

Brajesh Kumar Kaushik
Jan Zizka (Eds)

Computer Science & Information Technology

Fourth International Conference on Computer Networks &
Communications (CCNET 2017)
Dubai, UAE, April 29~30, 2017



AIRCC Publishing Corporation

Volume Editors

Brajesh Kumar Kaushik,
IIT-Roorkee, India
E-mail: bkkaushik23@gmail.com

Jan Zizka,
Mendel University in Brno, Czech Republic
E-mail: zizka.jan@gmail.com

ISSN: 2231 - 5403
ISBN: 978-1-921987-65-6
DOI : 10.5121/csit.2017.70501 - 10.5121/csit.2017.70509

This work is subject to copyright. All rights are reserved, whether whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the International Copyright Law and permission for use must always be obtained from Academy & Industry Research Collaboration Center. Violations are liable to prosecution under the International Copyright Law.

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

The Fourth International Conference on Computer Networks & Communications (CCNET 2017) was held in Dubai, UAE, during April 29~30, 2017. The Fourth International Conference on Signal Processing (CSIP 2017), The Fifth International Conference of Soft Computing (SCOM 2017) and The Fifth International Conference on Database and Data Mining (DBDM 2017) was collocated with The Fourth International Conference on Computer Networks & Communications (CCNET 2017). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CCNET-2017, CSIP-2017, SCOM-2017, DBDM-2017 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, CCNET-2017, CSIP-2017, SCOM-2017, DBDM-2017 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CCNET-2017, CSIP-2017, SCOM-2017, DBDM-2017.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Brajesh Kumar Kaushik
Jan Zizka

Organization

General Chair

Natarajan Meghanathan,
Brajesh Kumar Kaushik,

Jackson State University, USA
Indian Institute of Technology - Roorkee, India

Program Committee Members

Abdelghani Ghazdali	Hassan University, Morocco
Abdelilah Hakim	Cadi Ayyad University, Morocco
Aihua Mao	South China University of Technology, China
Amine Laghrib	Faculté des Sciences Beni-Mellal, Morocco
Àngela Nebot	Polytechnic University of Catalonia, Spain
Aouatif Amine	Ibn Tofail university, Morocco
Babar Shah	Zayed University, UAE
Bartholomew Placzek	University of Silesia, Poland
Berbra kamel	Saad Dahlab University Blida, Algeria
Bernardo Almada-Lobo	University of Porto, Portugal
Carlo Sau	Università degli Studi di Cagliari, Italy
Chaolu Feng	Northeastern University, China
Chen Haishan	Nanfeng College of Sun Yat-Sen University, China
Chin-Chen Chang	Feng Chia University, Taiwan
Chongyi Fan	National University of Defense Technology, China
Chuanzong Zhang	Aalborg University, Denmark
Chunshien Li	National Central University, Taiwan
Damian Ruiz	Universidad Politecnica de Valencia, Spain
Daniele Codetta-Raiteri	University of Eastern Piedmont, Italy
Deng Li Miao	China University of Petroleum, China
Duan Keqing	Wuhan Early Warning Academy, China
Elias Aboutanios	University of New South Wales, Australia
Emad Eldin Mohamed	Canadian University Dubai, UAE
Emilio Serrano	Technical University of Madrid, Spain
Fatih korkmaz	Cankiri Karatekin University, Turkey
Fatma Outay	Zayed University DXB, UAE
Guilherme Oliveira	Univates University Center, Brazil
Hamed Tabkhi	University of North Carolina Charlotte, USA
Hamid Alasadi	Basra University, Iraq
Hamzeh Khalili	Universitat Politecnica de Catalunya (UPC), Spain
Hao-En Chueh	Yuanpei University Of Medical Technology, Taiwan
Hari Krishna Garg	National University of Singapore, Singapore
Ismail Shahin	University of Sharjah, UAE
Issac Niwas Swamidoss	Nanyang Technological University, Singapore
Jagadish Nayak	BITS DUBAI,UAE

