won a decision over Samsung Electronics with the FPR type in the competition of 3DTV technologies, which motivated LG Electronics to overcome the ‘Chasm’. To succeed in the mainstream market, although LG Electronics targeted successfully the early market, we need totally different approaches differentiated from those in the early market, and thus propose three strategies for it: (1) Niche Marketing; (2) Mass Marketing; and (3) Mass Customization’s strategy

(1) Strategy of Bowling Alley (②): Product Leadership + Customer Intimacy
 The bowling alley is the period during which the new product gains acceptance in niche markets within the mainstream market, but has yet to achieve general, widespread adoption [5]. So, LG Electronics improves the level of completion in

manufacturing 3DTV's standardization with the focus on its strengths, both Film Patterned Retarder technology and Narrow Bezel Design, and increases Customer Intimacy through the verification of emotional evaluation on 3DTV from the consumer assessment conducted at KBS [6].

(2) Strategy of Tornado (☉): Product Leadership + Operational Excellence

The tornado is a period when the general marketplace switches over to the new technology [5]. LG Chemical has established a beachhead for distributional 3DTV since they developed the Film Patterned Retarder film, eventually bringing competitiveness to them in cost and development lead time. Also, together with the technology, automation of Panel Bonding and Inspection Equipment enabled them to supply the products faster.

(3) Strategy of Main Street (☑): Operational Excellence + Customer Intimacy

To secure a stable position in the mainstream market, LG Electronics needs to expose the 3D standard technology, the FPR type, to advertisement intensively to maximize Customer Intimacy, and should optimize operation efficiency by standardizing the production system of 3DTV.

4 Conclusions

4.1 Success Factors

Although Samsung Electronics encroached on the 3DTV market with its brand power and technological prowess, LG Electronics, fast follower, entered the market by appealing to the emotional aspect of users with the light and comfortable glasses. In the consumer assessment mentioned above, it is verified that the most influencing factors in the development of the high-tech product are convenience, accessibility, and emotional aspect, properties of the consumer [6]. Added to this, the dominant position in the high-tech market environment does not always guarantee favorable responses from consumers. Favorable responses on the emotional aspect from the market kindled sales to go up from the minus operating profit in 2010 to about plus 3% in 2011 4Q[www.lge.com].

The strategic collaboration among affiliates in LG Group can be the main success factor to LG Electronics. The successful joint work was possible due to those affiliates: LG Innotek - LED raw material, Heesung Electronics – BLU (Back Light Unit, LG Chem - FPR 3D Film, making possible to improve cost competitiveness and shorten development lead time, LG Display – FPR 3D Panel Bonding technology, and LG Electronics – final 3DTV assembly. With the joint work, the cheap polarized glasses which do not need battery contributed to the success by maximizing user convenience. Samsung and LG Electronics competed fiercely in the early market. Nevertheless, it is believed that this competition made both companies become the global first and second TV set makers in the world TV market. Finally, it is expected that LG Electronics fights well to regain its reputation in the past, a renowned TV set maker.

4.2 Implications and Limitations

This paper applied the dynamic model to the phased product and process innovation, and reinterprets its meaning. Also, the scope of interpretation for the model expanded to present a strategy to recover 'Chasm', a big gap in the technology adoption curve mentioned previously. Changes in the process innovation curve at the dynamic model show that the product life cycle of TV in the high-tech environment has shortened, and the sales from initial demand is strongly sensitive to the release date of the product[Fig.3]. Also, we can understand from the changes that a marketing strategy to overcome the 'Chasm' moves from product and technology-oriented in the early market of TV industry to market demand-oriented.

In conclusion, the application of the dynamic model and the technology adoption curve explains well the four areas of high-tech market environment such as product, technology, market, and a firm. However, the limitation of the model can be the difficulty of figuring out the start and end point of activity factors such as a dominant design from each time point in both the dynamic model and the technology adoption curve. Also, entry strategies to the mainstream market depending on a dominant design needs to be divided for the first mover and fast follower, respectively.

References

1. Abernathy, W.J., Utterback, J.M.: Patterns of innovation in technology. *Technology Review* 80(7), 40–47 (1978)
2. Afuah, A.: *Innovation Management*, 2nd edn., Oxford (2003)
3. Displaysearch, Monthly report.: 3DTV Industry (2010)
4. Dosi, G., Freeman, C., Nelson, R., Silverberg, G.: *Technical Change and Economic Theory*. Pinter, London (1988)
5. Jakki, M., Sanjit, S., Stanley, S.: *Marketing of High-Technology Products and Innovations*, 3rd edn. Pearson Press (2010)
6. Korean Broadcasting System.: *Consumer Report, KBS TV* (2011)
7. Lee, J.: *Display Industry*. Hana Daetoo Securities (2010)
8. Foster, M.: *Innovation* (1990)
9. Moore, G.A.: *Crossing the Chasm*. Harper Business (1991)
10. Merton, R.K.: *Social theory and social structure*, New York/London (1968)
11. Morison, E.E.: *Men, Machines, and Modern Times*. MIT Press, Cambridge (1966)
12. Porter, M.E.: *The Competitive Advantage of Nations* Macmillan, London (1990)
13. Sahal, D.: *Patterns of technological innovation*. Addison-Wesley (1981)
14. Schmookler, J.: *Innovation and economic growth*, Cambridge, Mass. (1966)
15. Schumpeter, J.A.: *The development economics*. Oxford Univ. Press, Oxford (1961)
16. Taton, R.: Reason and Chance in Scientific Discovery. *Academic Medicine* 33(6), 513 (1958)
17. Utterback, J.M., Abernathy, W.J.: A Dynamic Model of Process and Product Innovation, *Omega*. *The Journal of Management Science* 3(6) (1975)

The U-Work Utilization Analysis Using Groupware on Non-profit Organization

Kyeong Hui Du¹ and Chulung Lee^{1,2,*}

¹ Graduate School of Information Management and Security, Korea University, Korea

² School of Industrial Management Engineering and Graduate School of Management of Technology, Korea University, Korea
{mycathy, leecu}@korea.ac.kr

Abstract. Currently in Korea, a non-profit organization is separated from the government so that it is difficult to obtain sufficient statistics about it. The non-profit organization evolves constantly, the introduction of information systems has had a great effect. The need for research about it specifically has emerged. The non-profit organization has the management philosophy of "People-Service-Value". In other words, it is not only to pursue short-term profit companies for the purpose of simply but also for people and communities to increase the value. To do so, high personnel dependence, communication management, and administrative services, is essential. Therefore, in this study, I research the leading non-profit organization's information systems through use cases of u-Work. Through the actual deployment of IT services, I would like to help when introducing a national non-profit organization's information systems in the future.

Keywords: Groupware, U-Work, Non-profit Organization, IT utilization, Group socialization process, Weblog analysis, Ubiquitous, Information interaction, Accessibility, Voluntary participation, Co-operation.

1 Introduction

The non-profit organization is not only on a commercial basis but also on distribution of profit like private corporations. It does profit-making activities within the range that is not contrary to its identity. The scale of non-profit organization is in a thousand different ways. The one feature of non-profit organization in common is not to want monetary compensation but to help achieve each individual's innate personality and talent for beliefs and values. The costs and resources required to operate is based upon voluntary participation and depend mainly on the voluntary organizations. According to the 'Independent Sector' survey in 2001[1], 50% of all Americans are actively engaged in volunteer activities and they spend an average of 4 hours per week.

* Corresponding author.

The IRS in Korea announced that a public benefit corporation is 28,900 in 2009 and 62% of these are religious institutions.[2] Religious institutions not only serve their neighbors and inspire the conscience of the community but also motivate people and provide guidance and activities [3]. One of religious organizations in South Korea, Onnuri Community Church[4], the number of members increased to 5 million, but more than 20 years in the 12 families. in 1985. I analyze the u-Work environment utilization through information systems of the Korean Christian representative non-profit organization, Onnuri community church.

Whereas minority, managers and staff of the non-profit organization are tinged with internal accountability and work with professionalism as a salaried, voluntary organizations, voluntary organizations are the majority and contribute to the organization through the delegation of authority to an unpaid voluntary basis.[5]

The paid employees of Onnuri Community Church are divided into full-time staffs as the U-Worker and a group of pastors who are responsible for the ubiquitous ministry and counseling with smart phone and notebook. A voluntary organization consists of the leadership group and voluntary workers. The former having the right of electronic approval substantially participate in decision-making. The latter is delegated works under the ministry. I thought that the two of the wheel of the conflict is mediated by self-motivation and co-operation. I verified how the specific character of group socialization process makes an impact on ubiquitous trait, accessibility and information interaction, voluntary participation and co-operation.

The purpose of the IT is to support the business, to be used as a strategic tool, and to build a well-equipped unified architecture system for minimizing the gap between business and IT and to respond quickly for it. [6] I verify with the group socialization process trait of u-Workers, such as career and duty composed preceding variable by adding groupware trait and ubiquitous trait on study model in this study.

1.1 Theoretical Background

1) Group Socialization Process

Private non-profit organizations are intricately intertwined by the responsibilities and rights of the members with a variety of personalities. These organizations become source of conflict having ambiguity of responsibility, multiplicity of recognition and rewards system for volunteers.[7] To be most democratic, autonomous, and active organizations are most authoritative, inefficient and stuck in a rut [8]. Richard Moreland and John Levine presented a useful model about Group socialization process. It involves the process of initiation, affiliation, growth, and contribution for a common goal through relationship building within the group.

2) U-Work

The U-Work is a term that is a combination of 'Ubiquitous' and 'Work'. It is the new type of work that workers can perform efficiently under ubiquitous computing environment away from the constraints of time and place, utilizing information and communication technologies (Korea National Computerization Agency, 2005).

3) Groupware

Groupware is the concept which is raised from CSCW(Computer Supported Cooperative Work) in the U.S. in 1986. It supports communication and collaboration and coordination between the tasks.[9] The technical system such as groupware reduces the communication gap that always occur in larger and distributed teams and help to perform complicated tasks.[10]

4) Weblog Analysis

Bob White, VP Global Enterprise Initiatives at Ingram Micro Inc, IT distributor in Canada,. noting the attractiveness of a data integration of weblog information (click-stream) and actual information, applied the analysis information to the operation. In practice a variety of activities taking place on the Web site through weblog analysis improve availability and can be applied to the actual operating.

5) Information Interaction

Since the 1990s, personal word processor and independent information system was developed. It was difficult to share files scattered on each PC on non-standardized platform , stored and managed within its own organization.(Data Silo effect). The professor, Peter F.Drucker explained that the decision-making should be made at the lowest level within the organization, ie, at the nearest place executed the result of the decision-making, using a authority delegation model.

6) Team Co-operation

Non-profit organizations are the most labor-intensive in the third sector. It is not easy to integrate into the employee gathering is not easy because diffuse members are small- and medium-sized.[11] The use of groupware electronically based in real-time makes it possible.

2 Main Subject

2.1 Designing Process

1) Research Model

The model was constructed how group socialization process(career, age and duty) affects the typical characteristics of groupware such as ubiquitous trait, accessibility and information interaction, voluntary participation and co-operation of u-worker as shown in Fig. 1.

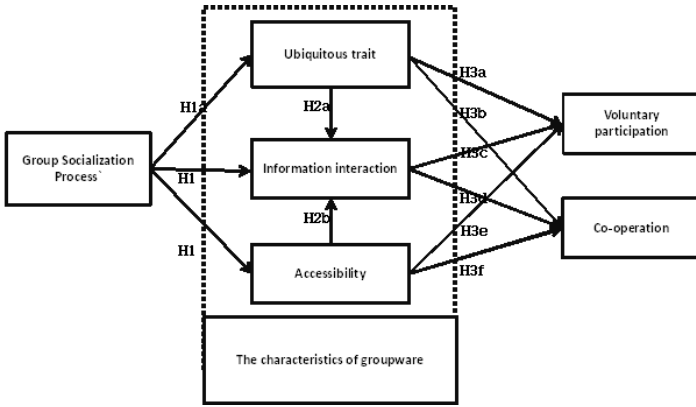


Fig. 1. Research model

2) Building Up Hypothesis

In this paper, I set up a hypothesis as following:

H1a: Group socialization process trait will have a positive(+) impact on ubiquitous trait, one of the typical characteristics of groupware.

According to Sung-min Cho and Sun-ro Lee(2008)’s study, remote working using ubiquitous computing environment is the new type of work based information systems. This study presented research model about having characteristics of remote working (Job Characteristics, absentee rate, turnover rate of expert group) an impact on ubiquitous trait.[12]

H1b: Group socialization process trait will have a positive(+) impact on information interaction, one of the typical characteristics of groupware.

H1c: Group socialization process trait will have a positive(+) impact on accessibility, one of the typical characteristics of groupware.

Non-profit organizations free from formalities and having greater autonomy will communicate transcending time and space when the network is connected. \

H2a: Ubiquitous trait, one of the typical characteristics of groupware will have a positive(+) impact on information interaction.

Businessperson under ubiquitous computing environment can freely connect to the network freely. He can retrieve information, report the work, and send a message while also moving outside the office. He will be highly aware of the need for information interaction.

H2b: Accessibility, one of the typical characteristics of groupware will have a positive(+) impact on information interaction.

Ease of logical approach has nothing to do with the actual distance but ease of access can be seen to represent the geographical characteristics.[14] Information interaction contributes to alleviating the conflict by enabling communication between members. Information interaction including formal and informal or public and private methods can be a means which members within the conflict of organization can respond covertly.[15]

H3a: Ubiquitous trait, one of the typical characteristics of groupware will have a positive(+) impact on voluntary participation.

The majority of users through a familiar Web browser on vacation, on a business trip or at home can participate in the decision making accessing the groupware framework(e-mail, real-time messaging, online conferencing).

H3b: Ubiquitous trait, one of the typical characteristics of groupware will have a positive(+) impact on co-operation.

H3c: Information interaction, one of the typical characteristics of groupware will have a positive(+) impact on voluntary participation.

Volunteers have sufficient ability and potential. In some areas they may have more information and knowledge and experience than paid staff.[16] Also, when you use frequently to the information system, by perceived usefulness they can avoid the hardness and heaviness of communication and get voluntary feedbacks.

H3d: Information interaction, one of the typical characteristics of groupware will have a positive(+) impact on co-operation.

On developing IT, researched by very advanced scientific analysis, the relevance between the variables about survey data and the reliability of the statistical relationship will help a working-level communicator to make better decision-making. [17]

H3e: Accessibility, one of the typical characteristics of groupware will have a positive(+) impact on voluntary participation.

H3f: Accessibility, one of the typical characteristics of groupware will have a positive(+) impact on co-operation.

Rodden and Blair described groupware systems into three categories according to the form of co-operation. First, simultaneous system through electronic conferencing and desktop conferencing system, second, non-simultaneous system such as E-mail, messaging, scheduling, and e-payment system, third, mixed system like document revision control system and electronics document management system.

2.2 Research Method

1) Sample Data Collection

In order to measure the factors affecting u-Work, I analyzed to reconstruct the weblog analysis and user DB on groupware database system. The period under this study is

12 business days on March 18, 2009 - March 30, 2009. The research subjects are 582 members of formal registered 67,012 members at Onnuri Community Church[18] They are key decision makers who can access to electronic approval system. I used by adopting the effective sample this in this study.

2) Demographic Characteristic

According to the demographic characteristic of the sample of 582, gender of survey were far higher 482 men (82.81%) than in 100 women (17.18%). The greatest rate of age group was 150 people in their 30's, accounting for 25.77 percent. By a narrow margin, following 140 people (24.05 %) in their 60's was 133 people (22.85 %) in their 50's, 66 people (11.34 %) in their 40's, and 33 people (5.67 %) in their 20's. According to educational background, following 283 people (48.62 %) in college graduates was 201 people (34.53 %) in a master's degree or in graduate school, 60 people (10.3 %) in a doctor's degree, and 38 people (6.52 %) in high school diploma below. For reference, at least MDiv(Master of Divinity) is depending on pastor's qualifications. According to work experience, following 137 people (23.53 %) in 16~20 years was 129 people (22.16 %) in 1~5 years, 119 people (20.44 %) in 11~15 years, 106 people (18.21 %) in 21~25 years, and 91 people (20.44 %) in 6~10 years. According to assigned duty, following 303 people(52.06%) in volunteers and 279 people(47.93%) in paid staffs.

3) Reliability and Validity Test

The result of performing a reliability test for each of the research variables used in this study by calculating Cronbach's alpha is shown in Table 1. It can be seen that the measurement items of the variables in this study are internally consistent and reliable if the value of Cronbach's alpha is above 0.7 as shown in Table 1.

This result of simple correlation analysis of research variable is shown in Table 2. Group socialization process trait, one of the characteristics of information system doesn't affect much in ubiquitous trait and accessibility. The ubiquitous trait and accessibility affects on information interaction. The ubiquitous trait has a high effect on voluntary participation. The accessibility has a low impact on information interaction.

The information interaction has a high effect on co-operation. The accessibility has a low impact on co-operation. The information interaction affects on voluntary participation and the information interaction has a high effect on co-operation. The ubiquitous trait acts on accessibility, voluntary participation as an important variable. The cross analysis process about the group socialization process trait affecting on accessibility described that 'full members(11~ 20 years)' are most accessible. The cross analysis process about sex affecting on information interaction described that man has higher information interaction. According to the group socialization process, following 'new member' is 'full member' and 'marginal member' in voluntary participation. The longer the work experience, the less the voluntary participation. According to the group socialization process trait, following 'new member' is 'full member' and 'marginal member' in co-operation trait. The longer the work experience, the less the co-operation except electronic approval function.

Table 1. Validity test

research variable	Cronbach'salpha
Group Socialization Process (GSP)	0.776
Ubiquitous(UBI)	0.706
Information Interaction(INF)	0.788
Accessibility(ACC)	0.755
Voluntary participation (VOL)	0.809
Co-operation(COP)	0.753

Table 2. Simple correlation analysis of research variable

	GSP	UBI	INF	ACC	VOL	COP
GSP	1					
UBI	-.434	1				
INF	-.201	.397	1			
ACC	-.382	.876	.349	1		
VOL	-.493	.945	.377	.846	1	
COP	-.188	.373	.735	.251	.337	1

** P<0.01.

3 Conclusions

Depending on the group socialization characteristics of non-profit organization, I added ubiquitous trait, accessibility, and information interaction of groupware characteristics to the research model and validated them in this paper. The information interaction, accessibility, and ubiquitous trait of u-Workers is useful and affects on information interaction. The non-profit organization has no retirement age, so average age band is quite high. But the member contributing to co-operation and voluntary participation is ‘new member’ and ‘full member’ such as pastors and paid staff. By using mandatory function such as electronic approval and vote system, and teaching IT education, I expect that ‘marginal member’ of volunteer organization will become active to be re-socialized.

Groupware is the communication system to be performed in the most basic first step to configure this integrated. Hereafter, beyond the framework of the existing combination of a ubiquitous trait and accessibility, it will support business as living communication of the entire organization, be used as a strategic tool and minimize gap between business to respond quickly.

This study has some limitations as follows. First, it was difficult to determine the time of sampling as the application period went 16 months or more after the opening because of the characteristics of organizational culture. Second, limited to a single organization, it should be extended and applied to a variety of non-profit organization. This will be complemented in the next study. Due to the rapid changes of the u-work surrounding environment, the communication of the non-profit organization and the development phase of information system such as an e-business intelligence system demand successive studies.

References

1. Giving and Volunteering in the United States, Independent Sector (2001)
2. Average and Median Amounts of Household Giving & Volunteering in 2002, Center on Philanthropy at Indiana University (2006)
3. Woo, J.-R.: Alternative jobs for the Active Senior: Start the Act 2 of your life in a non-profit organization, 54 p., Economist (2012)

4. Cha, M.: Analysis of the growth of the Church from John Fiske's popular culture standpoint - Onnuri Community Church, 1 p. (2008)
5. Anheier, H.K.: What is the 3rd sector: A comparative study on the non-profit organization, 211 p. Arkhe press (2002)
6. 2009 Spring Oracle Korea Magazine, 24 p. (2009)
7. Lee, S.-R.: Conflict Management in the private non-profit organization, 98 p., 343 p. Media Soop press (2007)
8. Moreland, R.L., Levine, J.M.: Group dynamics over time: Development and socialization in small groups. In: McGrath, J.E. (ed.) *The Social Psychology of Time: New Perspectives*, pp. 151–181. Sage, Newbury Park (1988)
9. Kang, J.-K.: Effective project management model using groupware under the distributed work environment, 7 p. Korea University (1998)
10. O'Connell, B.: *Independent Sector - The history of the nonprofit sector in the United States*, 225 p. Arkhe press (1997)
11. *Independent Sector - The history of the nonprofit sector in the United States*, 301 p. Arkhe press (1997)
12. Cho, S.-M., Lee, S.-R.: A study on the effects of information system characteristics on job satisfaction and job performance under u-Work environment. In: *2008 Proceedings of the Korea Management Information Systems Institute* (2008)
13. Vroom, V.: *Work and Motivation*. Wiley, New York (1964)
14. Kim, S.-H.: *Comparative analysis between groupware systems*. Korea University (1997)
15. *Teams, Markets and Systems*, 194 p. Ciborra, Cambridge University (1993)
16. Lee, S.-R.: *Conflict Management in the private non-profit organization*, 157 p., 343 p. Media Soop press (2007)
17. Yoo, Y.-D.: *Research on the communication performance of a non-profit organization*. Korea University (1994)
18. *Statistics of Information System Lab at Onnuri Community Church* (2009)

Analysis of the Risks of Overseas Advancement by Logistics Companies Applying AHP

Mihyung Kim¹, Ki-sung Hong¹, and Chulung Lee^{2,*}

¹ Graduate School of Information Management and Security, Korea University, Korea

² School of Industrial Management Engineering and Graduate School of Management
of Technology, Korea University, Korea
{mh_kim, justlikewind, leecu}@korea.ac.kr

Abstract. This study deduces the risk factors that may arise when logistics companies make inroads overseas ensuing globalization and calculates the importance of such risk factors. Risk factors were derived by reviewing literature, advanced research analysis, expert advice and marketing research; and the importance calculated applying AHP by surveying those in the logistics business with experience advancing overseas on the deduced risk factors to yield a means to manage high importance risks. In addition, representative overseas advancement strategies of logistics companies were analyzed and presented by region based on such risk factors.

Keywords: Logistics, AHP, Risks of Overseas Investment, Overseas Advancement Strategy, Risk Factor Analysis.

1 Introduction

Limitations in profit guarantee due to oversupply of the domestic logistics market and continued expansion of the global logistics market scope ensuing globalization of management and block formation is averting the gaze of the domestic logistics companies to the overseas logistics market. However, as company sales activities through overseas investment are achieved in an environment different from that in domestic market, far more risks compared to domestic sales accompany. Investing overseas without exhaustive research on the investment environment and risks beforehand will inevitably result in difficulties in local operations; and what's worse, failure may be result, withdrawing without even collecting on the principle investment if investment risks are handled negligently. Accordingly, there is a need to analyze potential risks exhaustively for successful advancement of logistics companies overseas.

This research analyzes decision making priorities on risk factors that may potentially arise when logistics companies advance overseas applying Analytic Hierarchy Process (AHP). First, the risk factors that may potentially arise when logistics companies advance overseas were determined by surveying groups of

* Corresponding author.

logistics company experts beforehand and the priorities of such risk factors were deduced by piecing together the opinions of experts through marketing research applying AHP on the selected factors. In addition, credible results based on analysis findings on strategy building are provided for successful foreign investment by logistics companies considering advancement overseas.

This paper is largely comprised of 4 sections. The 2nd section investigates the existing research on general overseas investment risks. The 3rd section deduces the risk factors of overseas investment by logistics companies and calculates the importance of such risk factors deduced applying AHP and presents pertinent suggestions. Lastly, the 4th section draws conclusions.

2 Risks of Overseas Investment

Overseas investment risks are mainly expressed as political risks in international administration and while investment risks are frequently debated in numerous literatures, no consensus on the definition exists. Generally, looking at literature or research papers published up until now, defined risk concepts can be divided largely into two groups. One being risks seen according to local government regulations on activities of foreign investment companies and the other being apprehensive political, social, economic factors of the local country that may disadvantageously impact foreign investors in engaging in sales activities overseas.

Choi [1] recently defined the risks, classifying general risks of overseas investment as either a business or management risk. Business risks were divided into 4 categories; namely, change in management environment, change in internal politics, intellectual property rights management and legal/regulation environment. Management risks, those arising from the interior value chain, are categorized as R&D, production, sales, finance/accounting and HR. Jun [2] categorized potential risks of logistics companies ensuing globalization into 6 categories, namely logistics environment, distribution finances, distribution costs, customer service, logistics activities and logistics information, depending on the characteristics of the risks.

3 Risks of Overseas Advancement by Logistics Companies and Analysis of Importance

This study deduces the risk factors of logistics companies in advancing overseas considering overseas investment risks deduced in preceding research papers and risks of advancing overseas from globalization of logistics companies as well as through expert advice and surveying. Based on the data collected through survey, importance of risk factors was deduced applying AHP. This study was carried out in the following process.

Firstly, the risk factors deduced by examining literature and advanced research involving risks of overseas investment and penetration were reclassified through a qualification process targeting domestic logistics companies and experts in the related field. Secondly, risk factors were quantified by surveying experts and persons

involved in logistics companies with experience advancing overseas. As for the analysis method, relative importance was deduced through pairwise comparison of risk factors through AHP analysis.

3.1 Risk Factors of Logistics Companies Advancing Overseas

In this study, the level of risks of logistics companies in advancing overseas were largely classified into 2 categories as management or environmental risk and the risk factors in each category were reconstituted. Risk factors were deduced in detail by category by studying literature as well as through expert advice and market research. Through such process, 11 risk factors were deduced in 2 fields as outlined in Table 1.

Table 1. Risk Factors of Overseas Advancement by Logistics Companies & Impact of Such Risks

Risk Category	Risk Factor	Risk Impact
Management Risk	Procurement of local funds	Difficulty in procuring local funds due to rising interest on loans or upward adjustment of cash reverse ratio
	Local manpower management	Deteriorated profitability from rising labor costs, technology leakage, snag in business operations or deteriorated quality
	Lack of local info	Difficulty in analyzing business validity or establishing an investment strategy due to lack of info beforehand
	Rise in local production costs	Deteriorated profitability from rising production costs such as increased oil costs, etc.
	Intensified regulations on foreign capital companies	Deteriorated profitability from rising taxes and reduced benefits, modification of initial investment plan goals such as extension plans, etc.
	Intensified competition with local companies	Market loss or deteriorated profitability from loss of price competitiveness
Environmental Risk	Lack of logistics infra	Deteriorated company profitability due to rising distribution costs from lack of logistics infra such as insufficient electric power, roads, etc.
	Poor business environment	Deteriorated company profitability from decline in social transparency or excessive government intervention

Table 1. (continued)

	Political risk	Deteriorated profitability due to relocation of government body in charge of logistics facilities
	National risk	Anxiety of laborers and immense cost of potential loss resulting by high uncertainty of political insecurity factors
	Security risk	Deteriorated corporate image due to failure to transport or rise in transport costs from security problems such as terror, etc.

Examining by category, management risks are composed of a total of 6 risk factors including procurement of local funds, local manpower management, lack of local info, rise in local production costs, intensified regulations on foreign capital companies and intensifying competition with local companies. As for environmental risks, there are a total of 5 risk factors including lack of logistics infra, poor business environment, political risk, national risk and security risk.

3.2 Analysis of Risk Factors

Priorities of the 11 risks deduced as risks of logistics companies in advancing overseas were derived calculating the weight of each risk by comparing each risk. The importance of the two types of risks as shown in Table 2 was compared by surveying logistics experts. Through a total of 33 surveys collected, AHP was applied based on the derived level of importance.

Table 2. Relative Comparison of the Risks of Logistics Companies Advancing Overseas

Existing Category	Importance									Comparison Category
	Absolutely Important (9)	Very Important (7)	Important (5)	Slightly Important (3)	Neither Nor (1)	Slightly Important (1/3)	Important (1/5)	Very Important (1/7)	Absolutely Important (1/9)	
Procurement of local funds			V							Management of local manpower

Analogously with the survey results, consistency was evaluated such that it appears effective and consistent. Consistency evaluation was obtained though the Consistency Rate. The consistency rate is a value divided by a Random Index acquired experimentally from the Consistency Index. In this study, resulting from consistency evaluation determining whether each survey result and value is effective and consistent, the consistency rate meaningfully appeared under 0.1 as 0.0851 (refer to Table 3). Accordingly, the importance value resulting through this survey can be considered meaningful. The weight of priorities by risk calculated applying AHP through such examination are as shown below in Table 3.

Analysis results of priorities by risk were examined in the order of ‘procurement of local funds’, ‘local manpower management’ and ‘intensified competition with local companies’, and management risks were determined high priority.

Table 3. Computation of the Priority of Each Risk Applying AHP

	1	2	3	4	5	6	7	8	9	10	11	Priority Vector	
1	0.49	0.43	0.52	0.27	0.15	0.58	0.48	0.29	0.22	0.18	0.29	0.36	3 rd
2	0.50	0.52	0.49	0.52	0.22	0.58	0.52	0.19	0.10	0.27	0.49	0.40	1 st
3	0.19	0.16	0.25	0.04	0.11	0.48	0.41	0.28	0.13	0.07	0.55	0.24	8 th
4	0.46	0.19	0.38	0.56	0.30	0.51	0.42	0.33	0.40	0.27	0.44	0.39	2 nd
5	0.09	0.09	0.10	0.51	0.08	0.53	0.20	0.07	0.09	0.07	0.53	0.21	10 th
6	0.09	0.73	0.10	0.30	0.07	0.37	0.26	0.37	0.18	0.07	0.52	0.28	4 th
7	0.18	0.08	0.09	0.26	0.07	0.42	0.32	0.31	0.47	0.09	0.48	0.25	6 th
8	0.09	0.10	0.08	0.18	0.07	0.22	0.25	0.21	0.13	0.09	0.28	0.15	11 th
9	0.28	0.45	0.07	0.19	0.07	0.34	0.69	0.26	0.10	0.09	0.17	0.25	7 th
10	0.11	0.42	0.47	0.23	0.06	0.37	0.22	0.27	0.23	0.09	0.42	0.26	5 th
11	0.07	0.06	0.06	0.16	0.07	0.18	0.10	0.06	0.60	0.09	0.32	0.16	9 th
1. Management of local manpower 2. Procurement of local funds 3. Weakened profitability ensuing rise in local production costs (wages, fuel tax increase, etc.) 4. Lack of local info 5. Intensified regulations on foreign capital companies (regulations on equipment investment slash in tax benefits, etc.) 6. Intensified competition with local companies (risks from excessive competition) 7. Environmental risks (insufficient logistics infra in the inlands, etc.) 8. Business risks from poor business environments (difficulty in setting up a local corporate body, etc.) 9. Political risks (shift in government, etc.) 10. National risks (anxiety from national affairs such as dispute cases, etc.) 11. Security risks (vulnerable marine transportation or ship stability, etc.)													
CI=0.1055, CR=0.0851													

3.3 Implications & Risk Management Measures

3.3.1 Risks of Procuring Funds Locally

The biggest problem for logistics companies in advancing overseas is the risk of producing funds ensuing overseas investment. In order for logistics companies to build an international logistics network, government assistance is necessary to procure such funds. Accordingly, it is necessary to facilitate the procurement of funds through bank loans or creation of logistics funds, apart from the method of procurement directly by the logistics company. With assured profitability of overseas investment,

the method of procuring funds through commercial banks through government-run bank support or guarantee of an export insurance corporation may be considered.

3.3.2 Risks of Insufficient Local Info

In order for domestic logistics companies to enter the overseas logistics market successfully, the capacity to acquire sufficient information beforehand is important. Insufficient information may be a stumbling block for successful operation or localization of domestic logistics companies. As such, government and support facilities are necessary to collect, analyze and provide the information necessary for logistics companies. Accordingly, there is a need to prepare a system that provides accurate information from government or support facilities for domestic logistics companies to penetrate the overseas market successfully.

3.3.3 Risks of Local Labor Management

To manage the risks associated with manpower, logistics experts with international work experience and logistics expertise management need to be cultivated and long-term products must be set up to cultivate global logistics professionals. Accordingly, a project to cultivate global logistics professionals must be established to become a power country when it comes to global logistics in the long-term run.

3.3.4 Intensified Competition with Local Companies

In order for logistics companies to succeed in advancing overseas, acquiring the quantity of goods transported must be the priority. However, one of the difficulties in advancing overseas is acquiring sufficient quantity of goods transported amidst competition with local companies. According to interviews with small- and medium-sized logistics companies that have penetrated Central and South America, a difficulty in operating business is that it is difficult for small- and medium-sized logistics companies with hardly any recognition to obtain orders for shipments from shippers as the brand image of logistics companies is considered often times when entrusting goods. As such, it is necessary to boost the brand image of domestic logistics companies to overcome such difficulties.

4 Analysis of the Overseas Advancement Strategies of Logistics Companies

Overseas advancement by domestic logistics companies are largely in the form of 1) advancement accompanying shipping companies, 2) M&A of overseas logistics companies, 3) joint venture with local companies, etc. (Kim [9]). Advancement accompanying shipping companies has involved natural advancement overseas along with the progression of globalization while M&A of overseas logistics companies was driven centering on large scale domestic logistics companies. Meanwhile, joint ventures with local companies were in the form of small- and medium-sized logistics companies lacking capital power investing jointly with local companies.

In this study, the risks of overseas advancement derived through the findings of this research were compared and analyzed surveying those involved in logistics companies and logistics experts to determine which overseas advancement strategies are meaningful by region (N. America, Europe & Asia) and the overseas advancement strategy for each risk by region. The analysis results are as follows.

Table 4. Comparison of the Overseas Advancement Strategies of Logistics Companies

Region	Overseas Advancement Strategy	Score Increase	Ranking
N. America	Co-advancement	1.326	1
	M&A	0.815	2
	Joint Venture	0.809	3
Europe	Co-advancement	0.937	2
	M&A	1.207	1
	Joint Venture	0.807	3
Asia	Co-advancement	0.818	2
	M&A	0.811	3
	Joint Venture	1.321	1

Co-advancement is evaluated the best strategy among strategies for logistics companies in advancing overseas in the N. American region. A representative example of co-advancement in the N. American region is the case of Glovis penetrating the US market. Glovis co-advanced into the N. American region accompanying Hyundai Motors. In the European region, M&A was evaluated the best strategy among strategies for logistics companies in advancing overseas. In the European region, relatively many cases of M&A of logistics companies were surveyed compared to other regions. A representative M&A with a local logistics company is the case of global advancement of large scale global logistics companies DP DHL with DB Schenker. Lastly, joint venture with local logistics companies was evaluated the best strategy among strategies for logistics companies to advance into the Asian region. In the case of establishing a joint venture company with local logistics companies, there have been numerous cases of logistics companies strategizing to make inroads into emerging nations recently, the case of local joint venture company established by Hanjin with the largest logistics company (local land transport company Central Asia Trans) in February 2013 being representative. Through joint venture, Hanjin built a complex logistics system linking Central Asia as well as Russia, Europe and the Middle East regions by truck-rail-air.

5 Conclusions

This study derived the risk factors that may arise when logistics companies advance overseas ensuing globalization and calculated the importance such risk factors. A total of 11 risk factors in 2 categories were deduced by reviewing literature and through advanced research analysis, expert opinion and market research, and the importance of the 11 factors deduced was calculated applying AHP by surveying those involved in the logistics business with experience advancing overseas. According to survey results, risks involving 'procurement of local funds', 'insufficient local information' and 'management of local labor force' were deduced relatively important. As well, measures to manage risks of high importance were provided.

Based on the deduced risk factors, representative strategies for logistics companies in advancing overseas, namely co-advancement, M&A and joint venture were compared regionally. According to analysis results, co-advancement strategy in N. America, M&A strategy in the European region and joint venture strategy in the Asian region were determined important.

Apart from such risks, the lack of an organization currently in charge of managing the risks was pointed out as a problem according to interviews with representative domestic logistics companies. However, re-recognizing the importance of managing risks, the necessity of forming an organization in charge of the risks has also been recognized internally by companies. Rather than settling simply for managing risks by forming an organization in charge of risks, there is a need for organization steering towards regulating risks.

References

1. Choi, B.H.: Top 6 risks in emerging/developing countries. LG Business Insight (2010)
2. Jun, C.Y., Kim, W.H., Kim, W.S., Kim, G.S., Kim, J.H.: Establishing strategies by country to promote collaboration and investment in the globally emerging logistics market – Based on ASEAN. Korea Maritime Institute (2009)
3. Choi, B.H.: Need to strengthen risk management, business in China. LG Weekly Economy (2007)
4. Cox Jr., L.A.: Risk analysis foundations, Models and methods. Kluwer's International Series (2002)
5. Molak, V.: Fundamentals of risk analysis and risk management. Lewis Publishers (1997)
6. Park, L.J.: Risks in China from deficit trade balance of China. LG Weekly Economy (2004)
7. Samil Pricewaterhouse Coopers: Total risk management – Consolidated framework. English Language Teaching Database (2004)
8. Son, S.G.: Research on management of risks of investing in China by Korean companies. Sungkyunkwan University (2010)
9. Kim, S.W.: Advancement into the overseas market by domestic logistics companies. Ocean Logistics Research (2010)

Development of an Operating System for Optimization of the Container Terminal by Using the Tandem-Lift Quay Crane

Sang-Hei Choi¹, Hyeonu Im², and Chulung Lee³

¹ Korea Maritime Institute, Logistics Technology Research Department,
1652, Sangam-dong, Mapo-gu, Seoul, Republic of Korea
shchoi@kmi.re.kr

² Korea University, Department of Industrial Management Engineering,
Anam-dong 5(o)-ga, Seongbuk-gu, Seoul, Republic of Korea
gusdndla@korea.ac.kr

³ Korea University, Management of Technology,
Anam-dong 5(o)-ga, Seongbuk-gu, Seoul, Republic of Korea
leecu@korea.ac.kr

Abstract. Since the beginning of containerizing the cargo in the mid-20th century, the trade volume between the countries is steadily rising. Ports with efficient facilities today are competing for mega ships. Advanced ports actively invest in the quay crane in order to maximize the efficiency of the port operating system. And the spreader of tandem-lift type, having good performance, has been operating in some large ports, or during a test. In this paper, we simulate conditions of Busan New Port Container Terminal using the tandem-lift quay crane. Thus, we optimize an operating system by analyzing the efficient ways of port operations.

Keywords: Tandem-lift quay crane, Alternative vehicle, Queue model, Simulation.




1 Introduction

Since the beginning of containerizing the cargo in the mid-20th century, the trade volume between the countries is steadily rising. Shippers were trying to reduce logistic costs because of securing their competitiveness in the market. One of these efforts is the emergence of mega ships. This is a result of the increasing flow of container traffic and the development of shipbuilding technology. Hence, ports with efficient facilities today compete for mega ships.

Advanced ports actively invest in the loading and unloading equipment affecting productivity in order to maximize the efficiency of the port operating system. It is the quay crane to show the biggest change of handling equipment in container terminal. The spreader of tandem-lift type, having good performance, has been operating in some large ports, or during a test.

In this paper, we simulate conditions of Busan New Port Container Terminal using the tandem-lift quay crane. And we estimate the waiting time between quay area and yard area according to the tandem ratio and the type of vehicles. Thus, we optimize an operating system by analyzing the efficient ways of port operations.

Table 1. Spreader types of quay crane (Source: Choi et al. 2013)

Spreader type	Single-lift	Twin-lift	Tandem-lift
			
Capacity	One 20ft container	Two 20ft containers One 40ft container	Four 20ft containers Two 40ft containers

2 Literature Review

In many studies, vehicle dispatching rules were used to simulate the container terminal. deKoster, Le-Anh, and van der Meer (2004) evaluated the performance of real-time vehicle dispatching rules by using simulation models. And researchers started to consider the tandem-lift quay crane according to the emergence of mega ships. Zhu, Fan, Cheng, and He (2010) proposed a new container terminal handling technology which is comprised of tandem-lift quay crane. Furthermore, Xing, Yin, Quadrifoglio, and Wang (2011) studied the problem of dispatching AGVs in the container terminal equipped with tandem-lift quay cranes. Meanwhile, Choi, Won, and Lee (2013) compared the performances of three alternative configurations of ship-to-yard vehicles in a conventional container terminal environment.

Erstwhile researches were mainly limited to partial optimization of the container terminal. What we need to know, however, is complete optimization of the container terminal. Therefore, this paper optimizes whole areas of the container terminal including quay area, travel area, and yard area.

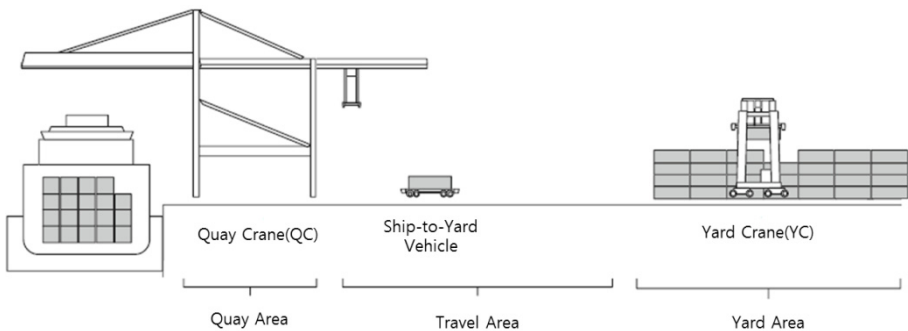


Fig. 1. A schematic view of the container terminal (Source: Bae et al. 2011)

3 Problem Description

3.1 Tandem-Lift Quay Crane

Busan New Port Container Terminal, PNC can accommodate 15,000TEU class and averages 35 moves per hour per crane. If PNC cannot increase the productivity of the crane when larger vessels are berthing, crane facilities should be expanded. Otherwise, PNC will not be able to meet the needs of ship owners. Thus, the solution is to use the tandem-lift quay crane.

The tandem-lift quay crane theoretically increases a double productivity compared to the twin-lift quay crane because four 20ft containers or two 40ft containers can be processed simultaneously at a time. Therefore, PNC has to consider that loading and unloading containers were equally planned. This paper separates the tandem ratio from 10% to 70% by an interval of 10%.

3.2 Ship-to-Yard Vehicle

The tandem-lift quay crane needs simultaneous operations of two single-stack trailers (SSTs) because four 20ft containers or two 40ft containers can be processed simultaneously at a time. Thus, the solution is to use double-stack trailers (DSTs) or serial dual-trailers (SDTs). This paper discriminates DST from SDT which is shown in fig. 2.



Fig. 2. Alternatives of ship-to-yard vehicles (Source: Choi et al. 2013)

4 Simulation Model

4.1 Quay Area

Ships are assigned to berth which takes the shortest working time. And vehicles are assigned to quay crane which has the most heavily workload within the assigned berth. In order to use the tandem-lift quay crane, containers have the same official number of ship and the same type of operation, and the total weight of containers must not exceed the maximum load of spreader.

Table 2. Input data in simulation

Port type	Warm-up time	Simulation end time	Total TEU	Inbound ratio
horizontal	7:00:00:00	30:00:00:00	575,491	0.29
# of berths	QC type	QC speed	Spreader type	Outbound ratio
4	single	0.75m/s	tandem	0.31
# of yards	YC type	YC speed	Filling rate	Transshipment ratio
28	twin	2.5m/s	0.8	0.4
# of vehicles	Vehicle type	Vehicle speed	Traffic factor	SDT length gain
128	SDT/DST	5m/s	0.7	15.8m

4.2 Travel Area

Vehicles entering the first have priority at every intersection. And there is no traffic congestion between external and internal vehicles. SDTs occupy the adjacent lane when they rotate through 90 degrees because of a large turning radius compared to other vehicles.

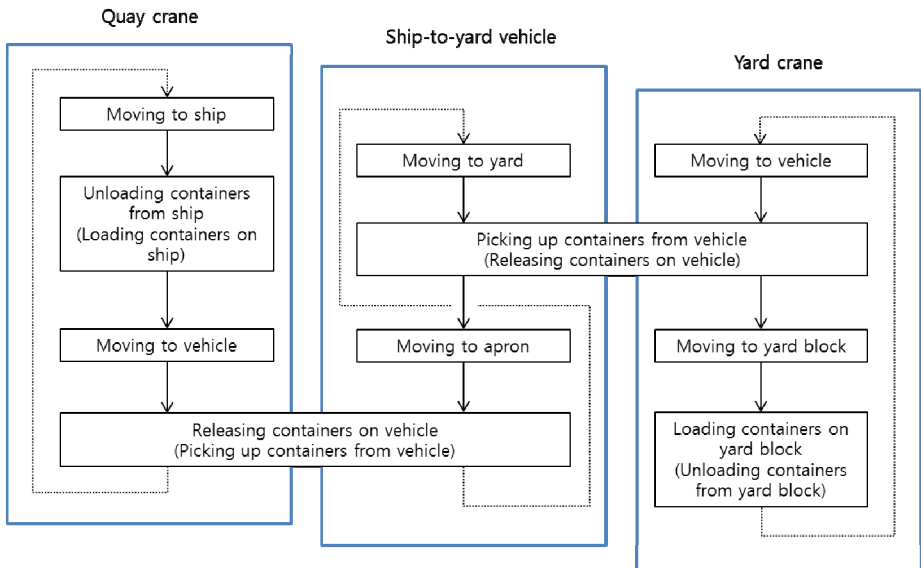


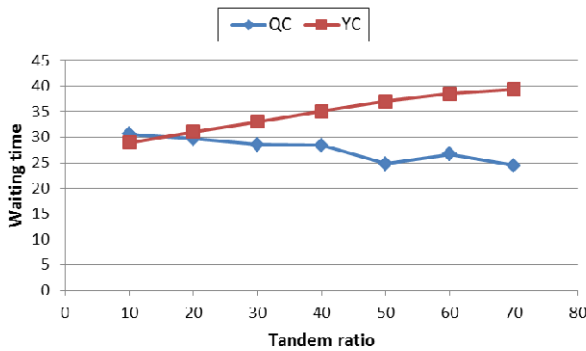
Fig. 3. Flowchart of the container terminal

5 Results and Conclusions

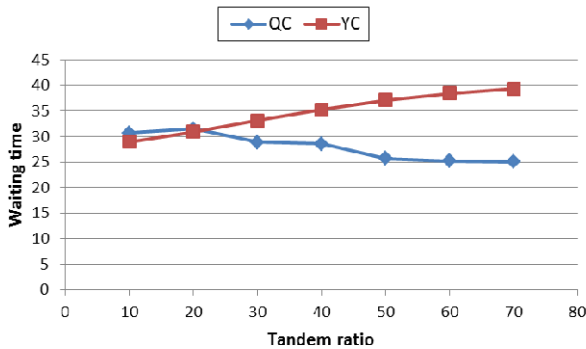
This paper discussed an operating system for optimization of the container terminal by using the tandem-lift quay crane. To analyze the efficient ways of port operations, we separated the tandem ratio from 10% to 70% by an interval of 10% and discriminated two types of ship-to-yard vehicles (DST, SDT).

The more the tandem ratio increases, the more waiting time of quay crane decreases and waiting time of yard crane increases. Because the probability of satisfying the conditions which have the same official number of ship, the same type of operation, and the total weight of containers not to exceed the maximum load of spreader is larger according to the increase of the tandem ratio. On the other hand, waiting time of yard crane is increased due to the increase of activities of quay cranes.

The result of comparing average waiting time of DSTs and SDTs shows that DSTs are efficient when the tandem ratio is less than 31.75, and SDTs are efficient when the tandem ratio is more than 31.75. If the more detailed parameters are added to simulation conditions of Busan New Port Container Terminal, the result will be improved a lot.



(a) Waiting time of QC and YC using DSTs



(b) Waiting time of QC and YC using SDTs

Fig. 4. Waiting time of QC and YC using DSTs and SDTs

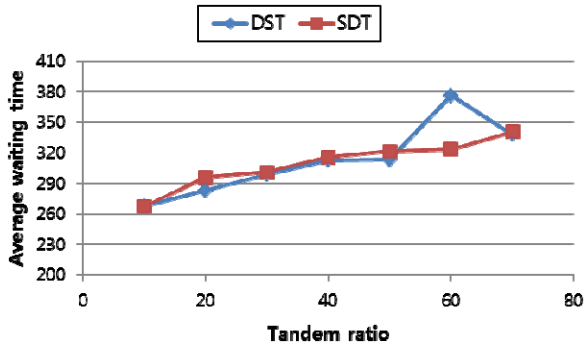


Fig. 5. Average waiting time of DSTs and SDTs

References

1. Bae, H.Y., Choe, R., Park, T., Ryu, K.R.: Comparison of Operations of AGVs and ALVs in an Automated Container Terminal. *Journal of Intelligent Manufacturing* 22(3), 413–426 (2011)
2. Bischak, D.P., Stevens Jr., K.B.: An Evaluation of the Tandem Configuration Automated Guided Vehicle System. *Production Planning & Control* 6(5), 438–444 (1995)
3. Choi, S.H., Won, S.H., Lee, C.: Comparison of Alternative Ship-to-yard Vehicles with the Consideration of the Batch Process of Quay Cranes. *International Journal of Industrial Engineering: Theory, Applications and Practice* 20(1-2) (2013)
4. Choi, Y.S., Yang, C.H., Choi, S.H., Won, S.H.: Analysis of Conceptual Model and Application Effects for High Efficiency Container Crane. *Ocean Policy Research* 22(2) (2007)
5. de Koster, M.B.M., Le-Anh, T., van der Meer, J.R.: Testing and Classifying Vehicle Dispatching Rules in Three Real-world Settings. *Journal of Operations Management* 22(4), 369–386 (2004)
6. Jordan, M.A.: Quay Crane Productivity. In: *Terminal Operation Conference Americas* (2002)
7. Stahlbock, R., Voß, S.: Operations Research at Container Terminals: a Literature Update. *OR Spectrum* 30(1), 1–52 (2008)
8. Xing, Y., Yin, K., Quadrioglio, L., Wang, B.: AGV Dispatching Problem Combined with Tandem Lift Quay Crane in Container Terminal (2011)
9. Zhu, M., Fan, X., Cheng, H., He, Q.: Modeling and Simulation of Automated Container Terminal Operation. *Journal of Computers* 5(6) (2010)
10. Pusan Newport Co., Ltd., <http://www.pncport.com/eng/>

Knowledge Sharing and the Forms of R&D Collaboration

Sang-Ho Kook

Graduate School of Management of Technology, Korea University,
Seongbuk-gu, Seoul 136-701, Korea
okmrkook@korea.ac.kr

Abstract. In the technological driven industry, the firm often tends to implement a R&D project with partners to share the cost and the risk, and it wants to make the driving force for the sustainable competitive advantage with a R&D project. In terms of knowledge sharing, a crucial key of these strategies is how to attain the technological knowledge from the partners and how to find the optimal ratio between its own R&D knowledge and the external acquired knowledge in the accumulated knowledge function, that's because the efficient way reducing marginal cost of product as well as coping with the rapidly changing business environment. Therefor, this article examines the factors of accumulated knowledge function and how the forms of R&D collaboration, inter- and intra firm's interactions affect accumulated knowledge.

Keywords: R&D collaboration, Knowledge sharing, Accumulated knowledge, Spillover, Vertical collaboration, Horizontal collaboration.