Jiahua Zhu	National University of Defense Technology, China
Jitender Grover	Maharishi Markandeshwar University, India
Joao Gama	University of Porto, Portugal
Lei ZHANG	University of Surrey, UK
Madya Dr. Yuhanis binti Yusof	Universiti Utara Malaysia, Malaysia
Mahmoud R. Delavar	University of Tehran, Iran
Mamoun Abu Helou	Al Istiqlal University, Palestine
Manish Mishra	University of Gondar, Ethiopia
Marta Ruiz	Polytechnic University of Catalonia, Spain
Masoud Vejdannik	Iran University of Science & Technology, Iran
Mike Turi	California State University-Fullerton, USA
Miloš Brajovic	University of Montenegro, Balkans
Mimoun Hamdi	Ecole Nationale d'Ingenieurs de Tunis , Tunisia
Mohamad Badra	Zayed University, Dubai, UAE
Mohamed Arezki Mellal	M'Hamed Bougara University, Algeria
Mohamed Waleed Fakhr	Arab Academy for Science and Technology, Egypt
Mohammad Amin Shayegan	Islamic Azad University, Iran
Mohammad Reza Jabbarpour Sattari	University of Malaya, Kuala Lumpur
Mohammad Siraj	King Saud University, Saudi Arabia
Nadjia Benblidia	Saad Dahlab University - Blida1, Algeria
Nayeem Ahmad Khan	University Malaysia Sarawak, Malaysia
Paul Honeine	Normandie Universite, France
Qinghua Guo	University of Wollongong, Australia
Riccardo Pecori	eCampus University, Italy
Robert S. H. Istepanian	Imperial College, London
Rocío Pérez de Prado	University of Jaén, Spain
Ronghuo Zheng	University of Texas, United States
Saeed Tavakoli	University of Sistan and Baluchestan, Iran
Salem Hasnaoui	National Engineering School of Tunisi, Tunisia
Sanjay Sharma	University of london, UK
Songling Huang	Tsinghua University, China
Tayebeh Askari	Bam University of Iran, Iran
Thai-Son Nguyen	Tra Vinh University, Vietnam
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Veton Kepuska	Florida Institute of Technology, Australia
Victor Banos	Technical University of Catalonia, Spain
Yang Li	Beihang University, China
Yao-Nan Lien	Asia University, Taiwan
Yingsong Li	Harbin Engineering University, China
Yuan Zhuang	Bluision Inc, USA
Yu-Chen Hu	Providence University, Republic of China
Yue Cao	Northumbria University, UK
Yueying Kao	Chinese Academy of Sciences, China
Zbigniew Banaszak	Koszalin University of Technology, Poland
Zhihui Zhu	Colorado School of Mines, USA

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Networks & Communications Community (NCC)



Digital Signal & Image Processing Community (DSIPC)



Organized By



Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

Fourth International Conference on Computer Networks & Communications (CCNET 2017)

Flooding Attacks Detection of Mobile Agents in IP Networks	01 - 12
<i>Jean Tajer, Mo Adda and Benjamin Aziz</i>	
Dynamic Curative Mechanism for Geographic Routing in Wireless Multimedia Sensor Networks.....	13 - 22
<i>Mohamed Nacer Bouatit and Selma Boumerdassi</i>	
Large-Scale Multi-User MIMO Approach for Wireless Backhaul Based HETNETs.....	43 - 52
<i>Mostafa Hefnawi</i>	
JCWAEED : Joint Channel Assignment and Weighted Average Expected End-To-End Delay Routing Protocol in Wireless Mesh Networks.....	53 - 62
<i>Wang Yi-rong, Wang Yan-ru, Zhang Hao, Liu Kai-ming and Li Nan</i>	
Reducing Frequency of Group Rekeying Operation.....	63 - 72
<i>YunSuk Yeo, Sangwon Hyun and Tai-Myoung Chung</i>	

Fourth International Conference on Signal Processing (CSIP 2017)

Secret Image Transmission Through Mosaic Image.....	23 - 32
<i>Shahanaz N and Greeshma R</i>	
Clustering Hyperspectral Data.....	73 - 80
<i>Arwa Alturki and Ouiem Bchir</i>	

Fifth International Conference of Soft Computing (SCOM 2017)

Managing Security and Compliance Risks of Outsourced IT Projects.....	33 - 42
<i>Moneef Almutairi and Stephen Riddle</i>	

**Fifth International Conference on Database and Data Mining
(DBDM 2017)**

A Process of Link Mining..... 81 - 87
Zakea Il-agure and Hicham Noureddine Itani

FLOODING ATTACKS DETECTION OF MOBILE AGENTS IN IP NETWORKS

Jean Tajer, Mo Adda and Benjamin Aziz

University of Portsmouth, School of Computing, Portsmouth, United Kingdom

ABSTRACT

This paper deals with detection of flooding attacks which are the most common type of Denial of Service (DoS) attacks in a Mobile Agent World. We propose a new framework for the detection of flooding attacks by integrating Divergence measures over Sketch data structure. The performance of the proposed framework is investigated in terms of detection probability and false alarm ratio. We focus on tuning the parameter of Divergence Measures to optimize the performance. We conduct performance analysis over publicly available real IP traces, in Mobile Agent Network, integrated with flooding attacks. Our analysis results prove that our proposed algorithm outperforms the existing solutions.