1 Introduction

In the technological driven industry, continuous R&D investments and efforts have been implemented as a source of innovation generating competitive advantages. On top of that, A firm has been trying to share marginal cost and risk with R&D collaboration¹ rather than individual R&D due to complicate of technology, uncertainty of market, transaction cost, appropriability regime, limitation of resource, etc. [6, 15, 24]. The results of many literatures, which are interactions through R&D collaboration raise accumulated knowledge and then affect the firm positive effects, are consistent to strategies of firm which prefers R&D collaboration [13, 17, 25]. Especially, since the size of firm and the level of technology apparently affect spillovers effects that play an important role of accumulated knowledge [1, 14], it is necessary to promote R&D collaboration of SME in Korea.

In spite of spillover effects, most of the firms as well as a lot of studies imply that the main purpose of R&D collaboration is usually transaction cost reduction or risk sharing in terms of economic theoretical [16, 18]. However, these theories and cases have an limitation to describe the phenomenon that is 'increasing returns of scales' in

¹ In this paper, a variety of forms (e.g. collaboration, cooperation, etc.) in order to research and develop R&D project together refer to collaboration, interchangeably.

knowledge-based industry [2, 14]. In addition, motives of R&D collaboration have been vigorously analyzed by means of resource-based theory, which is a firm should make use of the strategic R&D collaboration, when characteristics of differentiated resource that a firm possesses are complemented by those of partners [5, 10].

On the other hand, in terms of knowledge sharing, most scholars such like M. Spence and M. Sakakibara explain the motives of R&D collaboration is to create a new technological opportunity by acquiring and assimilating complementary knowledge of partners on condition that a firm’s own R&D investment improves its learning capability [5, 9, 20]. In addition, a firm can leverage its R&D expenditures to achieve a greater understanding of its technology than it’s able to have developed by counting on its own R&D capabilities[23].

Table 1. Comparison of R&D collaboration’s motives: Transaction cost, Resource-based, and Knowledge-sharing

	Transaction cost	Resource-based	Knowledge-sharing
Main purpose	Minimizing the sum of production and transaction costs for the new product (or process)	Maximizing firm’s value through acquiring and assimilating partner’s valuable resources	Creating a technological opportunity by maximizing spillover and complementary knowledge
Critical factors	Production cost, transaction cost	Resource types (i.e. property-based, knowledge-based)	Learning capacity, spillover, R&D investment

Table 2. The forms of Collaboration

Vertical collaboration	Horizontal collaboration	No collaboration
Suppliers or customers in the supply chain	Competitors, university, research center, etc.	Individual firms related with no partners

Furthermore, because spillover effects may have considerable differences among the forms of collaboration (No-collaboration (N.C.), Vertical Collaboration (V.C.), Horizontal Collaboration (H.C.)) [8, 25], we shouldn’t disregard a analysis for the motives of R&D collaboration. For example, J. Peters [8] indicated that spillover in V.C. make a difference for the technological opportunity and G. Atallah [3] found that R&D investment and welfare are more increased when the firm chooses a strategy that is V.C. rather than H.C.

2 Knowledge-Sharing Model Structure in Collaboration

2.1 M. Spence's Model

Following M. Spence [22], We assume that innovation performance by R&D collaboration refers to cost-reducing R&D and marginal production cost of firm i^2 is given as

$$\begin{aligned}
 c^i &= c^i(z^i) & (1) \\
 z^i &= z^i(M^i, M^j, \theta^j) \\
 &= M^i + \theta^j \cdot M^j & (2)
 \end{aligned}$$

where a function c^i is decreased with the amount of z^i which is a function of the accumulated knowledge, M^i and M^j are individual R&D investment for the each firm i, j and θ^j refers to spillover rate of firm i . It is assumed $0 \leq \theta^j \leq 1$, so the contributions of firm j 's R&D don't exceed M^i .

2.2 M. Sakakibara's Model

And then, M. Sakakibara [21] modeled that accumulation of knowledge can be also changed by absorptive capacity, when learning capacity by the firm own R&D investment is improved [9].

$$z^i = M^i + \gamma^i(M^i, \beta) \cdot \theta^j \cdot M^j \quad (3)$$

Here γ is a fraction of partner's knowledge that the firm can seize and assimilate, and thus refers to the firm's learning capability or absorptive capacity. β represents the degree of complementarity of the knowledge between the two partners. It is assumed $0 \leq \gamma \leq 1$, $0 \leq \beta \leq 1$, so firm i 's R&D investment value, M^i , can't be smaller than that contributed by firm j . In addition firm's learning capability, γ^i , is responded to M^i positively and the degree of complementary and technology distance, β , negatively [12]. That's because the complementarity or distance requires that firm i should learn essential knowledge to understand it easily.

Now, a firm is always trying to look for the optimal values in order to maximize its profits, P . When inverse demand function $D^{-1}(Q)$, where $Q = \sum_{i=1}^i q^i$ is the total quantity produced, $D^{-1} = a - b(q^1 + q^2)$, where $a \gg 0, b > 0$ and $s(\theta)$ is a cost of knowledge sharing.

$$\begin{aligned}
 P^i &= q^i \cdot (D^{-1} - c^i) - M^i - s(\theta^i) \\
 &= q^i \cdot c^i(z^i) \cdot \delta - M^i - s(\theta^i) & (4)
 \end{aligned}$$

Where $\delta = [a - b\{q^i \cdot (c^i(z^i)) + q^j \cdot (c^j(z^j))\}] - c^i(z^i)$

² Firm i and j regards as a receipt firm and donor firm, so complementary knowledge and spillover mainly flow firm j into firm i .

Eq (4) shows that firm's profits are inversely proportional to the square of marginal cost which is relevant to accumulated knowledge [4, 21], so it is absolutely important for a firm to reduce it.

3 Considering the Forms of Collaboration

3.1 The Ratio of Internal R&D Investment to Spillover Part

Both M. Spence and M. Sakakibara's models assume that a firm's own R&D investments value isn't smaller than that of spillover part in Eq (3). However, because the smaller a firm seriously depends on spillover effects [1], although firms are only similar to the amount of accumulated knowledge within industry, a firm's innovation performance changes inevitably as time and market environment (e.g. replacement of supplier or customer, the change of complementarity knowledge, etc.).

In terms of accumulated knowledge, the reason can be explained not the total amount of accumulated knowledge, but as a ratio (δ^i) of internal R&D investment to spillover part. It is also assumed that $0 \leq \delta^i \leq 1$.

$$\delta^i = \frac{\gamma^{i(M^i, \beta)} \cdot \theta^j \cdot M^j}{M^i} \quad (5)$$

Although increasing δ^i improves the R&D economic efficiency of firm i under the condition of same firm's own R&D investment, eventually innovation performance is sensitively influenced by the factors of numerator (i.e. spillover part) in the smaller firm.

As mentioned earlier, many literatures already demonstrated that V.C. has performance better than H.C. as well as the difference of performance occurs among firms that take the same form of collaboration. K. Kesteloot suggested that spillover is determined not only involuntary and uncontrollable (at a rate θ_{un}^j), but also, at least partially, intended (at a rate θ_{in}^j). It is assumed $\theta^j = \theta_{un}^j + \theta_{in}^j$, and then if $\theta_{un}^j + \theta_{in}^j = 1$, the spillover is perfect [19].

So we'll separate the spillover rate, θ^j , to consider these effects which are the forms of collaboration including inter-firm's interaction and intra-firm's activities in the spillover part, Eq (5).

3.2 The Propensities for the Forms of Collaboration

As discussed earlier, knowledge sharing directly related to R&D innovation performance plays an important role as a source of firm's sustainable competitive advantage. However, a lot of scholars regarded spillover as a main factors of specific knowledge capital and the state of aggregate knowledge [13] and paid no attention that the forms of collaboration, inter- or intra- firm's interaction may affect spillover.

If the form of collaboration between firm i and j , which firm j possesses complementary knowledge and wants to share know-how within industry, is changed against firm i on the supply chain, spillover rate, θ_{in}^j , may have different value as an each

form, and consequently results in different δ^i and z^i . For instance, when the small firm takes the V.C., innovation performance is superior to that of H.C. for the R&D project [3, 25]. That is, this result means that δ^i is relatively higher in the V.C. and suggests that θ_{in}^j should be considered.

If so, does the firm always have competitive advantage when δ^i is high? Suppose the firm i is a subcontractor³ with little internal R&D capacity and spillover rate, θ_{in}^j , by supplier or customer is extremely high, innovation performance of the firm i seems to be excellent. However, in fact, it doesn't promise not only to have sustainable competitive advantage, but also to be a global innovative R&D firm.

Therefore, θ_{in}^j should reflect mechanism as the characteristic for the forms of collaboration.

3.3 Degree of Integration

Because degree of integration of inter-firm is significantly responsible for making a decision, knowledge sharing's efficiency, scale of R&D project and performance [11], θ_{in}^j must be to have different value before and after the collaboration. For example, although firm i has kept strong integration with the partners during the project, in case of V.C., it is hard to anticipate the continuous support and sticky integration after project. In fact, the partner (e.g. buyer or supplier) intends to focus on this project to maximize its profits as well as spillover generally not interacts with but flows into the firm i . In the long run, θ_{in}^j is going to decrease gradually.

On the contrary, firm i and j need complementary knowledge of each other consistently, regardless whether the project keeps going or is done. So it's not easy to expect the dramatic increasing of θ_{in}^j , but the stronger integration as time goes and the higher θ_{in}^j .

Therefore, although degree of integration is proportional to θ_{in}^j , growth rate depends on diversely by the forms of R&D collaboration and the project's current state.

3.4 Extend to Innovation Behavior

Increasing accumulated knowledge, z^i , is beneficial to innovation activities of the firm i , but it is different matter to apply it to another innovation activities, in order to keep sustainable competitive advantage. If appropriability regime is tight or the partner can easily replace other partners, firm i has no choice but to be conscious of the partner during collaboration.

In other words, since the partners (e.g. university, research center and complementary firm) in H.C. expect to create new technological opportunities by activeness of innovation behavior, they agree implicitly to firm's active behavior. But the partners are concerned about the fluctuation of profits for their ongoing project due to the separation of firm i 's resource in V.C.

³ No partners want to collaborate with a firm that has little R&D capacity in H.C. in terms of knowledge sharing. Most of the firms prefer to collaborate with the partners with abundant complementary knowledge in order to create new technological opportunity.

Table 3. Interaction between accumulated knowledge and individual variables in V.C. and H.C.

	V.C.	H.C.
<i>Degree of Integration</i>	Positive Effect	Positive Effect
- Total amount of z^i Vs. Times	- High but gradually decreasing	- Low or Medium but gradually constant or increasing
<i>Extend of Innovation behavior</i>	Negative Effect	Positive Effect

4 Concept of Modeling for the Forms Effect

4.1 Define the Valuables

As discussed in the previous section, spillover rate is considered three variables and then apply to the model of M. Sakakibara[21].

$$\begin{aligned} \theta^j &= \theta_{un}^j + \theta_{in}^j \\ &= \theta_{un}^j + \theta_{in}^j(DC^i, DI^{i,j}, EI^i) \end{aligned} \tag{6}$$

DC: Degree of Consistency by the form of collaboration (V.C., H.C., No-collaboration,), $-1 \leq DC \leq 1$

DI: Degree of Integration between partners, $0 \leq DI \leq 1$

EI: Extend to Innovation behavior in the firm i , $0 \leq EI \leq 1$

$$\begin{aligned} z^i &= M^i + \gamma^i(M^i, \beta) \cdot \theta^j \cdot M^j \\ &= M^i + \gamma^i(M^i, \beta) \cdot \{\theta_{un}^j + \theta_{in}^j(DC^i, DI^{i,j}, EI^i)\} \cdot M^j \end{aligned} \tag{7}$$

It is assumed that firm i has no relationship with partners, *DC* is zero and θ_{in}^j is zero too. And when the form of collaboration perfectly is V.C. or H.C, *DC* is -1 or 1.

4.2 Discussion

These variables' values can be obtained through the financial structure, properties of R&D projects related with specific partner.

For example, Degree of integration can be calculated by the rate of sales (or purchase) related with specific partner of the total sales (or purpose), the numbers and scales (e.g. research cost, engineers, development period) of R&D projects related with specific partner. Second, extend to innovation behavior depends on the interaction of between internal and external factors such like number of new product (or process), firm size, market structures, industrial technology level [7]. Third, degree of consistency by the form of collaboration is measured by qualitative distance, in terms of supply chain, among the partners participated in R&D project.

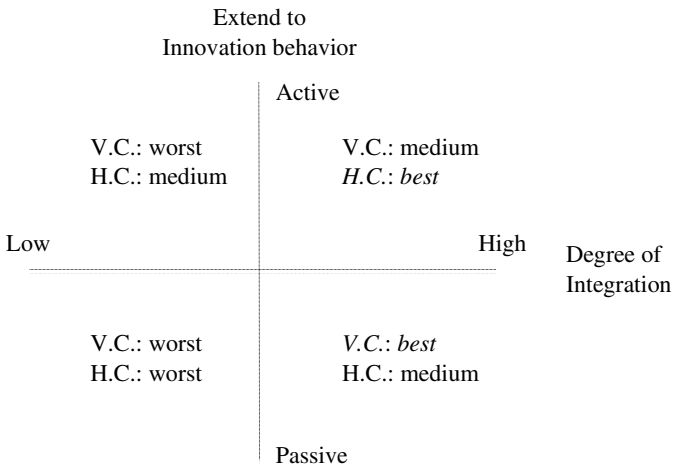


Fig. 1. Contribution to spillover rate for individual variables

5 Conclusion

Global innovative firms have built strategies that are to implement untiring R&D efforts and to acquire accumulated knowledge rather than to choose the way adopting economic profit in short term period. In addition it is the indispensable one of the strategies to keep and increase the accumulated knowledge.

This paper suggests that it is important for the firm to find the optimal rate between its own R&D capacity and the external acquisition of R&D capacity through the knowledge sharing in the accumulated knowledge function. In addition it is proposed that spillover rate, especially intended rate, that is the decisive factor of function depends on the forms of collaboration, degree of integration, and extend to the innovation behavior.

Next, I want to build detail mathematical model and prove it through empirical research in IT industry.

References

1. Acs, Z.J., et al.: R & D Spillovers and Recipient Firm Size. *The Review of Economics and Statistics* 76(2), 336–340 (1994)
2. Arthur, W.B.: *Increasing Returns and the New World of Business*. Harvard Business Review (1996)
3. Atallah, G.: Vertical R&D Spillovers, Cooperation, Market Structure, and Innovation. *Economics of Innovation and New Technology* 11(3), 179–209 (2002)
4. Bandyopadhyay, S., Pathak, P.: Knowledge sharing and cooperation in outsourcing projects — A game theoretic analysis. *Decision Support Systems* 43(2), 349–358 (2007)

5. Barney, J.: Firm Resources and Sustained Competitive Advantage. *Journal of Management* 17(1), 99–120 (1991)
6. Barratt, M.: Understanding the meaning of collaboration in the supply chain (2004)
7. Becker, W., Dietz, J.: R&D cooperation and innovation activities of firms—evidence for the German manufacturing industry. *Research Policy* 33(2), 209–223 (2004)
8. Becker, W., Peters, J.: R&D-Competition Between Vertical Corporate Networks: Market Structure and Strategic R&D-Spillovers. *Economics of Innovation and New Technology* 6(1), 51–72 (1998)
9. Cohen, W.M., Levinthal, D.A.: Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly* 35(1), 128–152 (1990)
10. Das, T.K., Teng, B.-S.: A Resource-Based Theory of Strategic Alliances. *Journal of Management* 26(1), 31–61 (2000)
11. Dodgson, M.: Learning, Trust, and Technological Collaboration. *Human Relations* 46(1), 77–95 (1993)
12. Griliches, Z.: Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics* 10(1), 92–116 (1979)
13. Griliches, Z.: The Search for R&D Spillovers. *The Scandinavian Journal of Economics* 94, S29–S47 (1992)
14. Hagedoorn, J.: Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy* 31(4), 477–492 (2002)
15. Hagedoorn, J.: Understanding the rationale of strategic technology partnering: Nterorganizational modes of cooperation and sectoral differences. *Strategic Management Journal* 14(5), 371–385 (1993)
16. Hennart, J.-F.: A transaction costs theory of equity joint ventures. *Strategic Management Journal* 9(4), 361–374 (1988)
17. Jaffe, A.: Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value 76(5), 984–1001 (1986)
18. Katz, M.L.: An Analysis of Cooperative Research and Development. *The RAND Journal of Economics* 17(4), 527–543 (1986)
19. Kesteloot, K., Veugelers, R.: Stable R&D Cooperation with Spillovers. *Journal of Economics & Management Strategy* 4(4), 651–672 (1995)
20. Sakakibara, M.: Heterogeneity of Firm Capabilities and Cooperative Research and Development: an Empirical Examination of Motives (1997)
21. Sakakibara, M.: Knowledge sharing in cooperative research and development. *Managerial and Decision Economics* 24(2-3), 117–132 (2003)
22. Spence, M.: Cost Reduction, Competition, and Industry Performance. *Econometrica* 52(1), 101–121 (1984)
23. Spencer, J.W.: Firms' knowledge-sharing strategies in the global innovation system: empirical evidence from the flat panel display industry. *Strategic Management Journal* 24(3), 217–233 (2003)
24. Teece, D.J.: Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy* 15(6), 285–305 (1986)
25. Zeng, S.X., et al.: Relationship between cooperation networks and innovation performance of SMEs. *Technovation* 30(3), 181–194 (2010)

Global Supply Chain Management Using Business Risk Re-alignment via the Change of the Transfer Pricing Methodology

Sang Min Ahn¹, Ki-sung Hong¹, and Chulung Lee^{2,*}

¹ Graduate School of Information Management and Security, Korea University, Korea

² School of Industrial Management Engineering and Graduate School of Management of Technology, Korea University, Korea

{hyojun44, justlikewind, leecu}@korea.ac.kr

Abstract. Traditionally, Supply Chain Management (SCM) indicates a strategy enabling an enterprise to achieve optimization of stock level and reduction of lead time through application of information technology. However, such traditional SCM strategy is in contemporary business world required to reflect business economics aspects such as accounting and taxation in order for the enterprise to enjoy maximum benefits.

This paper proposes a framework and methodology for optimal profit allocation to participants of multinational enterprises' global supply chain through re-alignment of certain risks assumed by each participant using financial engineering methods.

Keywords: SCM, Transfer Pricing, Value-at-risk, Earning-at-risk, Business Risk, Market Risk, Foreign Exchange Risk.

1 Introduction

The role of multinational enterprises (MNEs) in world trade has increased dramatically over the last 30 years, reflecting the increased integration of national economies and technological progress. MNEs are focusing on effective global supply chain management in order to obtain a competitive advantage and return a maximized value to shareholders. Undertaking into consideration the recent economic recession, it became imperative for MNEs to reduce costs through an optimized supply chain management strategy.

For MNEs, the tax imposed by the authorities on each part of the supply chain, significantly affects their net profit, which is one of the critical factors to evaluate the companies. The transaction price among the related parties on a supply chain is referred to as the 'Transfer Pricing' and this presents increasingly complex taxation issues for both tax administrations and the MNEs themselves. Since each country's tax regulations on MNEs cannot be viewed in isolation they must be addressed in a

* Corresponding author.

broad international context from the MNE's management's point of view. In the case of MNEs, the differing requirements of each country may result in high costs of relevant expenditures. Also they take the risk of double taxation on the same item of income by more than one tax jurisdiction.

In fact, many MNEs had been likely to reduce the global tax burden through the adjustment of transfer prices by focusing the different effective tax rate by each country. However, the tax authority of each country prevents the outflow of the tax revenue overseas by regulating the transfer pricing adjustment in a strict manner. According to Borkowski [1], appeals against transfer pricing adjustment conducted by the IRS in US federal courts was increased by 40% in 1999. Ernst & Young [2]'s survey results, based on a series of independent interviews of 877 MNEs across 25 countries, shows that 32% of all respondents identify transfer pricing as one of the most important tax challenges facing their group. 74% of parent respondents and 76% of subsidiary respondents believe that transfer pricing will be "absolutely critical" or "very important" to their organizations over the next two years.

In this study, the burden of the business risks (market risk and foreign exchange risk) among the participants of the supply chain was quantified, by the application of expanded financial engineering techniques. Then, through the rearrangement of the data, it was analyzed how that process would increase the value of the company which is one of many purposes of supply chain management.

2 Risk Re-alignment within Global Supply Chain

It is easier and less costly to transfer the risk of a company, rather than its functions or assets, through the contract among the participants of the supply chain or the change of the transfer pricing decision method. PricewaterhouseCoopers [3] classifies the typical business risks as market risk, inventory risk, product liability risk, credit risk and foreign exchange risk.

Among those, the market risk that products sales is not enough to make profits and foreign exchange risk the profitability changes due to the difference of the value of one currency between the billing point and the actual payment point have the most effect on the profit distribution among the participants of the global supply chain. Market risk usually exists when a company enters a new market, the number of competitor increases, and the consumer's need change. It is not possible to develop a new product in the target market. Foreign exchange risk is a typical business risk that can cause a serious disequilibrium of the profit distribution, when the foreign exchange fluctuation is large due to the recent financial market crash.

2.1 An Outline of the Model

To illustrate the model, it is considered that a multinational automotive parts company operates the global supply chain which has a manufacturing subsidiary in Slovakia ("Slovakian manufacturer") and sales subsidiary in US ("US distributor"), to sell the products to the US market. In the typical global supply chain model, one entity

purchases raw material/work-in-process emanating from an upstream entity, and sells products to a downstream entity which may manufacture another work-in-process for another downstream manufacturer. However, in the model, such typical supply chain model is simplified to a two-tier model: the single manufacturer to the single distributor.

Given that the corporate tax rate of Slovakia is 19%, and the US applies 40%, a MNE's group tax manager wants to allocate as much profit to Slovakia as possible. To do so, the tax manager may have three options pursuant to the OECD Transfer Pricing Guidelines [4]: transfer of functions performed, risks assumed or assets used from the US distributor to the Slovakian manufacturer. However, considering the infrastructure of Slovakia and the relevant additional cost, it may not be feasible to transfer the significant assets or functions from the US to Slovakia. If the Slovakian manufacturer raises the selling price (transfer price) to the US distributor without any reasoning, it will be more likely to be subject to the US tax authority's taxable income adjustment.

If, however, the transfer pricing adjustment was based on the reasonable supply chain management strategy that the burden of the actual market/foreign exchange risks were also transferred to Slovakia, the company may maximize the shareholder value by reducing tax, without additional investment since the assumption of increased risk shall be compensated by an increase in the expected return as stipulated by the OECD Transfer Pricing Guidelines.

This model analyzed the MNE group's global tax saving effects on the fiscal year 2007 ended April 2008 by transferring the market and foreign exchange risks from the US distributor to the Slovakian manufacturer.

2.2 Measurement of Market Risk: Earning at Risk (EaR)

Value at Risk (VaR), which was developed as a concept for the risk management of the financial instruments, refers the specified numerical value of the potential risk. In other words, VaR is a statistical method to measure the maximum amount of the loss on a portfolio of assets over a specific time, within a given confidence interval, in case that the market goes adverse. If the 1-year VaR of a company is ten million won at 95% confidence interval, it means that the probability of company's loss of more than ten million won during the next year is equivalent to 5%.

Myatt [5] indicated that earnings-at-risk (EaR), differently from VaR where exposures such as portfolio of securities and transactions is analyzed, analyzes future flows resulting from assets and liabilities, and associated hedging contracts. Thus, in this study, the EaR model was used to measure the market risk of the entire global supply chain, caused by the difference between the expected market sales price and actual one, and the following changes in assets and liabilities. To measure EaR for the market risk, the following calculation method is used in this paper.

$$EaRg = Gap \times \sigma_{daily} \times \alpha \times \sqrt{T} \quad (1)$$

where,

EaR_g : EaR on Gap

Gap : Net working capital (current asset –current Liability)

σ_{daily} : Daily standard deviation of operating income ratio (“OM”, the ratio of operating profit to sales)

α : Significant factor (Confidence level of standard normal distribution)

T : Period

To measure the average market risk of the same industry in US market, using the Osiris Database, the balance sheet and income statements of all (11) listed companies classified in US Industry Classification Code 501 (Motor Vehicles and Motor Vehicle parts and supplies), out of the entire listed company located in the North America region, had been selected.

Based on the collected financial data, the current assets and current liabilities were classified and then, the difference between the two items was calculated. The data used for that covers the three fiscal years(2005~2007), and the averaged Gap of each company for the three years was considered representing the entire US market.

To calculate EaR_g of each company(EaR_g^i), we need to calculate standard deviation of OM; hence, we derived past 3 years standard deviation of the annual OM of individual companies, and converted to a daily standard deviation using following formula:

$$\sigma_{daily} = \frac{\sigma_{annual}}{\sqrt{T}} \quad (2)$$

EaR_g^i s are derived from multiplying Gap by daily standard deviation of OM and then multiplied duration and $Z(1.65)$ value which is corresponding level of confidence.(risk management period: 1 year and 95% confidence level.) EaR_g^m is averaged value of EaR_g^i To quantify market risks on the US automotive parts market, it is necessary to calculate EaR on Sales (EaR_s) with EaR on operating income (EaR_{oi}). It is assumed that future operating income level will be same as current level, and the present value of future operating incomes of company shall be equivalent to the gap of company, which is similar assumption that the stock price of the company is equivalent to the present value of the company’s future dividends.

To convert EaR_g into EaR_{oi} , a market operating discount rate (ρ) derived using Gap and operating profit shall be multiplied. The formula is as follows:

$$\rho = \frac{oi}{Gap} \quad (3)$$

With the Gap data and income statements of the companies, we can obtain values of ρ^m for the period from the fiscal year 2005 to 2007. EaR on operating income can be obtained using the following formula:

$$EaR_{oi} = \rho \times EaR_g \tag{4}$$

In addition, the market averaged EaR on sales can be derived through the following formula:

$$EaR_s^m = EaR_{oi}^m \times \frac{s^m}{oi^m} \tag{5}$$

Based on the above formula, we could derive following averaged market risk of US Auto parts market for the period under analysis.

Table 1. Average market EaR of US Auto parts companies

(Unit: Thousands USD)

Items	Calculations	Items	Calculations
s^m	2,338,358	ρ^m	38.265%
oi^m	168,853	EaR_{oi}^m	960.66
EaR_g^m	2,511	$\frac{s^m}{oi^m}$	9.54
			$EaR_s^m = 9,162.32$

2.3 Measurement of Foreign Exchange Risk: Value at Risk (VaR)

Using the concept of VaR, foreign exchange risk can be measured based on the historical volatility data of the Euro and the Dollar which is the reference currency of the manufacturers and dealers. We selected Historical Simulation method which is one of the VaR measure methods in this study. Lam [6] explained that the historical simulation method allows the relaxation of both the linearity and normality assumptions of other two approaches since it uses historic data about price movements to generate scenarios. This method looks up VaR using market interest rate and historical change of market price which are arranged in order of size to obtain potential profit and loss distribution of the portfolio.

To measure VaR for foreign exchange risk, the following calculation method is used in this paper:

$$Var = \Delta P \times x \times \sigma_{daily} \times \alpha \times \sqrt{T} \tag{6}$$

where

ΔP : The change in the amount of the entity’s portfolio

x : Foreign exchange rate

σ_{daily} : Daily standard deviation for the past 3 years Spot exchange rate fluctuation

In this paper, to calculate the US market foreign exchange risk the US auto parts distributors may assume, average costs of sales of the US distributors reported in their

income statements from the fiscal year 2005 to 2007 after reflecting EaR results as calculated previous session was used as ΔP under the assumption that total purchases by the US distributor rely on the Slovakian manufacturer.

Given that average sales and gross profit of the distributor was 2,338,358 and 496,590 thousand US dollar, respectively, , the amount of averaged foreign currency exposed to the foreign exchange risk is 1,850,930 thousand US dollar based on following formula:

$$\Delta P^m = s^m - gp^m + EaR_s^m \quad (7)$$

where

s^m : market average sales

gp^m : market average gross profit

In this study, foreign exchange spot rate of Euro against US dollar for three years from April 30, 2006 to April 29, 2009 was used to analyze market F/X fluctuation. Based on actual spot exchange rate of daily Euro against US dollar exchange rate, the standard deviation of the spot exchange rate was 0.101779, and it comes to 0.003076 if it is converted to daily exchange rate standard deviation.

To derive VaR, one (1) year to risk management period and 95% confidence level were applied. Total market VaR is derived from multiplying total amount of foreign currency by daily standard deviation of exchange rate, and then multiplied duration and $Z(1.65)$ value which is corresponding level of confidence.

If calculate foreign exchange VaR using this value, that average exposure foreign exchange risk is 9.70% corresponding amount of purchase price(transfer price) when US sales company purchase products from Slovakia manufacturing company.

2.4 Analysis Results from the Transfer of the Risks

The aforementioned effect of the business risks' transfer within the supply chain can be achieved through the application of widely used OECD's Transfer Pricing Methods. The OECD Transfer Pricing guidelines suggest following five (5) transfer pricing Methods: Comparable Uncontrolled Price Method, Resale Price Method, Cost-plus Method, Profit Split Method, and Transactional Net Margin Method

Among above the methods, multinational companies normally choose the cost-plus method in order to firstly reward arm's length remuneration to related manufacturers within their supply chain, which adds the appropriate gross profit (example: the average gross profit level of the comparable manufacturers in the same region) to the manufacturing costs incurred by the manufacturers if the manufacturers perform routine manufacturing function. In this method, the remaining profit (or loss) shall be assumed by the related distributors within the supply chain. The exemplified model in this study has a same structure.

By switching the transfer pricing method from the cost-plus method to the resale price method where the transfer prices are equivalent to market (US) sales prices subtracted by arm's length level of gross profit level, the distributors shall be

rewarded a certain level of profit regardless of market environment and market price fluctuation (i.e., transfer of risks to manufacturer), and foreign exchange risk also can be transferred to manufacturer by changing denomination currency to selling region's currency (US dollar in the model).

For the supply chain, such as the model exemplified in this study, a MNE's global tax manager can theoretically optimize its global tax spending by applying the following method:

$$v = s - \beta + EaR_s + VaR_{f/x} \quad (8)$$

where

v : transfer prices:

β : average gross profit margin of independent sellers in the same market

Recall that the transfer price shall be established at the same condition that independent entities would have agreed. In dealings between two independent entities, compensation usually should reflect the functions that each enterprise performs taking into account assets used and risks assumed (OECD [4]). Having said that, an independent manufacturer may not assume business risks unrelated buyer would have to take if any consideration is not embedded in the prices between two entities. Therefore, the global tax manager shall reflect consideration for the risk transfer in the transfer prices in the model, following the change of the transfer pricing method. In other words, in the case that an unrelated seller takes some risks the buyer should have assumed, the unrelated seller will probably demand an increase of the sales price to the extent of covering potential losses.

In the model, to confirm how much cash saving would have been incurred if the risks were transferred to the manufacturer and the global tax manager decided to reflect effect of it in the transfer prices, following ex-post analysis comparing the model 1 and 2 is conducted:

$$v^{\text{model 1}} = s^m - gp^m = 1,841,768 \text{ thousand US dollar}$$

where,

Averaged sales and gross profit of the distributor were 2,338,358 and 496,590 thousand US dollar

Actual average gross profit value was used in this analysis, instead of β , in order to confirm the effects of the risk transfer

$$\begin{aligned} v^{\text{model 2}} &= s^m - gp^m + EaR_s^m + VaR_{f/x}^m = 3,161,626 \text{ thousand EURO} \\ &= 2,030,393 \text{ thousand US dollar} \end{aligned}$$

where,

Averaged sales and gross profit of the distributor were equal to the model 1

$EaR_s^m = 9,162$ thousand US dollar

$VaR_{f/x}^m = 279,450$ thousand EURO using 1.31875 spot Euro/US dollar spot rate as of April 29, 2009

The result shows that the average transfer prices could be increased by 188,625 thousand US dollar if the US market risk and F/X risk were transferred, and it can be interpreted that the MNE can save 39,611 thousand US dollar in annum considering difference of effective tax rates of two countries (US: 40%, Slovakia: 19%) if the suggested transfer pricing method was used.

3 Conclusion

This paper proposed a model for the determination of transfer prices possibly enabling a MNE to optimize global tax spending by transferring certain business risks. The study was done based on assumptions that the MNEs have manufacturing facilities in low-tax rate jurisdictions such as Slovakia and sales subsidiaries in a major market usually having high-tax rate such as the U.S. It was also assumed that the manufacturing subsidiaries perform routine manufacturing functions without any significant intangible assets. The model analyzed in this paper focused on global savings from the MNE's perspective, not of individual companies.

Economically, the increased risk burden is related to the increased expected return, but not always generates actual return increase. For example, an investor purchasing a junk bond with a face value of 100 US dollars might expect more return than those who have invested in the US Treasury bond with the same face value. However, in reality, the junk bond holder is generally exposed to a big loss.

In the suggested model, if the loss from the long term US market recession is imputed unilaterally to the Slovakian manufacturers, then the overall tax burden on the total supply chain of the MNE could increase. Therefore, this study is only applicable to MNEs with a competent operating activity level and sustainable supply chain. We believe that the suggested method will help the global tax manager to effectively structure the global supply chain of the MNE.

References

1. Borkowski, C.S.: Transfer Pricing Advance Pricing Agreements: Current status by country. *The International Tax Journal* 26(2), 1–16 (2000)
2. Ernst, Young: 2010 Global transfer pricing survey- addressing the challenges of globalization. EYGM Limited (2011)
3. PricewaterhouseCoopers: International Transfer Pricing 2001, PricewaterhouseCoopers: Chapter 1 (2001)
4. OECD: Transfer pricing guidelines for multinational enterprises and tax administrations. OECD (1998)
5. Myatt, J.: Other alternatives to VAR, ASIARISK, pp. 34–35 (February 2003)
6. Lam, J.: Enterprise Risk Management: from Incentives to Controls, p. 181. John Wiley & Sons, Inc., Hoboken (2003)

Research of Touchscreen Terminals Gesture Operation Error Based on Kansei Engineering

Rong Qin¹, Dongxiang Chen², and Xuelong Hou³

¹ College of Mechanical Engineering, Tianjin University, Tianjin, China
Pengxiang Apartment, Tianjin University, Tianjin, China
qinrongwykl@sina.com

² College of Mechanical Engineering, Tianjin University, Tianjin, China
Xinyuancun, Tianjin University, Tianjin, China
dxchen@tju.edu.cn

³ College of Architecture and Art Design, Hebei University of Technology, Tianjin, China
Dongqu Apartment, Hebei University of Technology, Tianjin, China
hx1000836192579@hotmail.com

Abstract. As mobile touchscreen terminal application is more and more diversification and individuation, gestures gradually become the main way of people interact with touch screen interface. But currently, there is a wide variety of mobile touch type terminal on the market, lead to each brand has its own unique style of gestures with no unified design criteria, confusion, operation error and learning difficulties often appear when in use by the user. In this paper, based on the Kansei Engineering related method, choose the optimal gestures form, to reduce the user operation error, and strengthen the availability of touchscreen terminals.

Keywords: gesture operation, mobile touchscreen terminals, wrong operation, Kansei Engineering.

1 Introduction

Currently on the market, touch screen terminal equipment type is various, each brand has its own unique style of operation, some even conflict to each other. This complex chaotic scenes usually let users difficult to identify and memory, cause some confusion and wrong operation inevitable. In the face of numerous and large operation specification, user is hard to remember that different platforms using norms and standards, therefore, learning difficulties and wrong operation appear. Some methods can solve such problems: model-based, accuracy is high, the effect is obvious; appearance-based, fast speed, can satisfy the real-time application requirements; Target tracking method and the rough set calculation method also has strong practicability. The results of these methods effectively improve the gesture operation forms of scientific and identifiable degree. But as a result of hand shape change high dimension, lead to these methods and their derivative algorithm great emphasis on virtual reality,

machine vision, pattern recognition, human-computer interaction, and other fields of exchange and cooperation in computer science, the demand of user's perceptual study is relatively less, the results of the study disconnect with the user's operation habit. In this paper, from the user's emotional instinct of operation, with gesture as the research object, effectively reduce user's operation error and enhance the availability of touchscreen terminals.

2 Gesture-Based Mobile Touchscreen Terminals Process Related Emotional Vocabulary Collection

Perceptual words collection is the foundation and the key of the Kansei engineering research, transform user emotion into understandable adjectives. Eventually representative perceptual vocabulary selection has a crucial influence to the scientificness of evaluation test and the conclusion, So should be to collect as much as possible to express gestures to the user brings emotional words. So should be to collect adjectives as much as possible that can express emotion that gestures operation brings to the user.

2.1 Vocabulary Collection

With mobile touchscreen terminals as object, through brainstorming, product specification and user manual, information that mobile terminal design professionals provide, magazine, Internet, etc. This case has collected 200 emotional adjectives.

2.2 Vocabulary Filtering

(1) Conduct a preliminary classification and selection to the 200 emotional adjectives according to the degree of similar meaning, eliminate meaning particularly close or relevant enough emotional vocabulary, eventually get 40 emotional adjectives, such as:

Concise, Precise, Clear, Harmonious, Generous, Effective, Striking, geometrical, calm, steady, advanced, individual, sportive, lovely, kind, quick, stereo, coherent, complicated, cursory, plain, unbalanced, stingy, ineffective, fuzzy, vulgar, messy, impulse, lively, base, popular, stillness, mechanical, distant, lengthy, plane, intermittent, strong sense of operation, weak sense of operation.

(2) Divide 40 emotional adjectives into 20 corresponding adjective phrases, making questionnaire, Invited 30 design major students participate in the survey, requiring people to select five most can describe gestures perceptual adjective phrases, analysis questionnaire, select the perceptual phrases that choosed most times, correct analysis the results of the survey, finally get 5 groups of representative adjective vocabulary.

5 groups of emotional adjectives are: Concise and Complicated, Clear and Fuzzy, Effective and Ineffective, Quick and Lengthy, Strong sense of operation and Weak sense of operation.

3 Gesture Operate Mobile Touchscreen Terminals Process Related Morphological Analysis

Final selection of representative sign sample have a crucial impact on the science of evaluation test and the conclusion, in this phase should be as much as possible to collect samples of gestures.

3.1 Gestures Form Collection

Through the research of gestures and the sample search from Internet, magazine and product brochure, after deleted some similar gestures samples, work up the pictures as the following form. See from table 1, for the same order, gestures from different operating platform have a variety of forms.

Table 1. Gestures table form














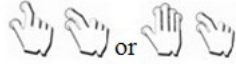
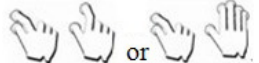

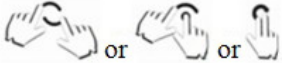
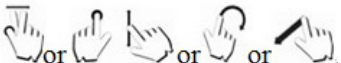

Command names	Gestures form
Click	
Double-click	
Press and tap	
Press	
Flick	
Bundle	
Duplicate	
Drag/Delete/Move/Scroll	
Fast scroll	
Pan/Move	
Move through list/Move	

Table 1. (continued)

Roll	
Press and drag	
Scale down	
Scale up	
Show	
Rotate/Adjust view	
Aim	
Move	

3.2 Shape Analysis

The perceptual cognition of the users need to be quantified. Above summarizes five pairs of adjectives: Concise and Complicated, Clear and Fuzzy, Effective and Ineffective, Quick and Lengthy, Strong sense of operation and Weak sense of operation. The implementation in the form of questionnaire survey, as the site questionnaire and the electronic questionnaire, a total of 30. Participants make 5 attributes evaluation (For example, the most concise, more concise, no obvious bias, more complicated, the most complicated) on the sample, Quantitative standards: -2 represent the most complicated, 1 represent more complicated, 0 represents no obvious bias, 1 represent more concise, 2 represent the most concise. Refer to table 2, the product sample SD perceptual scale.

With table 2 as scoring criteria, users score perceptual factor, for example, narrow command, can be completed by two types of gestures form, Users score two forms respectively, If the form considered the most concise, get 2 points, if the most complicated, get -2 points. Recycling 30 effective questionnaires. Calculated the total score of each item, the result is divided by 30, get a score is the final score for the commands in the form. The results in table 3.

Table 2. The product sample SD perceptual scale

Adjectives	Score					Adjectives
Complicated	-2	-1	0	1	2	Concise
Fuzzy	-2	-1	0	1	2	Clear
Ineffective	-2	-1	0	1	2	Effective
Lengthy	-2	-1	0	1	2	Quick
Weak sense of operation	-2	-1	0	1	2	Strong sense of operation

Table 3. Corresponding relationship survey of emotional vocabulary and design combination



































Emotional appeal		Concise	Quick	Clear	Effective	Strong sense of operation	Total score
Scale down		1.583	1.572	1.675	1.343	0.298	6.471
		0.572	0.772	0.969	1.237	1.263	4.813
Scale up		1.579	1.583	1.581	1.263	0.352	6.358
		0.593	0.754	1.101	1.336	1.173	4.957
Factor and drag		1.657	1.457	0.265	0.128	-0.195	3.312
		-0.527	0.199	1.754	1.652	1.687	4.765
Aim		1.201	0.992	0.283	0.643	-0.301	2.818
		-0.677	1.376	1.233	1.675	1.562	5.169
		0.741	1.201	0.913	1.198	0.243	4.296
		1.757	1.689	1.533	1.766	-0.725	6.020

Table 3. (continued)

Move		1.216	1.579	1.385	1.409	-0.289	5.300
		1.263	0.788	0.247	-0.333	-0.509	1.456
		1.346	1.369	1.480	1.298	0.221	5.714
		0.765	0.669	0.941	0.887	0.561	3.823
		0.965	0.864	0.782	0.774	0.103	3.488
		0.897	0.846	0.482	0.576	-0.462	2.339
		-0.561	1.119	1.599	1.639	1.487	5.283
Roll		1.401	0.566	1.648	1.521	0.321	5.457
		1.380	1.537	0.723	0.396	-0.109	3.927
Show		1.378	-0.207	0.137	1.284	0.560	3.152
		-0.254	-0.991	-0.098	1.498	1.279	1.434
		0.985	1.253	0.076	-0.372	-0.104	1.838
		1.673	1.518	0.882	-0.116	-0.671	3.286
Rotate		0.158	-0.223	0.099	0.477	0.981	1.492
		-0.103	0.378	0.959	1.338	1.587	4.159
		0.529	0.698	0.554	0.356	0.782	2.919

Through table 3 calculate perceptual factor scores, the highest score tertiary perceptual factor become the secondary emotion factor final result. The results in table 4.

Table 4. The highest score perceptual factor

Scale down	Scale up	Press and drag	Aim	Move	Roll	Show	Rotate
							

In Table 4, the form of gestures corresponding operation commands is the result of the highest score, is the gestures form most accord with user perceptual action of instinct, to sum up, research determine the right gesture contained by a set of touchscreen terminals operating system at least by perceptual engineering method. Use mobile touchscreen terminal gestures operation simulation system, input the results as the correct gesture operation form corresponding to the command, a random sample of 100 operators for validation test, the operator operating error rate significantly decreased, prove that the research results are effective.

4 Conclusion

In touchscreen terminals gestures, some commands and gestures in the process of operation form cannot determine corresponding, in this paper, Kansei Engineering was used to determine the gesture operation form contained in a set of touchscreen terminals system at least, effectively reduce the user wrong operation happen because of confusion, the research method in this paper is feasible.

References

1. Zheng, J.: The Small Touch Screen Terminal Interface Usability Research: [master degree theses], hubei province, HuaZhong Science and Technology University (2011)
2. Sun, C., Feng, Z., Li, Y., Zhang, M., Zhang, W., Pan, Z.: Human-computer Interaction Review based on Gesture Recognition (2010)
3. Liu, Y.: The Study to the Relationship between Gesture Operation and Touch Screen Motion Effects under the Rough Sets: [master degree theses]. Tianjin University, Tianjin (2012)
4. Wu, D.: Vision based on Gesture Recognition and Human-computer Interaction Research: [master degree theses]. Nanjing Aeronautics and Astronautics University, Nanjing (2010)
5. Li, Y., Wang, Z., Li, D.: Kansei Theory Research and Product Development of Engineering Application. Wuhan Technology University Journal (January 2010)
6. Li, Y., Wang, Z., Xu, N.: Kansei Engineering. Ocean Press, Beijing (2009)
7. Li, X.: Design Visual Language Interpretation. Art and Design (November 2010)
8. Liu, Q.: Thinking About Some Problems in The Design of Man-machine Interface. Luoyang University Journal (December 2004)
9. Feng, H.: Analyses The Man-machine Interface Design Software. Nanping Teacher College Journal (October 2006)
10. Li, Y., Guan, Z., Chen, Y., Dai, G.: Gesture-based Human-computer Interaction Research. System Simulation Journal (September 2000)

Heart Sound Feature Extraction Based on Wavelet Singular Entropy

Zhang Lu

School of Information Science and Engineering, University of Jinan, Jinan 250022, China
first.ise_zhanglu@ujn.edu.cn

Abstract. After analyzing the advantages and disadvantages of current methods of heart sound feature extraction, a new method based on wavelet transform, singular value decomposition and information entropy is put forward. In this method, firstly the heart sound is decomposed by wavelet transformation. Then the singular value of the sub-bands containing heart sound information is obtained by decomposition. Finally, according to the constructed wavelet singular entropy, the entropy of the above singular value is obtained. By comparing the wavelet singular entropy of normal heart sound signal with the several heart sound signals with pathological information, wavelet singular entropy can be found a good characterization of heart sound.

Keywords: Wavelet Singular Entropy; Heart Sound; SVD.

1 Introduction

Heart sound signal is from heart and collected by cardiotelephone. It is caused by mainly diastole and systole of the heart. It's the flow vibration from the body surface, caused by the impact of blood flow against heart valves, wall and large vessels [1]. There is important physiological and pathological information in heart sound signal, so doctors can determine the patient's disease characteristics by his heart sound signal. Normal heart sound is made up of by s_1 and s_2 that can be heard and s_3 and s_4 that can't be heard. If the heart is abnormal, there are murmurs in addition to s_1 and s_2 . For example, Coronary stenosis can cause murmurs in heart sounds diastolic period. This is because the flow caused by the coronary stenosis make surrounding tissues vibration and then sound is produced [2]. Therefore, to analyze the heart sound signals is an important subject about predictor of cardiovascular disease. Many researchers at home and abroad have tried a variety of methods to analyze the heart sound signals. B.El-Asir et al analyzed the heart sound signals with JTFA, and came to the conclusion that there were murmurs with different frequency at different times in different heart diseases [3]. Gauthier et al analyzed diastole with FFT, and draw the conclusion that diastole percentage in coronary heart disease heart sounds was higher than normal heart sound [4]. Above several studies have achieved some results, but there are some deficiencies in them. For example, in reference 4, because heart sounds is a typical non-stationary

signals, the effect of differentiate coronary heart disease and healthy people by the method of differentiate coronary heart disease and healthy people is not ideal.

In this paper, based on wavelet transform, singular value decomposition and information entropy, the three theories are organically combined together to form a new method of extraction of heart sound signals wavelet singular entropy. In this method, firstly the heart sound is decomposed by wavelet transformation. Then the singular value of the sub-bands containing heart sound information is obtained by decomposition. Finally, according to the constructed wavelet singular entropy, the entropy of the above singular value is obtained. By comparing the wavelet singular entropy of normal heart sound signal with the several heart sound signals with pathological information, wavelet singular entropy can be found a good characterization of heart sound.

2 Wavelet Singular Entropy and Principle

2.1 Singular Value Decomposition

Assume that the matrix $A \in \mathbb{C}_r^{m \times n}$ ($r > 0$), then there are unitary matrix U of order m and unitary matrix V of order n . They have the following equation :

$$U^H A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad (1)$$

In the above equation 1, $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, and $\sigma_i (i = 1, 2, \dots, r)$ are all non - zero singular values of matrix A . Change the equation 1, Singular Value Decomposition (SVD) of matrix A is obtained[5] :

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H \quad (2)$$

2.2 Singular Entropy

The singular value of the signal is to be described the characteristics of each frequency segment of the signal within the sampling time. The main features of the heart sound signals in various lesions signal appears for the difference of the singular values on the different frequency segment. To quantify the extent of this change, singular entropy is structured :

$$E_i = \frac{\sigma_i}{E} \quad (3)$$

In the above equation 3, $E = \sigma_1 + \sigma_2 + \dots + \sigma_r$, so $\sum_{i=1}^r E_i = 1$. This is in line with the initial normalization condition of entropy. According to the definition of entropy, singular entropy is calculated as:

$$H = -\sum_{i=1}^r E_i \ln E_i \quad (4)$$

3 Heart Sound Feature Extraction Based on Wavelet Singular Entropy

There are five steps in the heart sound feature extraction based on wavelet singular entropy: segmentation of heart sound, wavelet transform, singular value decomposition, calculating the singular entropy, characteristic value extracting.

1) Segmentation of heart sound. Heart sound signals used in the experiment are selected from Texas heart institute. Because the heart sounds are too long, they are envelope extracted. By this process, a complete heart sound signal with not only s1&s2, but also s3&s4. Envelope extraction method is HHT, and the obtained envelope is segmented. In Fig.1, the extracted normal heart sound is segmented, and a complete heart sound signal with not only diastole but also systole is obtained.

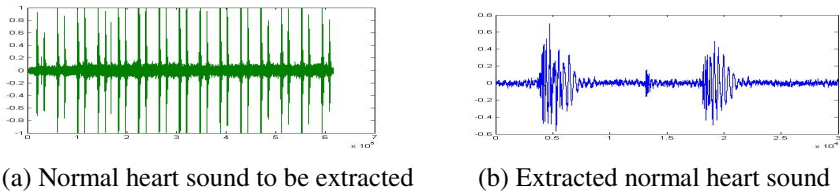


Fig. 1. Envelope extraction of the normal heart sound

2) Wavelet transformation. In this study, the DB6 heart sound signals are used as the mother wavelet in wavelet transformation. Wavelet decomposition scale and sampling frequency relate to the mother wavelet. According to the sampling theorem of wavelet, the more the decomposition scale, the greater the amount of computation is. So the decomposition scale should be selected according to the actual application and needs. The frequency of the heart sound signal used in this paper is 44100 HZ. Analyzing of heart sound signals by STFT, it's found that the main frequency of s1 is 50 ~150 HZ, and s2 is 250 ~300 HZ. So the decomposition scale is eight. The frequency of eighth order of decomposition detail signal cd8 is 172 ~344 HZ, and the frequency of eighth order of decomposition contour signal ca8 is 0~172 HZ.

3) Singular value decomposition. Because of the frequency of the heart sound signal is mainly in 0 to 300Hz, not only cd8 but also ca8 are decomposed by SVD. Then singular value matrix S_d and S_a are obtained.

4) Calculate the singular entropy. S_d and S_a are respectively substituted into the equation 3&4, the singular entropy H_a of eighth order of decomposition contour signal ca8 and the singular entropy H_d of eighth order of decomposition detail signal cd8 are obtained.

4 Analysis of Simulation Results

The simulation software is Matlab7.4. There are five simulation signals: a, Normal ; b, WSSS ; c, HOC ; d, VSD ; e, LAM.

The normal heart sound signal is contrasted with VSD. The Fig.2, Fig.3, Fig.4 respectively show comparison of extracted signal, eighth order of decomposition contour signal ca8 ,eighth order of decomposition detail signal cd8.