KEYWORDS

Mobile Agents, SYN flooding, Hellinger Distance, Chi-square, Sketch Technique, IP Networks

1. INTRODUCTION

Multi-Agent Systems (MAS) are designed using independent, autonomous known as agents which can perform their tasks independently or collectively in different types of environments [2]. The agents can be considered as processes with the ability to perform an action on the environment on behalf of user [32]. These systems allow distribution of complex tasks amongst agents. One of the basic properties of multi-agent system is its ability of self-organization which makes it utterly desirable for autonomous and flexible system designs such as graphical applications, logistics, transportation, search engines, network management etc [33].

Mobile Agent Systems can be divided based into programming language by which they are developed and use: Java and non-Java based. Around 85% of Mobile Agent systems available today are built using Java, due to its inherent support to Mobile Agent programming [19].

Mobile Agents are becoming a focus of modern research because of their applications in distributed systems which are replacing traditional client-server architectures rapidly [34]. However, one of the key concerns in practical implementation of Mobile Agent is the lack of protection against any threats.

The rest of this paper is organized as follows. Related work is provided in Section 2. Section 3 provides the security issues that a Mobile Agent can counter while visiting another host in the network. We will discuss Sketch data structure to provide grained analysis and to derive probability distributions and will introduce different divergence measures (Hellinger Distance and Chi-square divergence) in order to compare their performance if a flooding attack happens on

a Mobile Agent Network, in Section 4. Section 5 describes our proposed approach design. In Section 6, we present our experimental works and check the capability, reaction and performance of the mobile agents based on the developed design. Finally in Section 7, we present the conclusion and our future work.

2. RELATED WORK

From one side several researches have been proposed security solutions to detect and prevent attacks in real traffic. Most of these proposed solutions emphasize on many different detection and prevention strategies.

SYN flooding attack detection has been an interested issue for security researchers. The authors in [2] present the effects of correlation analysis on the DDoS detection. They propose a covariance analysis method for detecting SYN flooding attacks.

Existing methods for anomaly detection are based on different techniques, such as Haar-wavelet analysis [3], [4], Entropy based method [5] and Holt-Winters [6] seasonal forecasting method. Authors in [7] compare two different algorithms (CUSUM and adaptive threshold) for the detection of SYN flooding attack. They conclude that CUSUM performs better than adaptive threshold in terms of detection accuracy of low intensity attacks. However, both of these algorithms face problems of false alarm ratio under normal IP traffic variation.

Other work aggregates the whole traffic in one time series, and applies a change point detection algorithm to detect the instant of anomaly occurrence. The latter has a good performance in terms of spatial and temporal complexities, but presents the drawback of aggregating all traffic in one flow, where low intensity attacks cannot be detected. Furthermore, these methods use static threshold for detecting anomalies, which is not adequate with traffic variations, and may induce false alarm and miss detection.

Sketch data structure uses the random aggregation for more grained analysis than aggregating the whole traffic in one time series. It has been used to summarize monitored traffic in a fixed memory, and to provide scalable input for time series analysis. Authors in [8] propose the use of CUMulative SUM (CUSUM) over the sketch for network anomaly detection. Furthermore, they propose a new mechanism for Sketch inversion and malicious flows identification. We will exploit the Sketch data structure to derive probability distributions.

In addition, recent work experiments the histogram-based detector in order to detect the anomaly behaviors and changes in traffic distributions [9]. They apply Kullback-Leibler divergence between the current and previous measurement distributions.

Authors in [10] apply Hellinger distance (HD) on Sketch data structure, in order to detect divergence between current and previous distributions of the number of SIP INVITE request. In fact, HD must be near zero when probability distributions are similar, and it increases up to one whenever the distributions diverges (e.g. under Invite flooding attacks). In addition, they used the dynamic threshold proposed in [11] during their experimental analysis.

From other side, several researches have been conducted over mobile Agents.

Some Articles showed what exactly it is makes Java such a powerful tool for mobile agent development, also it highlighted some shortcomings in Java language systems that have implications for the conceptual design and use of Java-based mobile agent systems [19],[20].Some studies concentrate their work on the fault tolerance techniques in mobile agents,

network management applications based on mobile agent technologies and how the fault tolerance techniques can improve their performance [25], [26].

Other articles worked on an agent-based intelligent mobile assistant for supporting users prior to and during the execution of their tasks [27].

In addition, some works have been performed to integrate the mobile agents with the e-commerce. Some technical relevant issues are well presented [28], [29], [30].

Some researches concentrated their work on security concerns (i.e masquerading, denial of service, unauthorized access and repudiation) of mobile agents and how to protect them by several techniques like for example providing logical framework designed to support large-scale heterogeneous mobile agent applications, on safe code interpretation, digital signatures, path histories, State Appraisal and Proof-Carrying Code (PCC) [21], [22],[23],[24]

Our research combines the mobile agents and the detection methods. It emphasizes on how a mobile agent can detect a flooding attacks. We develop a general framework that increases the detection accuracy and reduces the false alarm by integrating different divergence measures over Sketch technique in a Mobile Agent world.