Fig.2a is a typical heart sound signal, the front half portion of which is s1 and the latter half portion s2. May also it is mean that the front half portion is systolic and the latter half phase. In Fig.2b, there is a clear difference between the systolic and diastolic. Predictably, there are surely murmurs in Fig.2b, and it shows the signal certainly with lesion signal .It can be judged Fig.2b is signal with lesions signal.

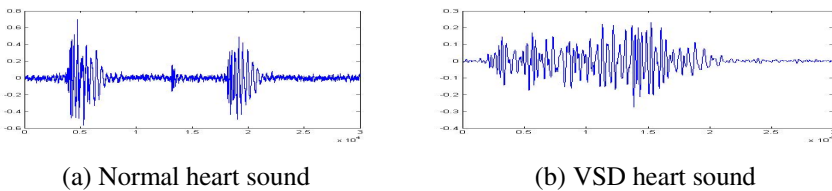


Fig. 2. Contrast between Normal and VSD heart sound

Then the Fig.3 is analyzed. The Fig.3 shows their contour signal, that is to say, it shows the difference between 0HZ to 172HZ.In the Fig.3a, two very clear contour signals are found, however in the Fig.3b, the composition of the signal can not be distinguished. So the difference between the normal heart sound and the VSD heart sound is very large.

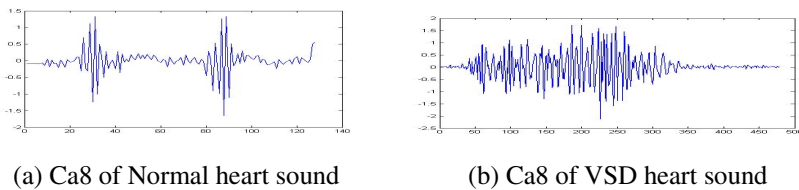


Fig. 3. Contrast between ca8 of Normal and VSD

Finally, the three detail signals are compared. That is to say, it shows the difference between 172HZ to 344HZ.Two signals with large amplitude and one signal with small amplitude are found in the Fig. 4a.Two signals with large amplitude are too found in the Fig. 4b,but signal with small amplitude can not be found. So there is difference between them, but the difference not large. According to reference 1, there is murmur in the whole systolic of VSD and the proportion of metaphase is the largest.

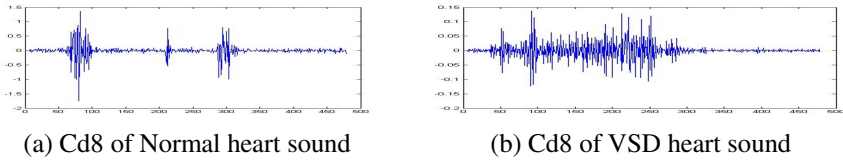


Fig. 4. Contrast between cd8 of Normal and VSD

The singular entropy H_a of eighth order of decomposition contour signal ca8 and the singular entropy H_d of eighth order of decomposition detail signal cd8 of the five heart sound are shown in the Table 1. Here the normal heart sound and the VSD heart sound are set an example. The H_a of the normal is 4.3374, while the H_a of VSD is 7.4331. The difference between the two is 3.0957. The H_d of the normal is 3.271, while the H_d of VSD is 4.305. The difference between the two is 1.034. So it's shown that the lesion signals of VSD is mainly between 0 to 172 HZ. In this band, there are mainly the first heart sound s1 and part of the second heart sound s2. The analysis result of the figure and table are unanimous. It is shown that wavelet singular entropy is a good description of the characteristics of heart sound.

Table 1. Wavelet singular entropy of heart sound

	H_a	H_d
Normal	4.3374	3.271
WSSS	9.8094	10.242
HOC	10.662	5.1194
VSD	7.4331	4.305
LAM	5.4886	9.7397

5 Conclusion

In this paper, wavelet transform and singular entropy decomposition are used to obtain the feature of heart sound. Through theoretical analysis and simulation test, this method can well express the heart sound signal lesions. What is improved in the method is how to link the singular entropy of the contour signal and detail signal.

Acknowledgment. This work was sponsored by Science Fund of University of Jinan (XYK1222), Science Fund of Jinan Young Star (20090206), Project of Shandong Province Higher Education Scientific Research Program (J11LF02).

References

1. Luo, J.-Z.: Cardiac Auscultation, vol. 3, pp. 32–50. People's Medical Publishing House, BeiJin (2000)
2. Dock, W., Zoneraich, S.: A Diastolic Murmur Arising in A Stenosed Coronary Artery. *The American Journal of Medicine* 42(4), 617–619 (1967)
3. El Asir, B., Khadra, L., Al-Abbasi, A.H., et al.: Time-frequency Analysis of Heart Sounds. In: 1996 IEEE TENCON. Digital Signal Processing Applications, vol. 2, pp. 553–558 (1996)
4. Gauthier, D., Akay, Y.M., Paden, R.G., et al.: Spectral Analysis of Heart Sounds Associated with Coronary Occlusions. In: 6th International Special Topic Conference on Information Technology Applications in Biomedicine, pp. 49–52 (2007)
5. Cheng, Y.-P.: Matrix Theory, vol. 1, pp. 225–232. Northwestern polytechnical University Publishing House, Xi'An (2000)

Research on MTMP Structure Chlorine Dosing Decoupling Control

Xie Peizhang and Zhou Xingpeng

Key Laboratory of Measurement and Control of CSE, School of Automation,
Southeast University, Ministry of Education, Nanjing, China, 210096
xpzseu@gmail.com

Abstract. In this paper, multi tunnels multi pools (MTMP) structure chlorine dosing control method is studied. First residual-chlorine decay model is proposed, then MTMP chlorine dosing process model is acquired. After that, wavelet neural network is introduced to identify the \mathcal{O} -th order inverse system so that pseudo-linearization system can be obtained. Then Time delay disturbance observer (DOB) control algorithm is designed for each decoupled subsystem, high performance and improved robustness are obtained. Simulation and application in tap-waterworks at Suzhou (China) shows that the algorithm is able to resist the model mismatch, disturbance and time delay, also the lower unit consumption of chlorine is obtained.

Keywords: Chlorine dosing, Neural networks inverse system, Pseudo-linearization, Time delay DOB, Multi tunnels multi pools (MTMP).

1 Introduction

A major objective of drinking water treatment is to provide water that is both microbiologically and chemically safe for human consumption, so it is a key process in water treatment. While The formation of potentially harmful trihalomethanes (THM) when using chlorine as a sanitizer in potable water supplies has led to tighter regulatory controls and hence a need for better control algorithm. Chlorine dosing is a complicated system with nonlinear, larger time-delay, time-varying and multi models; also the couple is introduced to the system due to the multi tunnels multi pools (MTMP) structure. Residual-chlorine must be controlled smooth and steady so that the THM can be reduced. A typical MTMP sanitizer dosing system is shown in Fig.1.

There are some papers [1,2] on residual-chlorine decay model. Based on these models, some control algorithm is studied. In paper [3, 4] a decentralization structure is proposed in the previous paper for robust model predictive control (MPC) of chlorine residuals in drinking water distribution systems (DWDS). These algorithms focus on the chlorine residuals in water distribution system, little thought is given to tank chlorine residuals, also time-delay, model mismatch and couple is out of consideration.

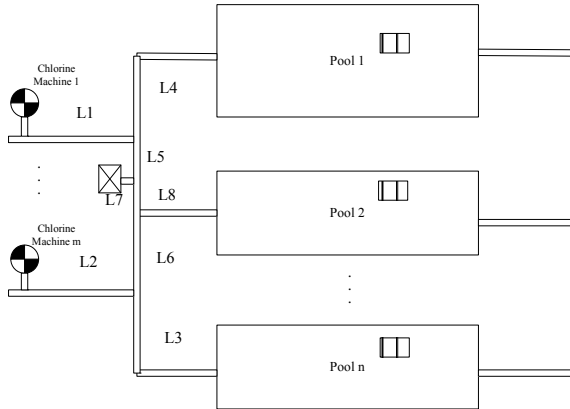


Fig. 1. Multi tunnels multi pools structure sanitizer dosing system

Time-delay DOB-NN [5-7] inverse system is proposed in this paper. Neural network inverse control method can realize the linearization decoupling control for MTMP structure chlorine dosing system. It constructs the α -th order inverse system using wavelet neural network and then cascades the original system so that the system can be transferred to a kind of normal decoupled system equipped with linear transferring relationship. Then Time delay DOB control algorithm is designed for each decoupled subsystem, high performance and improved robustness are obtained.

2 Multi Tunnels Multi Pools Structure Sanitizer Dosing System

There are kinds of sanitizer, chlorine is used the most usually because of its convenient use, storage and simple operation. Residual chlorine is the main parameter of disinfection performance in water treatment. There are some papers about the decay law of residual-chlorine in water [8].

In this paper first-order reaction kinetics equation is chosen, it can be described as follows: $C_B = C_A \cdot \exp(-kt)$

Where: C_A , C_B is the chlorine concentration of time A and time B, k is the chlorine decay factor.

The process can be divided into two parts, one is the rapid process and the other is the slow process as shown in Fig.2. During the rapid process, chlorine dosing to the clean water, the consumption of chlorine is very large, it is related to the initial dosage and the amount of NH3. Based on the theory, combined to the experiment, the approximate model can be acquired. The Rising Curve of chlorine dosing process is shown in Fig.3.

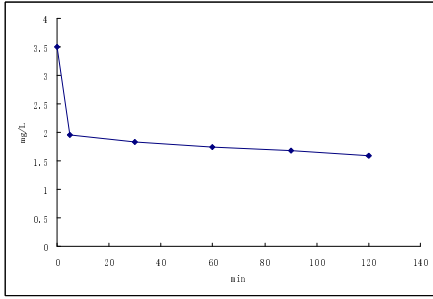


Fig. 2. Residual-chlorine decay curve

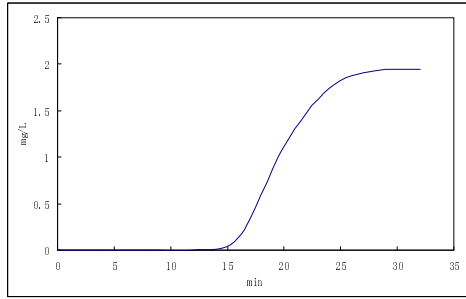


Fig. 3. Rising Curve of chlorine dosing process

In this paper, the model of multi tunnels multi pools structure chlorine dosing system is as follows:

$$\begin{cases}
 y_1(s) = \eta_{11}(X_1 - C_1) \frac{e^{-\tau_{11}s}}{T_{11}s + 1} + \eta_{21}(X_2 - C_2) \frac{e^{-\tau_{21}s}}{T_{21}s + 1} + \dots + \eta_{m1}(X_m - C_m) \frac{e^{-\tau_{m1}s}}{T_{m1}s + 1} \\
 y_2(s) = \eta_{12}(X_1 - C_1) \frac{e^{-\tau_{12}s}}{T_{12}s + 1} + \eta_{22}(X_2 - C_2) \frac{e^{-\tau_{22}s}}{T_{22}s + 1} + \dots + \eta_{m2}(X_m - C_m) \frac{e^{-\tau_{m2}s}}{T_{m2}s + 1} \\
 \vdots \\
 y_n(s) = \eta_{1n}(X_1 - C_1) \frac{e^{-\tau_{1n}s}}{T_{1n}s + 1} + \eta_{2n}(X_2 - C_2) \frac{e^{-\tau_{2n}s}}{T_{2n}s + 1} + \dots + \eta_{mn}(X_m - C_m) \frac{e^{-\tau_{mn}s}}{T_{mn}s + 1}
 \end{cases} \quad (1)$$

3 Design of Improved Time-Delay DOB Based on Pseudo-linearization System

The case study in this paper (waterworks at Suzhou in China) is a complicated MIMO system with large time delay, nonlinear and couple. It is a non-minimum phase system, and the original DOB can not deal with non-minimum phase system. Because of time delay, the inverse system can not be achieved due to advanced arguments. Wavelet Neural Network is proposed to identify the time-delay inverse system, and then this inverse system cascades the original system so that time-delay pseudo-linearization system can be obtained, after that this MIMO system can be transformed to time-delay SISO system without coupling.

1 Block of time-delay DOB based on pseudo-linearization system

The structure of time-delay DOB based on pseudo-linearization system is shown in Fig.4. P is the pseudo-linearization system consists of neural network inverse system and the original system. P is considered as follows,

$$P_i = s^{-n_i} e^{-t_i s}$$

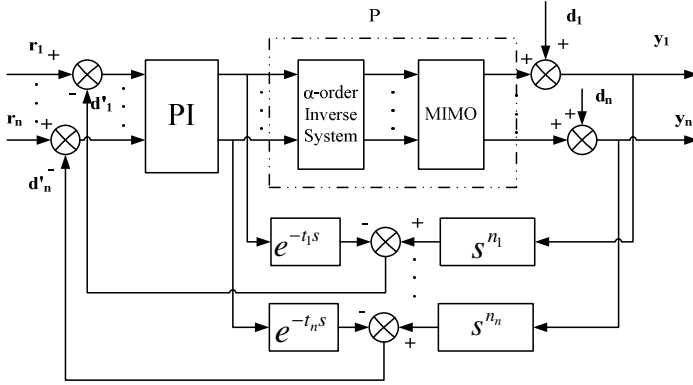


Fig. 4. Block of DOB based on pseudo-linearization system

Where n_i is the order degree of neural network input i . t_i is the minimum time delay of Channel i .

$$Y_i(s) = G(s)R(s) + G_d(s)D_i(s)$$

$$G(s) = \frac{G_p(s)G_c(s)}{1 + G_p(s)G_n(s)G_c(s) - e^{-t_i s}G_c(s)}$$

$$G_d(s) = \frac{1 - e^{-t_i s}G_c(s)}{1 + G_p(s)G_n(s)G_c(s) - e^{-t_i s}G_c(s)}$$

Where $G_p(s)$ is the pseudo-linearization system, $G_n(s) = s^{n_i}$, $G_c(s)$ is the transfer function of PI controller.

Thus, $d_i(s)$ can be rejected by time delay DOB based on pseudo-linearization system using WNN.

4 Case Study

In this paper, the parameters of the case study are shown as following:

$m=2, n=2. \eta = \begin{bmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{bmatrix} = \begin{bmatrix} 0.85 & 0.15 \\ 0.1 & 0.9 \end{bmatrix}$ is a matrix of flux ratio into pools

from tunnels. $\tau = \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{bmatrix} = \begin{bmatrix} 11 & 12 \\ 12 & 13 \end{bmatrix}$ is a matrix of time delay.

$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} 1.8 & 2.2 \\ 2.4 & 1.8 \end{bmatrix}$ is a matrix of parameters of inertial.

(1) Proof of the exists of the inverse system

2 inputs 2 outputs time-delay chlorine dosing model can be written as following:

$$\begin{cases} y_1(t + \tau_{11}) = x_1(t) = \eta_{11}u_1(t) \cdot \exp(-kt) + \eta_{12}u_2(t) \cdot \exp(-kt) \\ y_2(t + \tau_{12}) = x_2(t) = \eta_{21}u_1(t) \cdot \exp(-kt) + \eta_{22}u_2(t) \cdot \exp(-kt) \end{cases}$$

Remark 1, time delay of each single model can be treated as the same, and then time delay can be separated from the chlorine decay model.

$$\frac{\partial X}{\partial U} = \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix}, \text{ obviously } \det\left[\frac{\partial X}{\partial U}\right] \neq 0, \text{ so the inverse system exists.}$$

(2) Simulation of no disturbance with the algorithm proposed in this paper

Good results can be obtained with the algorithm using DOB based on wavelet neural network inverse system as is show in Fig.5. Compared with other algorithms, DOB-NN inverse system is much better and no system overshoot is observed.

(3) Simulation of mismatch disturbance

The result is shown in Fig.6, the controller designed using improved DOB based on wavelet neural network is able to resist the disturbance caused by mismatching disturbance or matching disturbance.

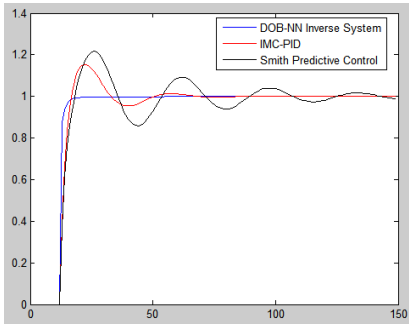


Fig. 5. Simulation of no disturbance

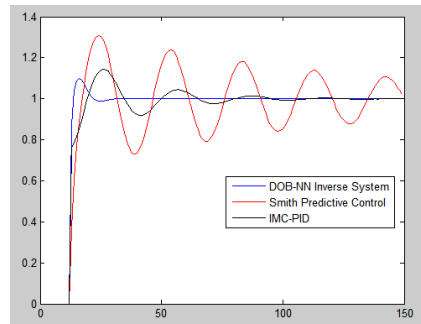


Fig. 6. Simulation of disturbance

5 Conclusion

In this paper, residual-chlorine decay law is presented, after that process model of MTMP chlorine dosing system is set up. Then Time delay DOB-NN inverse system control algorithm is proposed in this paper. Simulation and application in tap-waterworks shows that the algorithm is able to resist the model mismatch, disturbance and time delay.

References

1. Haiyan, C., Xinhua, Z., Yin, Y.: Research on residual chlorine decay model of reclaimed water network. In: 2011 International Conference on Electronics, Communications and Control (ICECC), pp. 3899–3902. IEEE (2011)
2. Clark, R.M., Sivaganesan, M.: Predicting chlorine residuals in drinking water: Second order model. *Journal of Water Resources Planning and Management* 128(2), 152–161 (2002)
3. Chang, T., Brdys, M.A., Duzinkiewicz, K.: Decentralized Robust Model Predictive Control of Chlorine Residuals in Drinking Water Distribution Systems. In: ASCE, pp. 1–10 (2004)
4. Duzinkiewicz, K., Brdys, M.A., Chang, T.: Hierarchical model predictive control of integrated quality and quantity in drinking water distribution systems. *Urban Water Journal* 2(2), 125–137 (2005)
5. Kim, B.K., Chung, W.K., Ohba, K.: Design and performance tuning of sliding-mode controller for high-speed and high-accuracy positioning systems in disturbance observer framework. *IEEE Transactions on Industrial Electronics* 56(10), 3798–3809 (2009)
6. Jung, S.: On the unified approach to the disturbance observer. In: 2012 12th International Conference on Control, Automation and Systems (ICCAS), pp. 573–578. IEEE (2012)
7. Rahman, A.A., Ohnishi, K.: Robust time delayed control system based on communication disturbance observer with inner loop input. In: IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, pp. 1621–1626. IEEE (2010)
8. Feben, D., Taras, M.J.: Studies on Chlorine Demand Constants. *Journal of American Water Works Association* 43(11), 922–932 (1951)

The Research on Electronic Data Forensic Institutions Equipment Configuration Standards

Mai Yonghao¹, K.P. Chow², Zhou Gang³, Lu Zhengwu⁴, and Zhang Jun⁵

¹ Hubei University of Police, Hubei Wuhan

² The University of Hong Kong, Hong Kong

³ Wuhan Engineering Science and Technology Institute, Hubei Wuhan

⁴ Communication University of China, Beijing

⁵ Hubei University of Police, Hubei Wuhan

Abstract. Equipment configuration standard is the important content of forensic¹ standardization, standardization. The paper briefly describes the background to carry out the research, elaborated on the basis of the combination of electronic data forensic architecture model of the process control and quality management chain of evidence equipment configuration standards oversight concepts into the relevant principles, propose reference standard for electronic data accreditation bodies equipment configuration, and the analysis shows.

Keywords: Electronic Data Forensic, System Architecture, Equipment, Configuration Standards.

1 Background

With the extensive application of electronic products, the importance of electronic evidence in litigation activities become increasingly apparent. Gradually with the National Forensic Reform, standardization of electronic data forensic work is gradually expand.

In 2006, the Ministry of China Justice forensic Authority issued a "the forensic institutions basic equipment configuration standards (provisional)" (hereinafter referred to as the "the configuration standard"). Recently, the forensic management department of the Ministry of Justice seek for opinions towards the society on configuration standards widely. The original Identification of audio and video materials were identified three types of sound recordings identification, image data identification and electronic data. To promote the standardization and standardization of electronic data accreditation bodies equipment configuration development in the form of the will of the state.

¹ Fund Project : Law Society funded projects in Hubei Province in 2010: Project No.SFXH210.

2 Configuration Principle

Electronic data forensic means electronic data forensic appraisers get use of computer science and technology or specialized knowledge to identify and judge of litigation involving and provide expert opinion activities. From the legal perspective, the electronic data forensic is a serious litigation activity, more as rigorous scientific research, as shown in Figure 1 electronic data forensic architecture. Preflight work is mainly by means of photographs, video, and the live test, true record samples before accepting a variety of state and electrical characteristics; receiving samples including checksum after the formal acceptance of the samples (ie, hash values) calculate in mutual confirmation archive hash value; consistency samples cloning mainly through "bit for bit" cloning and hash value comparison checksum copy of the original samples; data analysis is mainly through a variety of forms, and more kinds of means, a variety of levels of granularity to complete work on a comprehensive analysis of the data information; extract fixed, expert conclusions and court question refers to extract and fixed valuable evidence to the formation of the expert conclusions on the basis of the analysis of the data, and do a good job to appear accept the preparation of the question. This seven steps reflect the entire process of electronic data forensic work environment, evidence chain supervision and equipment evaluation access is a concrete manifestation of the electronic data forensic process control and quality management.

Electronic data forensic case studies			
Electronic data forensic legal programming			
Electronic data forensic technology programming			
Electronic data forensic program development			
Work environment	Pre-screening and accepting		Supervision of the chain of evidence
	Samples receiving		
	Samples cloning		
	Data Analysis	Time analysis, spatial analysis, structural analysis, particle size analysis	
		Functional analysis, correlation analysis, data analysis, code analysis	
	Extract fixed		
	Appraisal conclusions		
	Court question		
	Access to instruments and equipment evaluation		
Legal qualification of the personnel and agencies			
The technical basis of the electronic data forensic			
Legal basis for electronic data forensic			

Fig. 1. Electronic data architecture

Therefore, the configuration standard of the equipment should be based on the identification of the entire process, fully covering technology, management and legal aspects of the suit the characteristics, the basis of the characteristics of the prominent identification process control and quality management, and to reflect the doubts reproduce the full chain of evidence supervision, taking into account the mandatory introduction of the national evaluation of the access mechanism to ensure the standard-setting, scientific, normative and broad applicability.

3 Configuration Standards and Instructions

3.1 Essential Equipment

Based on the steps of electronic data forensic, there should have essential equipment such as cameras, and video cameras at scene investigation and samples preflight, samples cloning equipment and evidence of a fixed link must have read-only interface, data cloning tool and school the experience the code calculation tool, which is the fundamental guarantee of the electronic data the forensic process control and quality management. Read-only, clone and check the three basic technologies, to ensure that the electronic data forensic not only qualitatively but also quantitatively, and can accurately reproduce the distinctive feature of electronic data; evidence discovery and analysis identified links essential equipment inspection dedicated computer and electronic data recovery, search, analysis software, to ensure to find and extract the evidence and analysis in depth to form a complete chain of evidence. In the same time, the equipment is also essential for the quality management of the entire identification process electronic evidence storage cabinets and other related tools. Summed up above the necessary equipment is divided into eight types. The following analysis shows:

Cameras and Camcorders

Camera or video is a necessary part of handover of the samples scene forensics as well as identification of inspection bodies and the principal. The camera or video focuses on some of the important details of the samples, including the appearance of the samples, models, interface type, storage capacity, SN encoding, whether the package or affixed with a seal, numbered, write text and other important characteristics are important evidence of state and the integrity of proof samples for censorship. In addition the camera also necessary to audit whether the identification process legitimate, an important means of, for example, when no live witnesses or online forensics test whether the operator identification standards to operate.

Read-Only Interface

Read-only interface device protect electronic data samples is not "pollution" through the digital read-only technology, which can protect samples data from the acceptance to the identification of the completion of the original and integrity of the entire process.

Read-only tool is divided into software read-only and hardware read-only. Software read only rely on the operating system processes instructions to achieve the read-only. Hardware read only rely on increasing hardware devices in the identification computer and storage devices, and shielding writing into the signal data read-only operations. Samples to conduct an inquest must be read-only interface device. In order to guarantee the origin and integrity of electronic evidence, to circumvent identification inadvertently change the seized material, add or delete operation. Due to the diversity of the seized material interface type, you need to be equipped with a variety of commonly used types of interface read-only interface tools, such as IDE read-only interface, SATA read-only interface, SCSI read-only interface, SAS read-only interface. In a number of configurations, all kinds of read-only interface device shall be equipped with at least three sets the routine identification requirements to ensure that the disk array, such as raid5 disk array is composed of at least three hard disk, restructuring is required to ensure that the three sets of read-only interface.

Data Cloning Tool

Firstly, the data copying tool is different with data cloning tool. We usually call copy relative to the operating systems, a simple copy of the stored data contents of the medium, the smallest unit of the copying operation is a file, while a clone is a storage unit for each of the data storage device, is a "bit mirror bit "overall. The difference of operating granularity is very large. Such as the data contents of C in the hard disk (A) are copied into the empty hard disk B, the hard drive of the contents A and B are both C . But the data obtained in the data recovery operation on the B disk is likely not as C. The cloning operation can not only ensure that both A and B of the hard disk, what content data obtained by the recovery operation for the C, and two hard disk check code are also the same.

Secondly, the clone in the assessment process does not destroy and modify the data integrity of the original samples, clone hard disk duplicator or cloning software in the original samples to the target samples. Because of this clone is a bit of cloning the contents of the original samples in the target samples completely reflected, not lost, missing, it will not be modified, including deleted files, unallocated space Print buffer, data residual District and the file fragmentation, so that you can only forensic identification samples on target after cloning.

Again, cloning is evidence reliable backup carrier, is an important part of the preservation of evidence. When the evidence is being questioned, we can reproduce the identification process through cloning Samples way to eliminate questioned.

Finally, in some cases, there are many unpredictable operation of the check of the original samples, experiments, may affect the original electronic evidence. Some samples cannot be taken away from the scene, or the need to place the disk array data the recombinant transfer to forensic hard drive, or need simultaneous identification of many of the same samples, when samples data cloning is very necessary, not only can

improve work efficiency and effectively avoid causing irreversible destruction of the seized material data, you can also ensure the effective control of the identification process.

Checksum Calculation Tools

Checksum comparison is one of the three basic technologies of electronic data forensic process control and quality management, both assisted cloning technology to accurately reproduce the identification process. And it can be used for the preservation of electronic evidence, achieved through checksum hash value. For example, MD5 and SHA-1 are two common hash algorithm, they can be verified for any type of file, hard disk partition or the entire hard disk, output 128bit and 160bit fixed-length hash value. Even a single punctuation or hard data changes in the same file and hard disk partition will result in different hash value of the output. It mean this is the checksum comparison. Files, hard disk partitions or entire hard disk capacity is large, conventional hard to find is "tampering", but by the checksum hash value, we can verify that the files, partitions or entire disks in any case hash values checksum object to determine whether it has been tampered with. The identity identification of intellectual property is the use of this principle.

Checksum comparison is an important inspection means of the evidence chain supervision. This is because the checksum value than the samples, you can test whether all aspects of the entire judicial process caused by "pollution" of samples to ensure that the evidence original and consistency. While some have electronic data recovery, search and analysis functions integrated forensic software usually integrated checksum calculation function, but some only need to calculate the checksum, with specialized compact and practical checksum calculation tool becomes efficient and necessary. There are also views that the comprehensive forensics software integration checksum calculation function should no longer be listed separately, but in view of the checksum calculation tool of special significance and practical role, I believe that should be singled out as one of the essential tools.

Electronic Data Inspection Dedicated Computer

Electronic data verification computer is dedicated to the examination and analysis of electronic data computer. It should be equipped with two in generally, and a common identification can be connected to the Internet, the other one cannot be connected with any network, specifically identifying confidential data. Some electronic data inspection workstations, integrated directly into the read-only lock, card reader, clone machines, printers and other hardware devices, installed all kinds of tests and analysis software and verification tools, and electronic data test can also be used as a dedicated computer.

Integrated Electronic Data Recovery, Search, Analysis Software

Recovery, such as the common Encase, FTK, X-Ways, Winhex, DataCompass and comprehensive electronic data search, analysis software, such integrated software

complete electronic evidence display, search, recovery, extraction, validation, analysis, mirroring a variety of functions, for electronic data forensic institutions should be equipped with at least 1-2 kinds. At the same time should pay attention to such software feature-rich and highly professional and has a critical influence on the identification results for evaluation access to ensure the identification of the legality, accuracy and impartiality is particularly important.

Electronic Evidence Storage Cabinet

The electronic evidence storage cabinet is mainly used for the safe storage of the samples. It is an important part of the accreditation bodies process management and quality management, accreditation bodies to manage samples necessary equipment. And It should have a certain degree of anti-theft feature.

Other Necessary Tools

Some common gadget in the electronic data forensic work is also essential, due to the variety and miscellaneous. Here summarized with example: various interfaces riser card and adapter bridge (SATA rpm IDE, Micro-SATA to SATA, SAS to SATA, etc.), all kinds of memory card reader (common storage media, such as SD card, MMC card, CF card, TF card reader), a large-capacity hard disk for storing clone data (500GB or more), all kinds of data cable and power cord, screwdriver and within the screw from the (To demolition notebook), anti-static bag and gloves, CD burner, labels and markers, and so on.

3.2 Optional Equipment

A wide range of optional equipment, different functions, electronic data forensic institutions can be flexibly configured according to their own business and their own economic strength. Optional equipment can be divided according to the categories of electronic data forensic level for technical support equipment, the applications support equipment and other tools and equipment. At the same time, taking into account the optional equipment too much, in order to avoid making the configuration standard directory is too long, the only technical support for devices and applications support equipment as optional equipment for reference.

Technical support device: password cracking system, professional data recovery tools, disk array recombinant equipment, mass data storage systems.

Application support equipment: comprehensive forensic analysis of real-time communication tools, virus and malicious code analysis tools, electronic documents and data messages Comprehensive analysis tools, data comparison tools, on-site forensic tools, online forensics tools, cell phone data extraction, recovery, analysis tools, MAC/LINUX system inspection tools.

Other tools and equipment: data destruction equipment and storage media repair tools, industrial clean room, welding sets, magnetic force microscopy, voice recorder, scanners, shredders, disc destruction machine, video surveillance equipment, switches, routers, hardware firewalls, and so on. The specific configuration standards are shown in Table 1.

Table 1. Electronic data forensic institutions equipment configuration standards

03	Electronic data identification	Camera	1	Essential	Electronic data inspection system shall be equipped with measures to protect against computer viruses and other malicious code and network intrusion
		Video camera	1	Essential	
		Read-only interface	3	Essential	
		The data clone tool	1	Essential	
		Checksum calculation tools	1	Essential	
		Electronic Data Inspection dedicated computer	2	Essential	
		Integrated electronic data recovery, search, analysis software	1	Essential	
		Electronic evidence storage cabinet	1	Essential	
		Other necessary tools	1	Essential	
		Password cracking system	1	Optional	
		Professional data recovery tool	1	Optional	
		Disk array reorganization equipment	1	Optional	
		Mass data storage system	1	Optional	
		The instant messaging comprehensive forensic analysis tools	1	Optional	
		Virus and malicious code analysis tools	1	Optional	
		Comprehensive analysis tool for electronic documents and data messages	1	Optional	
		Data comparison tool	1	Optional	
		Live forensics tools	1	Optional	
		Online forensics tools	1	Optional	
		Cell phone data extraction, recovery, analysis tools	1	Optional	
MAC/LINUX system inspection tools	1	Optional			

4 Epilogue

The basic configuration standards of conduct electronic data forensic institutions equipment, can not only provide a direct reference to configure the equipment for the identification of bodies, and can effectively contribute to the identification of the study of the theory and technology, research and development of tools and equipment, as well as the evaluation of the establishment of the access mechanism, resulting in technical, scientific conclusions, by managing the technology standardized by law to promote the standardization of electronic data forensic identification legalize and standardize research.

References

- [1] Mai, Y., Sun, G., Xu, R., Dai, S.: Computer forensics and forensic, pp. 6–9. Tsinghua University Press, Beijing (2009)
- [2] Mai, Y., Zhao, Y., Ji., B., et al.: Electronic data forensic practice, pp. 8–22. Law Press, Beijing (2011)
- [3] Forensic Administration of the Ministry of Justice of the P.R.C. The forensic institutions basic equipment configuration standards (Interim) (September 1, 2006)

Design of an RDFizer for Online Social Network Services

Junsik Hwang, Hyosook Jung, Sujin Yoo, and Seongbin Park*

Korea University, Seoul, Korea

{js.seth.h,est0718,mynameislydia,hyperspace}@korea.ac.kr

Abstract. Recently, the number of online Social Networking Services (SNSs) such as Facebook or Twitter has been increasing and many users have various activities such as talk, image sharing and, making recommendations, etc. In addition, they can write their own profiles that contain a lot of personal information and values. If we can apply Semantic Web technology such as Resource Description Framework (RDF) to the information existing on SNSs, it can help analyzing valuable information in a machine-understandable way. In this paper, we propose an RDF Schema that defines social activities on SNSs. Then, we implemented a system by which data existing on SNSs can be converted into an RDF document using the proposed RDF schema.

Keywords: Semantic Web, Social Network Service, Semantic SNS.

1 Introduction

The number of data produced in social networks has recently increased and the types of data format are different among the social networks. In this situation, some services are created for the purpose of using SNS with efficiency. For instance, social comment services combine questions and answers with SNSs for better customer services. There are services that can spread an URL which contains information, pictures, video or online stores as a way of social marketing. In fact, SNSs tend to get connected each other more and more and it becomes complex owing to interoperability based on Open APIs. To solve this complexity, a service emerges that manages distributed bookmarks over SNSs. And a social network federation system that tries to integrate multiple SNS has been introduced [1][2].

In this paper, we present a system that transforms the data on SNSs which are written in various formats into an RDF document. First, we designed an RDF Schema where social activities can be defined. Then, we developed a system that converts data on SNSs into an RDF document (i.e., RDFizer [3]). By using our system, users do not have to spend much time on formatting data existing in different formats. In addition, both experts and end-users can easily convert data about social activities on SNSs into an RDF document.

* Corresponding author.

This paper is structured as follows. Section 2 describes related works to our research. Section 3 explains the proposed RDF Schema as well as the system in detail. Section 4 concludes the paper.

2 Related Works

SNSs have priceless data. In social search, Aardvark, data from twitter have been extracted [4]. In addition, some efforts have been made to apply Semantic Web technology to data of SNSs [1][2][5]. These researches tend to exploit well-known RDF Schema such as FOAF [6] and SIOC [7]. However, since SIOC focuses on describing “online communities”, it is not a proper schema to describe the structure of SNS [8].

FlickrWrapper [9] is using Flickr and DBPedia and it is some kind of meta search which merges results into an RDF file. FlickrWrapper searches pictures over Flickr[10] by just keyword and user activities are not considered at all. It is the purpose of our research to make a semantic environment of SNSs by converting SNS data to RDF documents using an RDF Schema and an RDFizer. Precisely, it includes user activities that consist of who, when, where, what etc.

In general, the main data of SNS is who, when, with whom and what did. The “who” is the account of the SNS, “when” is the time whose article is written, “what did” presents the action of publishing the text. “Who” is divided by the target conversation as the author of answered question or the public when text has not a specific audience. The main difference between our system called Semantic SNS and existing studies of SNSs with Semantic Web technology lies in the fact that we focus on social activities and try to integrate SNSs. Studies related to the characteristics of SNS[11][12][13] show that SNS have distinct features on their use. It is a tool for social communications and contributes sharing many contents even used in marketing. User’s experience over SNS occurs in a short time and is very intensive event.

3 Proposed System

In this section, we explain the RDF schema (RDFS) that we propose. Then we describe the structure of the proposed system as well as how it can be used. A proposed RDFS adopts well-known RDFSs such as Dublin Core [14], FOAF and SIOC. Dublin core is used for describing a common web resource, and FOAF is used for the profile of a user. The proposed RDFS defines vocabularies for external link (ss:externalLink) and message receiver (ss:receiver). An external link usually has a form of URL in text, message receiver is provided in different way in each SNS site. In Twitter, special tag is used for marking user id as listener. In the case of Facebook’s comment, it is inferred from who is writer of replied text. Commonly these messages have short lengths and replied each other, then a bunch of text composes a dialog (ss:dialog). Semantic SNS considers these features of SNS data. Figure 1 shows the structure of the proposed RDF Schema.

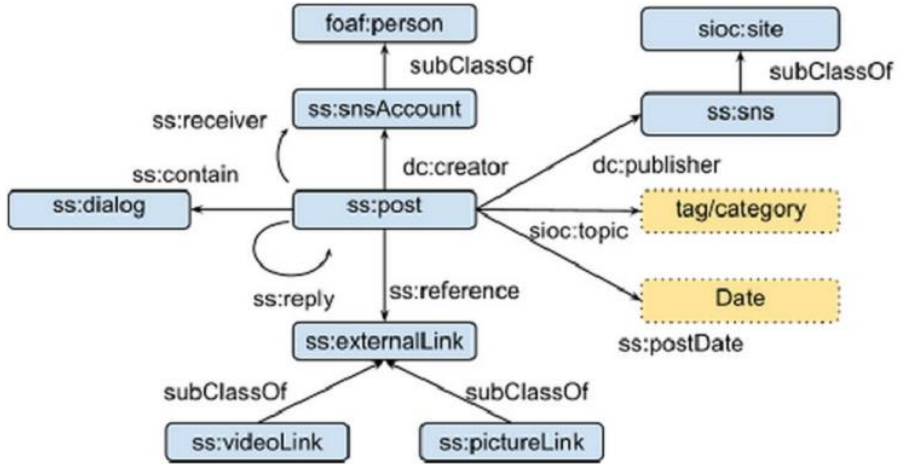


Fig. 1. Structure of the proposed RDF Schema

As an RDFizer, the main functionality of Semantic SNS is to generate an RDF document that contains data which comes from an SNS. In this research, social activities, written on SNSs as user’s action or its result are mainly data which is a form of text, image, reply, recommender etc. Most SNSs maintain these data and Open APIs provide these data with internal identifier. In case of Twitter and Facebook, we can make a URI of data from these identifiers hosted by Twitter or Facebook. For instance, Twitter’s URI has the following form, ‘https://twitter#!/SCREENNAME/status/IDENTIFIER OF STATUS’, where SCREEN NAME is account of user and IDENTIFIER OF STATUS is contained in Open API’s results.

As an RDFizer, Semantic SNS reuses these URIs and attaches properties to them. Consequently, a user can get an RDF document with one’s social activities. Semantic SNS needs proper authorization for accessing. Both Facebook and Twitter, provide authorization processes through OAuth [15]. After Semantic SNS has user’s permission, it accesses data through Open APIs[16][17] which are provided by both SNSs, respectively. In the current version, our system tries to have authorization of Facebook and Twitter one at a time. So, a user must have accounts on both systems.

Figure 2 shows the components of Semantic SNS. We used ‘node.js’[18] as a platform which works like a web server application. And a MVC Web framework, ‘Express.js’[19], is used for handling http request and response. And it needs OAuth, which means that the system manages some requests of HTTP on difference servers (exactly, Semantic SNS, Facebook and Twitter) over connectless context. When it gets authorization, it has to format data from SNSs. The Web Request Handler is built for processing http requests. The OAuth Manager deals with authorization states and manages OAuth sequence. In current version, OAuth Manager handles Facebook and Twitter and these functionalities are supported by facebook-js[20]and twitter-js[21]. The RDFizer collects data and sets its format.

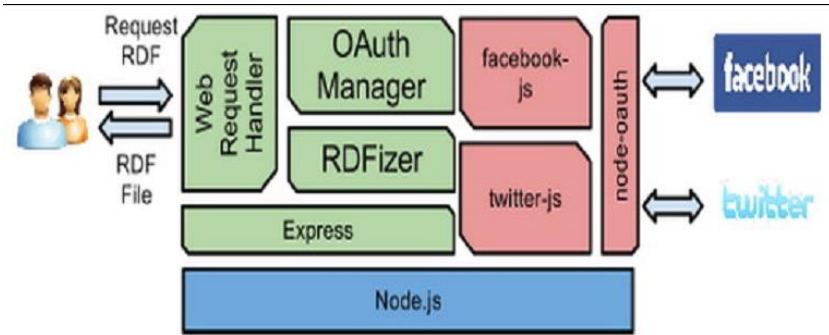


Fig. 2. Semantic SNS architecture

In order to generate an RDF document from data on SNSs, we have to deal with different data formats of SNSs. Facebook proposes concept of Open Graph[22] as a data scheme, and its data is queried by FQL (Facebook Query Language) [23]. For instance, if a user wants to query one’s facebook’s status history, the user just sends a request with a query sentence like ‘SELECT uid, status id, time, source, message FROM status WHERE uid= me()’. Twitter affords Twitter REST API[24], it consists of urls for getting Twitter’s data. In Semantic SNS, the RDFizer needs tweets written by a user, which are obtainable from timeline. So we use an API, ‘statuses/home timeline’, designed for getting online user’s tweets. Figure 3 is an activity diagram of the proposed system.

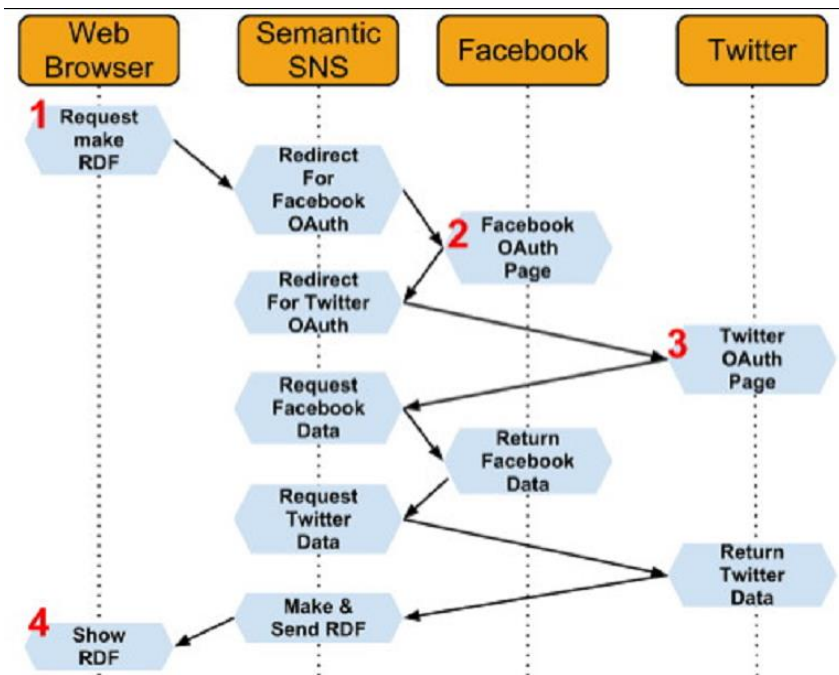


Fig. 3. Activity diagram of Semantic SNS

In this activity diagram, the starting point is a user's request which is regarded an order for making an RDF document. Clicking the "start" button in the web page lets the system try to get authorization for accessing SNSs. It means that a web page redirects to Facebook OAuth page. At that moment, the user will see a Facebook's page with question about permission. If the user agrees with accessing of Semantic SNS, then the system goes to the next step (i.e., starting with Twitter). Like Facebook, the system needs a permission to get data from Twitter. So Twitter's OAuth page will be shown. After user's agreements, Semantic SNS starts working. It starts to gather user's data contained in social activities, through Open API of each SNS. It sequentially demands data of Facebook and Twitter. After data collecting, Semantic SNS makes an RDF document and sends it to user's web browser. Then, the user will see an RDF document that contains information about social activities.

4 Conclusion

In this paper, we propose an RDF Schema that defines social activities existing on SNSs. We also implemented a system that converts data on SNSs into RDF documents using the proposed RDF Schema. A recent survey shows that the percentage of member overlaps between social network sites is relatively high [25]. This means that we might need to consider different SNSs for a single user in order to exploit social activities of the user. This aspect has been reflected in our system from the start of our research. Therefore the proposed system can help integrating data which come from different SNSs in a machine readable format (i.e., RDF document). Currently, we are working on ways by which the proposed RDF Schema can be extended. We plan to extend our system so that it can deal with social activities data from different SNSs than Facebook and Twitter.

References

1. Zhou, B., Wu, C.: Semantic Model for Social Networking Federation. *Advances in Information Sciences & Service Sciences* 3(11), 1 (2011)
2. Chao, W., Guo, Y., Zhou, B.: Social networking federation: A position paper. *Computers & Electrical Engineering* 38(2), 306–329 (2012)
3. RDFizers (2008), <http://simile.mit.edu/wiki/RDFizers>
4. Horowitz, D., Kamvar, S.D.: The Anatomy of a Large-Scale Social Search Engine. In: *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, USA, April 26-30*, pp. 431–440 (2010)
5. Rowe, M., Ciravegna, F.: Getting to Me Exporting Semantic Social Network Information from Facebook, *Social Data on the Web Workshop*. In: *Proceedings of the ISWC 2008 Workshop on Social Data on the Web (SDoW 2008)*, Karlsruhe, Germany (October 27, 2008)
6. FOAF, FOAF (2009), <http://www.foaf-project.org/>
7. SIOC (October 24, 2006), <http://sioc-project.org/>
8. Berrueta, D., et al.: SIOC Specification (March 25, 2010), <http://rdfs.org/sioc/spec/>

9. Flickr Wrapper (2007),
<http://www4.wiwiss.fu-berlin.de/flickrwrapp/>
10. Flickr, <http://www.flickr.com/>
11. Schaefer, C.: Motivations and Usage Patterns on Social Network Sites. In: 16th European Conference on Information Systems, ECIS 2008, Paper 143, Galway, Ireland (2008)
12. Hargittai, E.: Whose Space? Differences Among Users and Non-Users of Social Network-Sites. *Journal of Computer-Mediated Communication* 13(1), article 14, 276–297 (2007)
13. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of the 19th International World Wide Web (WWW) Conference, Raleigh, USA, April 26–30, pp. 591–600 (2010)
14. Dublin Core, <http://dublincore.org/>
15. Hammer-Lahav, E.: OAuth community, OAuth (September 5, 2007),
<http://oauth.net/>
16. Twitter developers (January 24, 2011), <http://dev.twitter.com/>
17. Facebook developers (August 2006), <http://developers.facebook.com/>
18. Node.js (2009), <http://nodejs.org/>
19. Express, <http://expressjs.com/>
20. Facebook-js, <https://github.com/masyllum/facebook-js>
21. Twitter-js, <https://github.com/masyllum/twitter-js>
22. Facebook open graph,
<http://developers.facebook.com/docs/opengraph/>
23. Facebook query language (FQL),
<http://developers.facebook.com/docs/reference/fql/>
24. Twitter REST API, <https://dev.twitter.com/docs/api>
25. Patriquin, A.: Connecting the Social Graph: Member Overlap at OpenSocial and Facebook (2007),
<http://blog.compete.com/2007/11/12/connecting-the-social-graph-member-overlap-at-opensocial-and-facebook/>

A Holistic Cloud-Enabled Robotics System for Real-Time Video Tracking Application

Bingwei Liu¹, Yu Chen¹, Erik Blasch², Khanh Pham³, Dan Shen⁴, and Genshe Chen⁴

¹ Department of Electrical and Computer Engineering,
Binghamton University, SUNY, Binghamton, NY, USA
{bliu11, ychen}@binghamton.edu

² Air Force Research Laboratory, Rome, NY, USA
Erik.Blasch@rl.af.mil

³ Air Force Research Laboratory, Kirtland AFB, NM, USA
khanh.pham@kirtland.af.mil

⁴ Intelligent Fusion Technology, Inc. Germantown, MD, USA
{dshen, gchen}@intfusiontech.com

Abstract. Future distributed sensor fusion applications will require efficient methods of information management such as Cloud computing. Using a server-based cloud-enabled software architecture would increase performance over hardware constraints (e.g., power, memory, and processors). In this paper, we propose a comprehensive framework for information fusion demonstrated for Cloud Robotics, which possesses user favorable features such as good scalability and elasticity. Robots are connected together to form a networked robotic system that is able to accomplish more computationally intensive tasks. Supported by the emerging Cloud computing technology, cloud-enabled robotic systems (CERS) provide even more powerful capabilities to users, yet keeping the simplicity of a set of distributed robots. Through an experimental study, we evaluate the memory, speed, and processors needed for a video tracking application.

Keywords: Cloud computing, image tracking, robot networks.

1 Introduction

Today's robots or robotic systems can perform complex tasks in real-world, dynamic environments, thanks to the advances of both microprocessors and generic CPUs. These systems are usually expected to achieve high mobility and are sensitive to communication time delays. When a task is beyond a single robot's capacity, multiple robots are required to accomplish the task. Networked robots, as distributed information fusion systems, require advances in resource management [1], Cloud-computing [2], and target tracking [3] to support effective situation awareness [3], [4] while at the same time providing secure communications [5]. Typically, robotic systems utilize image processing systems [6] for coordinated target tracking that have hardware limitations from intense measurement processing for image segmentation [7],

state estimation [8] and target assessment [9]. To address these issues, the paper describes a Cloud-enabled environment to increase video-tracking performance.

Traditional standalone robots are limited by constraints such as power consumption, computing ability, storage space, etc. Offloading part of a task to a remote server or distributing a complex task to a group of robots could potentially reduce response time, achieve more accurate decisions and consume less energy. Following the original idea of Internet-based tele-operated robots, the term “Networked Robots” was adopted by the IEEE RAS Technical Committee on Networked Robots in 2004 [10]. There are two different types of networked robots, tele-operated and autonomous [10]. Since networked robots are connected via a network, tele-operated robots can be accessed over a wider area. Autonomous robotic systems, on the other hand, allow Robots and sensors to coordinate through a network to perform complicated tasks that are difficult for a single robot. However, computation, storage and knowledge sharing are still limited at the distributed local network of robots.

As a new computing paradigm, Cloud computing (CC) has attracted researchers from the distributed computing community and information technology (IT) service providers. The well-known attractive features of CC include on-demand scalability of highly available and reliable pooled computing resources, secure access to metered services from anywhere, and displacement of data and services from inside to outside the organization. Due to the low cost of storage services provided in a Cloud, compared with purchasing and maintaining a storage infrastructure, it is attractive to companies and individuals to outsource applications and data storage to public Cloud computing services.

Cloud computing allow users to focus on their application without worrying about IT infrastructure plan. Central Processing Unit (CPU) cycles, storage space and even network services can be purchased on demand. When combining with the Cloud, robotic systems are able to take advantage of the almost unlimited parallel computing and vast storage space of Cloud computing. Cloud robotics was introduced by James Kuffner at Google to describe this new approach to robotics [11]. There is active research on Cloud Robotics in both Cloud and robotics communities [12], [13], [14], [15], [16], [17]. Most of these papers focus on special applications of Cloud robotics, with limited consideration about holistic system design, implementation details, or information fusion opportunities enabled from the Cloud environment.

In this paper, we propose a comprehensive distributed Cloud-enabled robotics framework for information fusion and provide a preliminary performance evaluation through a case study based on a video tracking application. In this framework, we considered the implementation of both the Cloud and the robot networks with additional security features, leading to a holistic framework. In addition, at the Cloud side, we include a virtual machine (VM) cluster and a physical machine cluster into our framework as a dynamic computing clusters.

The rest of this paper is organized as follows. Section 2 briefly discusses related work in networked robots and Cloud robotics. Then we introduce a holistic Cloud-enabled robotics system (CERS) framework in Section 3. Section 4 reports our preliminary performance evaluation result obtained through a case study of an image processing application. And we conclude this paper in Section 5 with some discussions about our on-going efforts.

2 Related Work

Chen et al. defined the concept of Robot as a Service (RaaS) based on Service-oriented architecture (SOA) [12]. The authors presented the idea of combining robot services with a Cloud, using Microsoft Robotics Developer Studio (MRDS) and Visual Programming Language (VPL). RaaS was implemented and tested on two processors, Intel Core 2 Duo and Atom.

Agostinho et al. [13] proposed a Cloud computing environment for networked robotics applications. VMs are assigned different roles in the Cloud environment. A layered workflow management system was used for scheduling purposes.

Kehoe et al. [14] illustrated a system architecture for Cloud-based robot grasping using a Google object recognition engine. A prototype and initial experiments for a Cloud-based robot grasping system were implemented in their work.

Arumugam et al. [15] proposed a distributed agents with collective intelligence framework, in which heterogeneous robots can work together in large environments. A Robot Operating System (ROS) platform [18] was used for sensor data collection and communication. In their implementation, Hadoop was used as a high performance computing and storage platform. A grid based FastSLAM algorithm was implemented as a Hadoop Map/Reduce task [15].

Hu et al. [17] suggested using gossip protocols for communication between robots within a highly dynamic mobile robotic network. No route discoveries and maintenance are needed in this system. It is simple to implement and has low computation and memory requirements. However the latency of message exchange could be expensive in this ad hoc wireless network. The authors also considered energy consumption in the decision of whether to offload computation to the Cloud.

Compared to previous works in Cloud robotics, we highlight the following contributions of this paper.

1) *Relatively complete framework*: The architecture of both the Cloud side and the local robot network side are considered in our framework. Previous work either built the Cloud environment on a single machine using a virtualization software instead of using a Cloud platform, or didn't provide enough specifications of the Cloud architecture.

2) *Dynamic computing cluster*: Dynamic computing cluster: We combine a virtual machine cluster with a physical cluster in our dynamic computing cluster infrastructure. This architecture has potential flexible scalability and efficient resource allocation.

3) *Working flow prototype*: We investigate the working flow of a user requesting a robot service from the web interface provided by the Cloud as a proof of concept example.

4) *Performance comparison between a physical workstation and multiple virtual machine instances*: We evaluate the performance of a workstation and different size of virtual machine instances in an video tracking algorithm in the processing of simultaneous requests.

3 A Holistic Cloud Enabled Robotics System

3.1 Cloud Robotics

Cloud robotic systems take advantage of all the benefits of the Cloud, while at the same time providing users who want to focus on the functionalities of robots in an economic way to investigate their robotic system. There are at least the following benefits when we integrate multiple robotic systems within a Cloud architecture.

1) *Faster robotic applications development*: Developers in robotics can cooperate in the platform provided by the Cloud to develop robotic applications in a more efficient way. Once application developers deploy their applications into the Cloud, they can take advantage of the fast provision technology of the Cloud to serve almost unlimited users.

2) *Easier to get started*: A user doesn't need to configure the development environment in order to have an initial idea about robots in general, or a specific type of robot. All they need is a web browser to access these services in Cloud robotics.

3) *More efficient robot resources usage*: Robot owners can also reduce their cost in maintenance by charging a small amount of fee to each user.

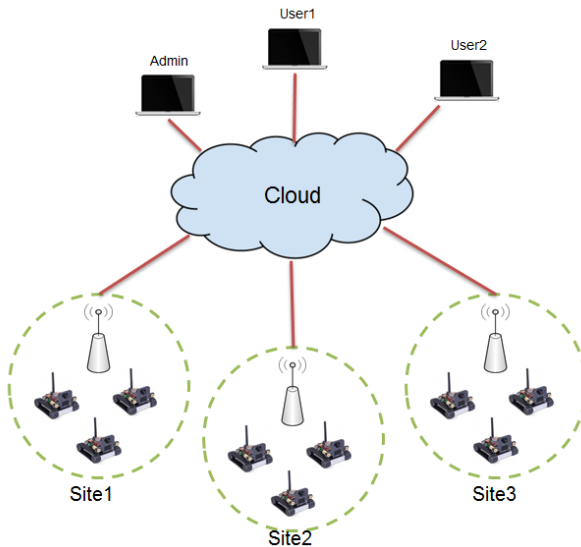


Fig. 1. A typical scenario of a Cloud Robotic System

In this paper, we propose a more general holistic framework for Cloud-enabled Robotic System (CERS). Fig. 1 is a typical scenario of a CERS. Robotic systems that are able to provide robot services register with the Cloud as robotic service providers. The Cloud provides a uniform web interface for all customers. Developers communicate and cooperate with each other through the web portal. Administrators also manage robotic systems through a specific secure website hosted in the Cloud. Robotic systems and the Cloud exchange messages through ROS messaging mechanism [18].

3.2 An Abstract Architecture

As a general discussion of CERS, we first consider the abstract architecture in Fig. 2. At the top of this system is an application layer, including three types of applications (APPs) for different purposes. The Management Apps consist of authentication and access control functionalities, as well as the management of computing, storage, network, auditing and QoS etc. The services Apps include customized applications for different robot systems, robot resource database and robot systems monitor etc. The user APPs provide applications directly interacting with end users, such as a web page that can see the video captured by a robot in real time. The Application Programming Interface (API) layer is the middleware between applications and underlying layers where developers use them for their application development.

Under the API layer is the computing, storage and databases platform layer. Basic databases such as user registration and robot states will be created for the management of the system. The Cloud provides elastic computing and storage resources on demand and schedules jobs or tasks according its load balance policies. High availability can also be provided to desired clients.

In order to provide generic services in Cloud, we suggest that heterogeneous robotic systems use the widely adopted Robot Operating System (ROS) [18] as the robot platform. The Cloud communicates with ROS directly to acquire data and send commands to robots. Each Robot team must have at least one ROS master to take care of message exchange, robot services registration and robot control. The residence of ROS masters is flexible. An ROS master can run on a local computer which locates at the same area as managed robots. It can also run on a virtual machine in the Cloud.

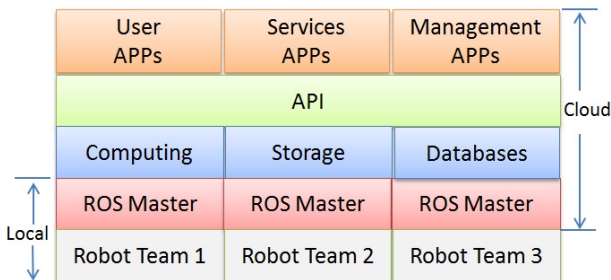


Fig. 2. Abstract Architecture of a Cloud Robotics System

3.3 Robot Network

In the ROS system, the term “nodes” are processes running on the robot system. Multiple nodes can run on the same ROS system. The ROS master serves as a name server for all other nodes so that they can find each other. Once the existence of other nodes are discovered, each node (robot) can communicate with any node directly in ad hoc wireless mode or through a centralized access point. A good choice to host a ROS master is a computer that is in the same local network with all other distributed robots. An access point is also needed in order to communicate with the Cloud. Another option is to run the ROS master on a virtual machine in the Cloud. In this case, a robot must have a public Internet Protocol (IP) address or a local manager node is added to route the traffic.

There are two running modes under our proposed CERS:

- 1) **Local mode:** *When the network connection is not satisfied, the local robotic network works under local mode. A manager node will be chosen and host the ROS master. This manager node can be a laptop or simply a robot that has most powerful processing ability.*
- 2) **Cloud mode:** *In order to take advantage of the Cloud, when the round-trip delay time (RTT) of the communications message to the Cloud is suitable for the services provided by the local robot system, the ROS master will switch to Cloud mode.*

Robot states are stored in a local ROS master as well as Cloud ROS master, and a clone of the local ROS master is also registered in the dedicated Cloud database.

3.4 Cloud Side Consideration

Fig. 3 illustrates a possible implementation at the the Cloud side in order to provide robot services to the public. The underlying Cloud infrastructure employs a dynamic virtual machine (VM) computing cluster and storage system. The following aspects are worth noting when designing a Cloud robotic system.

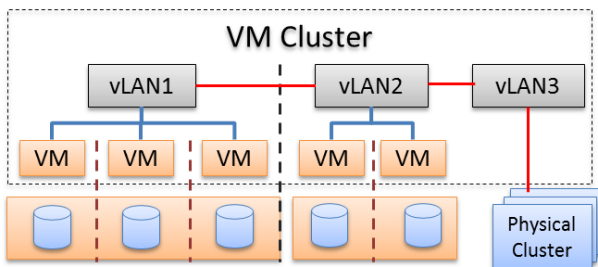


Fig. 3. Dynamic computing cluster and storage system

1) *Dynamic Computing Cluster*: The dynamic cluster is a computing cluster consisting of VM cluster and physical cluster. According to the system load, the Cloud will allocate appropriate computing resources to different tasks. The dashed lines in Fig. 3 represent the appropriate isolation between every two VMs. One of the VM in the cluster is chosen as a computing scheduler. This scheduler monitors the performance and load of all VMs. We assume that this VM has strong secure environment that is difficult to be compromised. This is reasonable since we can always use a private IP for this VM and further protect it behind a secure firewall. This elastic computing structure is able to process sophisticated computing intensive tasks.

2) *Separated Storage*: Every VM can have one or more associated storage volumes. These volumes are assigned by the hypervisor and cannot be modified without root privilege in the hypervisor. User data security and integrity can also be included in the system by adopting storage integrity auditing service. A distributed storage system like Hadoop Distributed File System (HDFS) [19] can be employed as the underlying storage infrastructure.

3) *Distributing Functionalities into VMs*: Every individual functionality in our model can be implemented on a virtual machine with proper computing and storage resources to handle user requests, message exchanges or performance monitoring. This is an efficient and economic way to implement a complex system. The elasticity of the Cloud will also assure high availability of services. Once any functional VM goes wrong, a new VM can be deployed in minutes to replace the bad VM.

3.5 Web Interface

Providing computing services through the web has been proved by public Cloud provider like Amazon to be a successful service delivery mechanism. We also recommend using RESTful web services [20] to implement Cloud applications.

The Cloud provides a uniform web interface for administrators, developers and regular users to access data and perform tele-operation to robots. Administrators monitor the health of VMs, states of robots and user behavior. Developers store program codes on the revision control and source code management system provided by the Cloud and can easily cooperate with each other on the hosted repository. Regular users visit the web interface to access history data, tele-operated robots, request data processing and monitor VMs with owner privilege when desired.

3.6 Security

Security is usually the first concern to both a service provider and a user. The provider wants to assure their properties are safe in the Cloud. Important and sensitive data in the Cloud should be stored securely. The Cloud should also be able to appropriately defense attacks to web server and user data. The study of security in Cloud computing is still an open area, hence it needs to be considered carefully when putting robot

resources online. A compromised server can send malicious commands to robots, causing tremendous lost for robot owners.

To protect user data and ensure the security of the system, several policies need to be enforced:

1) *VM Isolation:* As we mentioned above, VMs must be isolated by the hypervisor even when they are running on the same physical machine.

2) *Secure Storage:* The Cloud must implement secure storage of user data before deploying robot services. The Cloud can provide an auditing service for users to guarantee data integrity. A third party auditor (TPA) can be employed to audit the Cloud in data integrity. The auditing procedure is usually a challenge-response style. A user or the TPA challenges the Cloud with the integrity of his data. The Cloud then responds with a message to prove that it is actually possessing the user's data and all data blocks are intact in the Cloud storage infrastructure.

3) *Network Management:* Each virtual Local Area Network (vLAN) has strict rules to prevent unauthorized access even within the same vLAN. Usually, the hypervisor will take care of the packet routing and virtual network optimization. If necessary, a trusted network manager can be deployed to monitor the network with proper permission.

3.7 Work Flow Example

As a proof of concept example, we show a working flow of a user requesting a service through the Cloud robotic system in Fig. 4. The user first visits the web interface and requests for the service (1). The web server then call the user authentication and access control module (2), which then queries the user database (3) and grants the access if

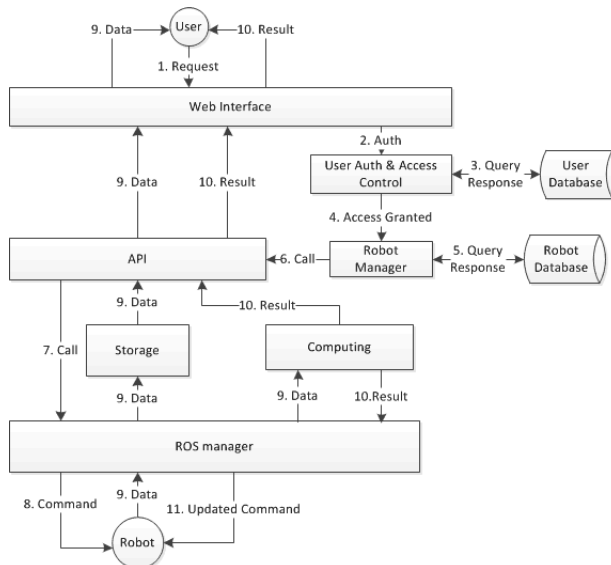


Fig. 4. Work flow when a user request a service from the web portal

the user has the right permission (4). The robot manager then checks out the robot database to make sure there exists a robot that can provide this server (5). Next, the API issues a message to the ROS manager (7). The ROS manager then sends command to the appropriate robot (8) and retrieves the requested data (9). The data is then passed all the way to the user. The computing module also performs necessary computations and returns the result to other modules and the user (10). Finally, new commands are sent by the ROS manager for further control of the robot.

4 Performance Evaluation

To evaluate the performance of offloading image processing in video tracking tasks to the Cloud, we designed a Cloud-enabled distributed robotics prototype, consisting of a remote robot network and a Cloud testbed in our datacenter. The robot network includes three SRV1 robots and an AR. Drone 2.0 flying robot. The flying robot collects image sequences with its on board camera and transfers to a local server or to the Cloud in real time. The Cloud runs a simple motion detection algorithm to track the motion of ground robots [8], [21]. We compare the performance in executing the video tracking task between virtual machine or physical machine by simulating up to twenty simultaneous requests of video signal processing and collects the performance data.

4.1 Experiment Setup

1) *Cloud Testbed*: The Cloud testbed consists of 16 servers in our data center, all using Xen Cloud Platform (XCP) 1.6 [22]. We choose Citrix XenCenter as our management software, which provides convenient management features for our experiment purpose. We can easily create, clone and move a virtual machine within the XenCenter. Live migrate, within the same pool or across different pools, is also an attractive feature. Each Cloud server is equipped with two Intel Xeon E5405 Quad-core processors at 2.0GHz, 32GB memory and 3TB storage. For this experiment, we only use a pool of four servers (Fig. 5). Fig. 6 shows the real-time monitor of the CPU, memory and network performance of the large instance in the experiment.

2) *Authentication and Access Control*: We have implemented part of the functionalities of the web interface in our framework. The administrator uses a web interface to monitor all robots' camera images. All users including the administrator are authenticated by a username password scheme before they can access any resource.

3) *Machines under Comparison*: To evaluate the performance differences among a local machine and virtual machines in the Cloud, we compared the local machine with three instances in the Cloud. Table 1 lists the specification of local machine and the three virtual machines instances we setup for performance comparison. We denote the VMs as small, medium and large instance according to their computing capacities. The small instance has comparative configuration with the local machine. The medium and large instances double the number of virtual CPUs and memory each time. All machines, physical or virtual, use a Ubuntu 12.04 LTS operating system.

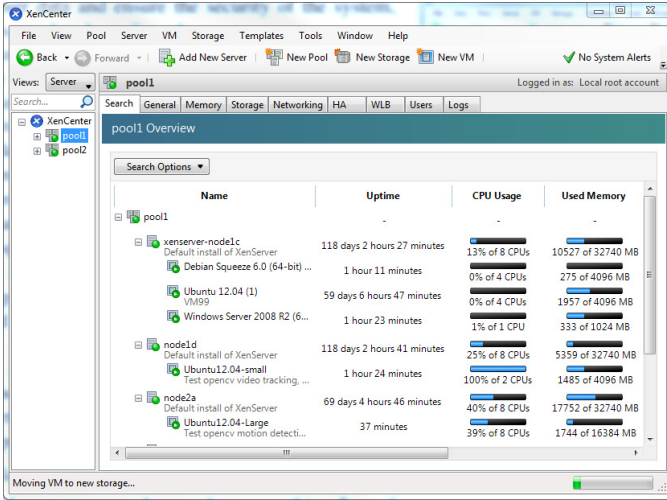


Fig. 5. The pool of Cloud servers

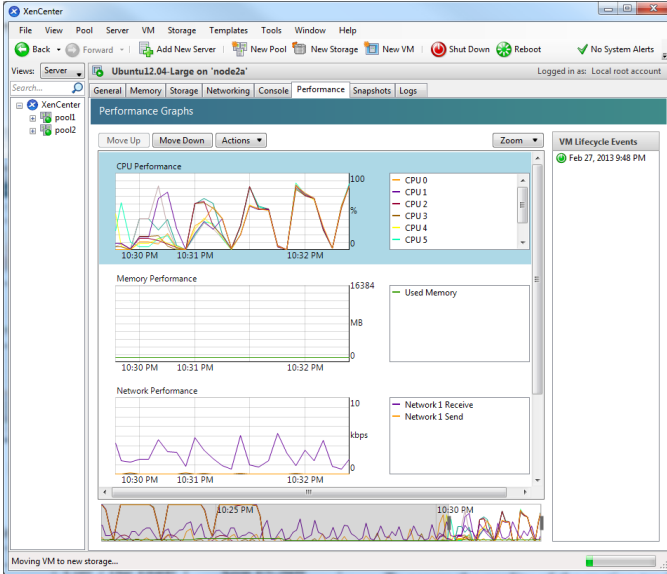


Fig. 6. VM performance monitoring

Table 1. Specification of Compared Machine

Item	Local Workstation	VM1	VM2	VM3
CPU Type	Core 2 Due E8400	Xeon E5-2609		
CPU Frequency (GHz)	3.0	2.4		
Number of Cores	2	2	4	8
Cache(MB)	6	10		
Memory(GB)	4	4	8	16

4) *Task Analysis:* The task of this simple prototype is to track the motion of ground robots using computing vision technologies. The image processing procedure is computationally expensive on a workstation or a laptop. We want to investigate the feasibility of offloading computing tasks to the Cloud and assess system scalability. To assure the quality of the motion tracking application, the frame per second rate must not fall under certain threshold so that video quality will not cause too much delay. The time to process a frame after captured by a camera consists of the time to transfer back and forth between a robot and the server as well as the time for the server to process this frame. Since the transfer time could vary under different network conditions, in this simple test, we only consider the frame per second that a server can process as the indicator of robust performance.

5) *Algorithm:* We tested a straightforward algorithm of target motion detection in this experiment. For each frame sent by the Robot, first the tracker finds the edge of every object using simple threshold algorithm. Then we find the contours in the edge frame, and find an approximate polygon for each contour and calculate a bounding box for it. Next we merge bounding boxes that are close to each other and finally get bounding boxes for all objects. Fig. 7 shows a frame of the result. The image processing algorithm for video tracking was written in C++ using openCV 2.4.4 libraries [23]. The same algorithm was implemented in all physical and virtual machines. The frame per second (fps) data was collected every a machine runs the processing algorithm. All fps values were average of 20 trials under the same condition for a number of simultaneous request.



Fig. 7. A frame with tagged targets of the video tracking algorithm

4.2 Results

The results of our experiment are illustrated in Fig. 8. Overall, the medium and large instances had higher performance than the others since they have more CPUs to process the images. All machines have almost the same fps for 1 to 3 requests, which means that handling 3 requests will not affect the CPU performance in any case of our

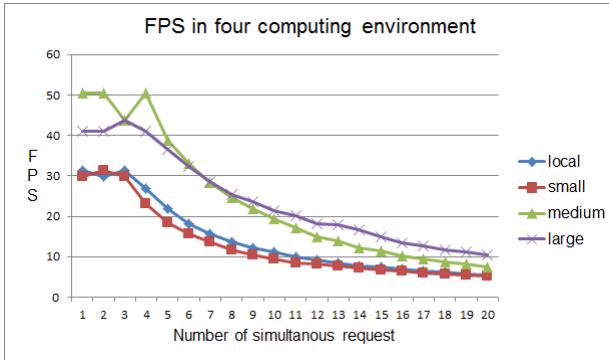


Fig. 8. Compare of four image processing environment: Local machine (local), small instance (small), medium instance (medium) and large instance (large)

experiment. The medium instance surprisingly had high fps than the large instance which has 8 virtual CPUs. One reason could be that the tracking algorithm was not optimized for parallel processing. A few requests at the same time will not take full advantage of the number of processors.

Started from 8 requests, the large instance had higher fps (25.27) than the medium one (24.58). Considering the general fps of the image processing algorithm, the large instance had fair performance in this experiment with over 10 fps for up to 20 requests. When there are more than 20 requests, the Cloud can easily assign more VMs to serve the users. An interesting phenomenon we observed in our experiment is that the amount of memory has no direct impact of the fps after 2 GB.

Because of the time consuming transfer of video back to the robot, we suggest that the server only return the coordinates of targets in each frame. This way the transfer delay from the Cloud to the robot is negligible, requiring only the transfer of image sequences or video streams through the Cloud. In our case, only 24B of data is needed for each frame. The Cloud architecture also enables the display of the results on a web page to the user.

5 Discussion and Conclusion

In this paper, we proposed a general architecture for a distributed Cloud-enabled robotic information fusion system. A performance evaluation was investigated on an video tracking task for a robot network. Our results show that offloading computation to the Cloud is feasible and is especially beneficial when there are a lot of robot networks requesting image processing tasks.

We conducted all the experiments in our Cloud testbed consisting of 16 servers running on a Xen Cloud platform. A web portal with authentication module has been implemented and allows the user to communicate with the Cloud. Our current work is just a beginning of exploring the possibilities in combining robotic systems with Cloud computing to accomplish more computationally intensive information fusion tasks.

There are still a lot of interesting performance and implementation questions to be investigated in the future, such as seamless integration with the ROS system to control robots in real time, efficient scheduling of computing resources and dynamic bandwidth allocation according to the load of the system.

References

- [1] Blasch, E., Bosse, E., Lambert, D.A.: High-Level Information Fusion Management and Systems Design. Artech House Publishers (2012)
- [2] Blasch, E., Chen, Y., Chen, G., Shen, D., Kohler, R.: Information Fusion in a Cloud-Enabled Environment. In: Choi, B.-Y., Han, K., Song, S. (eds.), Springer Publishing (2013)
- [3] Blasch, E., Kadar, I., Salerno, J., Kokar, M.M., Das, S., Powell, G.M., Corkill, D.D., Ruspini, E.H.: Issues and challenges in situation assessment (level 2 fusion). *J. of Advances in Information Fusion* 1(2), 122–139 (2006)
- [4] Blasch, E., Seetharaman, G., Pal, K., Ling, H., Chen, G.: Wide-area motion imagery (wami) exploitation tools for enhanced situation awareness. In: IEEE Applied Imagery Pattern Recognition Workshop. IEEE (2012)
- [5] Mazur, S., Blasch, E., Chen, Y., Skormin, V.: Mitigating cloud computing security risks using a self-monitoring defensive scheme. In: Proceedings of the 2011 IEEE National Aerospace and Electronics Conference (NAECON), pp. 39–45. IEEE (2011)
- [6] Lee, K.-M., Zhou, Z., Blenis, R., Blasch, E.: Real-time vision-based tracking control of an unmanned vehicle. *Mechatronics* 5(8), 973–991 (1995)
- [7] Shen, D., Blasch, E., Pham, K., Chen, G.: A clustering game based framework for image segmentation. In: 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), pp. 818–823. IEEE (2012)
- [8] Mei, X., Ling, H., Wu, Y., Blasch, E., Bai, L.: Efficient minimum error bounded particle resampling 11 tracker with occlusion detection. *IEEE Trans. on Image Processing (T-IP)* (2013)
- [9] Wu, Y., Cheng, J., Wang, J., Lu, H., Wang, J., Ling, H., Blasch, E., Bai, L.: Real-time probabilistic covariance tracking with efficient model update. *IEEE Transactions on Image Processing* 21(5), 2824–2837 (2012)
- [10] IEEE society of robotics and automation's technical committee on networked robots, <http://www-users.cs.umn.edu/isler/tc/>
- [11] Kuffner, J.J.: Cloud-enabled robots. In: IEEE-RAS International Conference on Humanoid Robotics, Nashville, TN (2010)
- [12] Chen, Y., Du, Z., García-Acosta, M.: Robot as a service in cloud computing. In: 2010 Fifth IEEE International Symposium on Service Oriented System Engineering (SOSE), pp. 151–158. IEEE (2010)
- [13] Agostinho, L., Olivi, L., Feliciano, G., Paolieri, F., Rodrigues, D., Cardozo, E., Guimaraes, E.: A cloud computing environment for supporting networked robotics applications. In: 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC), pp. 1110–1116. IEEE (2011)
- [14] Kehoe, B., Matsukawa, A., Candido, S., Kuffner, J., Goldberg, K.: Cloud-based robot grasping with the google object recognition engine. In: IEEE International Conference on Robotics and Automation. IEEE (2013)

- [15] Arumugam, R., Enti, V., Bingbing, L., Xiaojun, W., Baskaran, K., Kong, F.F., Kumar, A., Meng, K.D., Kit, G.W.: Davinci: A cloud computing framework for service robots. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 3084–3089 (May 2010)
- [16] Goldberg, K., Kehoe, B.: Cloud robotics and automation: A survey of related work. EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2013-5 (2013)
- [17] Hu, G., Tay, W., Wen, Y.: Cloud robotics: architecture, challenges and applications. *IEEE Network* 26(3), 21–28 (2012)
- [18] Robot operating system, <http://www.ros.org>
- [19] Hadoop, <http://hadoop.apache.org/>
- [20] Fielding, R.T., Taylor, R.N.: Principled design of the modern web architecture. *ACM Transactions on Internet Technology (TOIT)* 2(2), 115–150 (2002)
- [21] Ling, H., Bai, L., Blasch, E., Mei, X.: Robust infrared vehicle tracking across target pose change using l1 regularization. In: *Int. Conf. on Info Fusion*, vol. 1 (2010)
- [22] Xen cloud platform, <http://xen.org>
- [23] Opencv, <http://www.opencv.org>
- [24] Mell, P., Grance, T.: The nist definition of cloud computing (draft). NIST special publication, vol. 800, p. 145 (2011)
- [25] Wang, L., Liu, M., Meng, M., Siegart, R.: Towards real-time multi- sensor information retrieval in cloud robotic system. In: 2012 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 21–26. IEEE (2012)
- [26] Mei, X., Ling, H., Wu, Y., Blasch, E., Bai, L.: Minimum error bounded efficient l1 tracker with occlusion detection. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1257–1264 (June 2011)

Improving Network Health Monitoring Accuracy Based on Data Fusion for Software Defined Networking

Sejun Song

Texas A&M University
College Station, TX, USA
sjsong@tamu.edu

Abstract. Software defined networking (SDN) provides programmable centralized control architecture by separating and abstracting the control plane from the underlying forwarding system. In addition to enabling new applications such as traffic engineering and network virtualization, SDN also simplifies and automates network administrations to achieve increased network reliability and security. However, since events that occur within the network should be inferred by a centralized remote SDN controller, to make an accurate decision in time, the network event management should be able to effectively monitor and analysis the events. In this paper, we propose an efficient network health monitoring based on data fusion strategies to improve the decision accuracy for the centralized remote SDN controller.

Keywords: Network Health Monitoring, OpenFlow, Software Defined Networking, Data Fusion.

1 Introduction

Network health monitoring [1] against network failures, congestions, mis-configurations, and security attacks is an integral part of building dependable network services. Network management system (NMS), then, should be able to determine a detected problem's root cause in real time in order to isolate problem within the contained area. For example, a network failure can be caused by an actual network device failure, a configuration error, a software fault, network congestions, or a deliberate security attack. If it is not detected, identified, and managed timely, it will impact network reliability and may eventually cause significant service discontinuation.

Despite the effort by research communities as well as network operations to build an effective network health monitoring and management system, the current practices are still unreliable, inaccurate, and not scalable. It is mainly due to the traditional network infrastructure that is an ossified black box and network health monitoring and management practices take mainly remote approaches. As problems are pushed out to the intelligent edge devices and servers, diagnosis is postponed, network problems may be accumulated, and potential damages can be enlarged.

Recently, to enhance the problems caused by the network ossification, Software Defined Networking (SDN) [2] has been proposed by many network industries and

researchers. SDN provides programmable centralized control architecture by separating and abstracting the control plane from the underlying forwarding system. In addition to enabling new applications such as traffic engineering and network virtualization, SDN also simplifies and automates network administrations to achieve increased network reliability and security. Particularly, fueled by increasing data center networking and cloud computing, SDN has been building up significant momentum toward the production network deployment.

Since all the events that occur within the network should be inferred by a centralized remote SDN controller, the network event monitoring and management system on the controller should be able to effectively detect and analysis the network events to make an accurate control in time. However, as the underlying network is an inter-related complex system, it is not straightforward to identify a root cause of a problem. For example, a single problem may issue huge amount of related syslog events as well as some faults may induce a failure which is seemingly not directly related to the original source of the problem. These insignificant event reports may result false negative or false positive decisions which may cause yet another network service problem. In this paper, we propose an efficient network health monitoring based on data fusion strategies to improve the decision accuracy for the centralized remote SDN controller. Specifically, we discuss a couple case studies, new flow attack detection and event storm detection, to use data fusion strategies for improving decision accuracy and, in turn, for predicting the failures of network applications to ensure the high availability of the high performance data center and clouds.

2 The Case of New Flow Attack Detection

Unlike the traditional network system, an SDN facility such as OpenFlow moves the control plane from an individual switch to a remote controller. Hence, for unknown incoming flow packets, an OpenFlow switch needs to send a new flow request to the remote controller via a secure channel (SSL). In practice, adversaries can inject randomly generated New Flow packets into an OpenFlow switch port to saturate CPU usage of a switch (named New Flow Attack). A FlowVisor [3] is a remote controller designed to ensure the resource isolation within the OpenFlow switch. FlowVisor remotely monitors the switch's new flow packet count to control the CPU utilization of each switch. However, we have identified a few New Flow Attack scenarios that the remote FlowVisor cannot handle properly. For example, as presented in Figure 1, when a switch's CPU is already saturated by an instantaneous new flow attack, the actual outgoing number of new flow requests (i.e. 100 pps) sent to FlowVisor can be far less than the real incoming packet counts (i.e. 4000 pps). It results a false positive decision that FlowVisor cannot accurately control the remote switches.

As the network system becomes more complex, it is not sufficient to rely on a single type of data to cope with the network problems. The network health management system should have a data fusion facility that can collect various types of data and efficiently correlate the data in real-time. As illustrated in figure 2, we propose to use

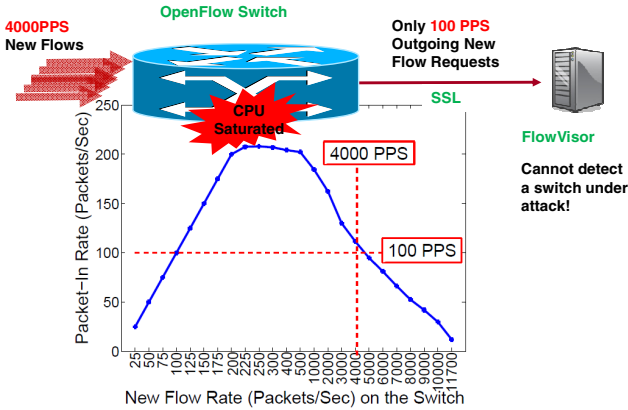


Fig. 1. Remote Controller Fails to Detect New Flow Attack

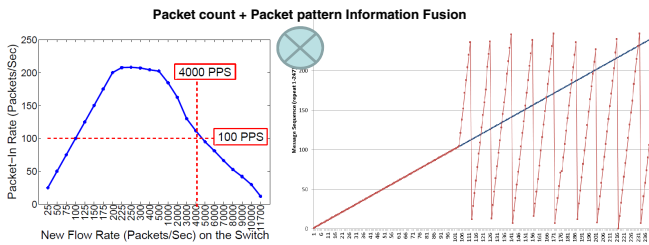


Fig. 2. Example of Network Health Data Fusion

the control data patterns (packet jitter and variation information) in addition to the packet count. Although a remote controller such as FlowVisor may not be able to detect the remote switch problem by only counting the incoming requests, a data fusion facility can make a different decision by accounting the incoming data patterns. For the same incoming packet count, the saturated switch presents very different incoming data patterns.

3 The Case of Event Storm

The network objects may have a containment relationship among each other. For example, as illustrated in Figure 3, a line card object contains many physical interfaces (ports). In turn, each physical interface also contains many logical or virtual interfaces. If objects are in a containment relationship, a status change on an object causes status changes on all the objects it contains. It, in turn, may produce intensive status change notifications to cause an event storm. For example, a port failure may trigger thousands of logical interface failure events. With thousands of event notifications, it may cause tremendous overheads on the switch itself as well as the network and management services. If the event storm is not handled properly, it may cause significant problems

in the OpenFlow network due to the related overheads as well as the root cause of the problems. If a remote controller has a data fusion facility in support of the object containment relationship, the decision process can be fast and simplified and the processing system overhead can be greatly reduced.

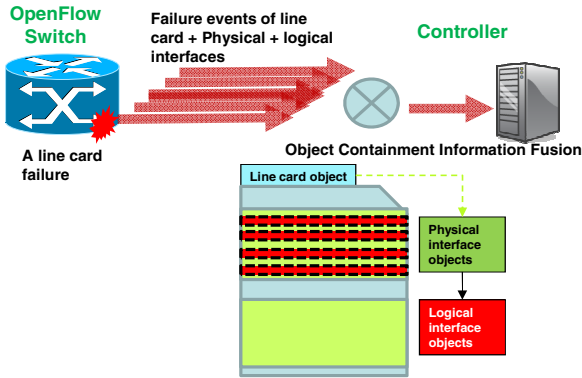


Fig. 3. Object Containment Relationship Data Fusion

4 Conclusion

The traditional networks, as well as the proposed SDN, mainly take remote approaches to monitor network problems. In the proposed data fusion case studies for the SDN architecture, we have shown that they are ineffective and vulnerable to various network problems including new flow attacks and event storm. We have developed a couple of data fusion strategies for detecting network problems and shown the effectiveness of using the data fusion facility compared to the plain FlowVisor environment.

References

- [1] Song, S., Huang, J.: Internet Router Outage Measurement: An Embedded Approach. In: Proceedings of the IEEE/IFIP NOMS 2004, Seoul, Korea (April 2004)
- [2] The OpenFlow Switch Consortium, <http://www.openflowswitch.org>
- [3] Sherwood, R., Gibb, G., Yap, K.-K., Appenzeller, G., Casado, M., McKeown, N., Parulkar, G.: Can the production network be the testbed? In: Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation, OSDI 2010, pp. 1–6. USENIX Association (2010)

Improving Diagnostic Accuracy Using Multiparameter Patient Monitoring Based on Data Fusion in the Cloud

Zhanpeng Jin^{1,2}, Xiaoliang Wang¹, Qiong Gui¹, and Bingwei Liu¹, and Sejun Song³

¹ Department of Electrical and Computer Engineering
Binghamton University, State University of New York
Binghamton, NY 13902-6000

² Department of Bioengineering
Binghamton University, State University of New York
Binghamton, NY 13902-6000

{zjin, xwang90, qgui1, bliu11}@binghamton.edu

³ Department of Engineering Technology and Industrial Distribution
Texas A&M University
College Station, TX 77843-3367
song@entc.tamu.edu

Abstract. Accurate clinical decision making in medical monitoring relies on the strategic fusion of multiparameter physiological signals and usually demands a wide variety of complex machine learning approaches and a large set of knowledge data-base. However, those requirements impose great challenges on computing and storage capabilities, which make it impossible to execute on a single portable computing platform. Leveraging emerging cloud computing technologies, we propose to strategically manage the workloads on the mobile medical monitoring device and migrate the highly intricate multiparameter data fusion and training procedure to the cloud. The mobile device transmits all sensing data acquired from wearable body sensors to the cloud, which now provides a large pool of easily accessible dataset for the training procedures. The well-trained configurations will be sent back to the mobile device and update its existing machine learning based implementations.

1 Status and Challenges of Medical Monitoring

The skyrocketing healthcare expenditure and the fast aging population impose serious economic burdens to the entire society, while people has constantly pursued high-quality medical services. All of these challenges highlight the needs for innovative solutions supporting more accurate, affordable, flexible, and reliable medical diagnosis and treatment. A critical and costly part of the existing healthcare systems is the monitoring of patients' vital signs and other physiological signals, which play significant roles in facilitating physicians' diagnostic practices and tracking people's daily health status.

Most of diseases are not a single symptom, but rather a grouping of signs reflected from highly inter-correlated physiological measures. Current medical monitoring systems often act in an isolated fashion and trigger an alarm when a specific

physiological measurement shows abnormalities, without any dependence on other related signals or an individual's prior medical information [1]. Consequently, the overall health state of a specific person cannot be accurately represented and an extremely high rate of false alarms is reported in clinical settings, e.g., up to 90% of alarms are false positives and the vast majority of alarms have no real clinical impact [2]. Thus a main challenge in medical monitoring is to build a new fusion system involving heterogeneous electro-physiological data for improving the detection of patient states. Given the extracted spatial, spectral, and statistical physiological features, the whole society has been investigating novel approaches that allow synergistic combination of a number of diverse medical parameters, according to the personalized and context-dependent multi-criteria evaluation. The ultimate goal is to compile a unified Health Index (HI) [3] to produce an accurate overall picture of individual's health status.

In medical decision making, data fusion consists of combining data, reducing its complexity and designing a synthetic representation to be more easily interpreted. This requires the integration of spatially or temporally distributed information-gathering systems of high levels of abstractions. In principle, fusion of multiparameter physiological data provides significant advantages over single-parameter data, especially for highly intricate medical diseases, which usually involve the correlated rhythms and changes of a set of physiological behaviors. In addition to the statistical advantage gained by combining same-source data (e.g., obtaining an improved estimate via redundant observations), the use of multiple types of physiological information may increase the accuracy with which a quantity can be characterized. For instance, multi-channel ECG signals accompanied with respiratory rate, blood pressure, and SpO₂ can be used to provide a more accurate picture about the health status of a cardiovascular patient, although approximate diagnostic assessment can be performed using only selected ECG channels.

Observational physiological data may be fused at, from the feature level to the decision level. Feature-level fusion involves the extraction of representative features. An example is the use of statistical metrics (e.g., mean and correlation coefficient) and spectral metrics (e.g., frequency and power spectral density (PSD)) to represent a certain type of physiological measurement. Such features extracted from multiple medical observations will be combined into a single concatenated feature vector which is input to standard pattern recognition approaches. A higher level fusion is also applicable even after each physiological activity has been used to deduce a preliminary determination of an individual's medical condition. Diagnostic results based on one physiological signal may be either consistent, irrelevant, or contradictory to the behavior revealed from another medical measurement. Thus a synergistic information fusion is imperative at a decision level.

2 Opportunities of Data Fusion in the Cloud

Multi-parameter monitoring has been widely used in today's clinical diagnostic processes, including various bedside monitors in hospitals or ambulatory devices used for personalized healthcare. Unfortunately, existing computer-assisted diagnostic

analysis based on monitored signals still remains at a rather simple and intuitive level. Besides being visually screened and evaluated by physicians or nurses, most of patient monitors rely on some certain “calling criteria” to activate the alarms, which can be as simple as a few fixed threshold values, or a systematic mechanism involving all monitored variables. Generally, these criteria can be referred to as “physiological track and trigger warning (TTW) systems” [4]. TTW systems can be further categorized as single-parameter, multi-parameter, aggregate weighted scoring or combination systems [4]. The simplest one, single-parameter systems [5], raise alarms if the extreme abnormality observed in any one of monitored signal. In contrast, multi-parameter systems only indicate warnings if extreme observational values occur in two or more signals. Aggregate weighted scoring systems [6] allocate points to each vital sign variable and trigger the alarms according to the sum of the allocated points, namely Early Warning Scores (EWS) [7]. The combination TT warning systems integrate the elements of single-/multi-parameter systems with aggregate weighted scoring.

Recently, sophisticated machine learning techniques have been extensively investigated for higher-level medical data fusion and used in identifying imminent medical problems by automatically recognizing the abnormal behaviors from a huge amount of physiological signal data, such as fuzzy inference systems, artificial neural networks (ANNs), support vector machines (SVMs), and Bayesian Networks. For instance, according to our earlier studies, the ANN-based electro-cardiogram (ECG) signal processing algorithm can achieve over 95% accuracy in detecting various cardiac arrhythmia [8], and the SVM-based multiparameter vital sign analysis can significantly reduce the amount of false alarms [9].

However, to achieve a satisfactory training performance, those machine learning based clinical decision support approaches usually demand a large set of *a priori* knowledge as the training dataset and iteratively perform the computation-intensive training processes, which make it impossible and infeasible to execute on a single portable computing platform in the ambulatory setting that is increasingly demanded from the perspective of pervasive healthcare. For example, recent advances in wearable body sensors and mobile computing technologies have enabled and promoted the use of mobile-based health monitoring and alert systems (usually referred as “mHealth”), aiming at providing real-time feedback about an individual’s health condition to either the user or to a medical center, while alerting in case of possibly imminent health-threatening conditions. However, the limited computational power and battery life of existing mobile devices, significantly limit their ability to execute resource-intensive applications. Emerging cloud computing provides an alternative to facilitate the conventional clinical decision support systems and to transform the way how future healthcare will be practiced and delivered in a more effective and efficient manner. Some prior studies have been conducted to explore the possible use of cloud computing in healthcare [10].

To address the increasing demands of more sustainable use of mobile devices and more accurate diagnostic decision-making in medical monitoring, we propose to strategically manage the workloads on the mobile devices and to migrate the highly intricate multiparameter data fusion and the supervised training procedure to the cloud. The mobile devices will also transmit all sensing data acquired from wearable body

sensors or portable physiological monitors to cloud storage, which now can provide a large pool of easily accessible dataset for the supervised training procedures. However, the operations of the machine learning algorithms deployed on mobile devices will continue their regular classification processing without any halt, based on the latest trained configurations. Once the supervised training on the cloud is finished, the well-trained configurations will be sent back to the mobile devices and update the existing implementations of the machine learning algorithms on the mobile devices.

In addition to the substantially reduced power consumption and extended battery life of mobile devices, this bidirectional, dynamic workload balancing and migration approach can constantly improve the performance of the deployed machine learning techniques by regularly updating them according to the most recent training results. As new sensing data of a subject is continuously processed on the mobile devices and backed up on the cloud, the whole system holds the potential to gradually evolve itself toward an even higher diagnostic accuracy through unrelenting individual-specific training and adaptation. However, though the synergistic combination of cloud computing and data fusion has shown great promise in transform future clinical decision support systems, there still are some research questions regarding how to determine the optimal interval for launching a new training process and accordingly ensure a balanced tradeoff between the diagnostic accuracy and the computation efficiency, as well as how to avoid the over-fitting issue in the iterative training procedures.

References

- [1] Clifford, G., et al.: Robust parameter extraction for decision support using multimodal intensive care data. *Phil. Trans. R. Soc. A* 367(1887), 411–429 (2009)
- [2] Imhoff, M., Kuhls, S.: Alarm algorithms in critical care Monitoring. *Anesthesia & Analgesia* 102(5), 1525–1537 (2006)
- [3] Alemzadeh, H., et al.: An embedded reconfigurable architecture for patient-specific multiparameter medical monitoring. In: *Proc. of the EMBC*, pp. 1896–1900 (2011)
- [4] DH and Modernisation Agency, Critical care outreach 2003: Progress in developing services, Department of Health, United Kingdom (2003)
- [5] Smith, G., et al.: A review, and performance evaluation, of single-parameter ‘track and trigger’ systems. *Resuscitation* 79(1), 11–21 (2008)
- [6] Smith, G., et al.: Review and performance evaluation of aggregate weighted ‘track and trigger’ systems. *Resuscitation* 77(2), 170–179 (2008)
- [7] Kyriacos, U., et al.: Monitoring vital signs using early warning scoring systems: a review of the literature. *J. Nursing Management* 19(3), 311–330 (2011)
- [8] Jin, Z., et al.: Predicting cardiovascular disease from real-time ECG monitoring: An adaptive machine learning approach on a cell phone. In: *Proc. of the EMBC* (2009)
- [9] Wang, X., et al.: Leveraging mobile cloud for telemedicine: a performance study in medical monitoring. In: *Proc. of the NEBEC* (2013)
- [10] Shen, C.-P., et al.: Bio-signal analysis system design with support vector machine based on cloud computing service architecture. In: *Proc. of the EMBC*, pp. 1421–1424 (2010)

Author Index

- Abbas, Haider 285, 297
Abdul-Kareem, Sameem 163
Abdullah, Mohd. Taufik 277
Abdullah, Zaridah 231
Aboutajdine, Driss 77
Abubakar, Adamau 163
Ahmed, Ishtiaq 197
Ahn, Sang Min 413
Akbiil, Brahim 77
Alghaffi, Khawla 329
Alghathbar, Khaled S. 277
Ali, Qasim 285
Al-Rashid, Nalisa Alia Amin 231
Assar, Saïd 285
- Bai, Cuixia 35
Behl, Sanjiv 151
Blasch, Erik 455
- Chang, Hsi-Ya 43
Chang, Ling-Hua 151
Chen, Dongxiang 421
Chen, Genshe 455
Chen, Hung-Yi 343
Chen, Mu-Song 343
Chen, Xiaojun 323
Chen, Yu 455
Chen, Yuan-Joan 343
Chiroma, Haruna 163
Cho, Kyungeun 7
Cho, Seongjae 7
Choi, Sang-Hei 399
Chow, K.P. 323, 441
Chung, Che-Lun 337
Chung, Tae Choong 197
- Deghantanha, Ali 277
Ding, Qiufeng 259
Doi, Hiroshi 139
Du, Kyeong Hui 383
Duan, Wei 245
- Encheva, Sylvia 29, 93, 107
- Fadzil Hassan, M. 55
Fan, Song Yong 337
Ferchichi, Abdelwaheb 185
- Gang, Zhou 441
Gantsou, Dhavy 145
Ghafoor, Abdul 133
Gouider, Mohamed Salah 185
Guan, Donghai 197
Gui, Qiong 473
- Halboob, Waleed 277
Han, Jong Wook 99
Hani, Ahmad Fadzil M. 55
Ho, Tze-Yee 343
Hong, Il young 315
Hong, Ki-sung 391, 413
Hou, Xuelong 421
Huang, Frank Fu-Yuan 351
Huang, Kuo-Chan 43
Huang, Tony Cheng-Kui 171
Huang, Tse-Chi 43
Hwang, Jun-oh 375
Hwang, Junsik 449
- Im, Hyeonu 399

- J., Divya Udayan 205
 James, Joshua I. 361
 Jang, Yunsik Jake 361
 Jeong, Chang Won 1
 Jeong, Young Sik 1
 Jiang, Gangyi 35, 127
 Jin, Xianli 259
 Jin, Zhanpeng 473
 Jones, Andrew 329
 Joo, Su Chong 1
 Jun, Zhang 441
 Jung, Hyosook 449

 Kang, Won Min 237
 Kao, Da-Yu 351
 Kechadi, Tahar 307
 Kim, Hyoyoung 213
 Kim, HyungSeok 205
 Kim, Hyunwoo 367
 Kim, Jee-In 205
 Kim, JeongNyeo 99
 Kim, Kyungbaek 225
 Kim, Mihyung 391
 Kim, Sung Min 85
 Kook, Sang-Ho 405
 Koong, Chorng-Shiuh 63, 157

 Laskar, Md. Nasir Uddin 197
 Latif, Rabia 285
 Lee, Changhoon 315
 Lee, Chulung 367, 383, 391, 399, 413
 Lee, Jinseok 1
 Lee, Jun 205
 Lee, Kilhung 49
 Lee, Ki Won 237
 Lee, Sangjin 315
 Lee, Seung Chul 85
 Lin, Yu Min 337
 Liu, Bingwei 455, 473
 Liu, Bo 21
 Liu, Peipeng 323
 Liu, Rujuan 177
 Liu, Weidong 69
 Lu, Zhang 429
 Luo, Ting 127

 Ma, Li 245
 Mahmud, Ramlan 277
 Martin, Thomas 329
 Masood, Rahat 133

 Mohri, Masami 113, 119, 139
 Mubarak, Khalid 269

 Namgung, Jaeung 315
 Naruse, Takeru 119
 Naveed, Rida 297
 Niu, Zhendong 177

 Ou, Chin-Chih 151

 Paputungan, Irving Vitra 55
 Park, Jin Wan 213
 Park, Ji Soo 237
 Park, Jong Hyuk 237
 Park, Jungheum 315
 Park, Kyoungju 219
 Park, Seongbin 449
 Park, Sung Yun 85
 Park, Su-wan 99
 Peizhang, Xie 435
 Peng, Cheng-Yu 351
 Peng, Zongju 35, 127
 Pham, Khanh 455

 Qin, Rong 421

 Saa'id, Shahn'il Asmar 231
 Sato, Makoto 139
 Scanlon, Mark 307
 Seo, Jun Seok 85
 Shao, Feng 35, 127
 Shao, Zehui 259
 Shemali, Mouza Ahmed Bani 269
 Shen, Dan 455
 Shi, Jinqiao 323
 Shibli, Muhammad Awais 133
 Shieh, Tung-Ho 151
 Shim, Byonghyo 13
 Shin, Sang-I 219
 Shiraishi, Yoshiaki 113, 119, 139
 Sim, Sungdae 7
 Son-di, Patrick 145
 Song, Jiaxing 69
 Song, Sejun 469, 473
 Song, Wei 7
 Sun, Guozi 251, 259

 Tomida, Koji 113
 Tsai, Mu-Jung 43
 Tseng, Chien-Chao 63, 157

- Udzir, Nur Izura 277
Um, Kyhyun 7
- Wang, Fangju 191
Wang, Jian 13
Wang, Shih-Jeng 351
Wang, Xiaodong 21
Wang, Xiaoliang 473
Wang, Yigang 35, 127
Won, Chee Sun 7
Woo, Jongwook 49
- Xia, Xiao 21
Xingpeng, Zhou 435
Xu, Fei 323
Xu, Kuo 13
- Yang, Chyuan-Huei Thomas 337
Yang, Tzu-I 63, 157
Yang, Yitao 251
Yeun, Chan Yeob 269
Yin, Feiran 69
Yonghao, Mai 441
Yoo, Sujin 449
Yu, Mei 35, 127
- Zemerly, Mohamed Jamal 269
Zhengwu, Lu 441
Zhou, Xingming 21
Zhou, Yanjie 245
Zu, Zhiyue 251