3. MOBILE AGENTS SECURITY THREATS AND COUNTER MEASURES

Security is one of the key factors of MAS. In fact, a MA is one of the potential threats to computer systems and vice versa, from the host system to the MAS itself. In this part, we will talk about the main security issues related to MAS.

The security threats for MASs could be divided as follows:

- **IP spoofing:** consists of sending packets with a faked IP source address. The server should believe that the packets come from another host, probably a host that is allowed to establish connections with the attacked host, if the real one is not allowed
- **Sniffing:** it is the observation and analysis of network traffic in order to obtain relevant information (such as IP addresses and host functionalities) to perform other attacks.
- **UDP flood attack:** this kind of flooding attack consist on sending many UDP packets to different port of a target in random way. This target will check if there's any application on the relevant port, if not, he will be occupied to send ICMP replies and can't treat requests from legitimate clients.
- **SYN flood attack:** it consist on sending many TCP connection requests to a target. This latter will accept the establishment of the connection and notify the client. Except that, this one will never use them. Thereby, the server will be drown by unused connections and, eventually, will not reply to legitimate users requests.

There are many security services that can be used for securing the agents systems, for example; authentication, integrity, confidentiality and authorization.

In case of the authentication, the host needs to know the sender of the delivered agent. The agent authentication process includes verifying the entity that programmed the agent and also verifying the entity that dispatched it to the host. Basically, the agent and the host need to know with whom they are talking and dealing with, here the public-key encryption or passwords can be used,.

For integrity, checking the integrity of the agents is a technique that makes sure no one has made any changes to the agents, the agents travelling from host to another, and communicates and exchanges their data with other hosts and other agents. In this case, we need to make sure that the agents have not been tampered with in relation to their state, code or data. Moreover, the agents could carry different types of data, for example some private data. These data should only be readable from a specific host or agents. This technique is very important to avoid an eavesdropping threat. The last service which helps to protect the agents and the hosts is authorization; the incoming agents should have a specific right to access the host information, so different agents have different authority, to protect the hosts and also to protect themselves.

4. THEORITICAL BACKGROUND

4.1. Sketch technique

In this section, we review the K-ary Sketch data structure. Using Sketch data structure makes our framework flexible and scalable for grained analysis. No matter how many flows exist in the traffic, Sketch generates fixed-number of time series [3], [4] for anomaly detection. Sketch provides more grained analysis than aggregating whole traffic in one time series.

The Sketch data structure is used for dimensionality reduction. It is based on random aggregation of traffic attribute (e.g. number of packets) in different hash tables. A Sketch S is a 2D array of $H \times K$ cell (as shown in Figure 1), where K is the size of the hash table, and H is the number of mutual independent hash functions (universal hash functions). Each item is identified by a key κ_n and associated with a reward value v_n . For each new arriving item (κ_n, v_n) , the associated value will be added to the cell $S[i][j]$, where i is an index used to represent the hash function associated with i th hash table ($0 \leq i \leq d - 1$), and j is the hash value ($j = h_i(\kappa_n)$) of the key by the i th hash function.

Data items, whose keys are hashed to the same value, will be aggregated in the same cell in the hash table, and their values will be added up to existing counter in the hash table. Each hash table (or each row) is used to derive probability distribution as the ratio of the counter in each cell to the sum of whole cell in the line. The derived probability distributions (we get K probability set, one per line) are used as inputs for divergence measures

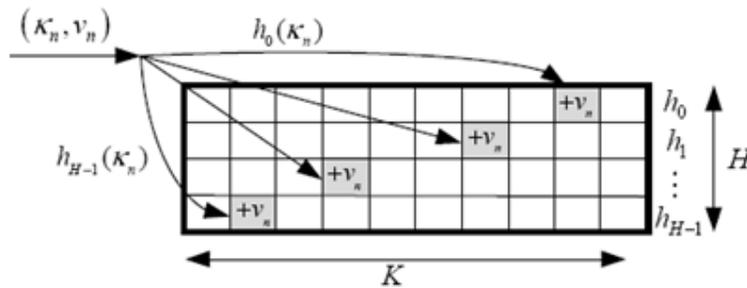


Figure 1: Sketch Data structure

4.2 Divergence Measures

These measures are used to detect the DDoS attacks based on the deviation of traffic distribution. In fact, the idea is to compare the prior distribution derived from Sketch counters in previous time slot, with the currently obtained distribution. One can use this change to detect flooding attack, because the counter of one cell will increase significantly with the number of sent requests, and the probability distribution deviates at the start and stop instants of the flooding attack.

4.2.1. Hellinger Distance (HD)

Hellinger Distance (HD) is used to measure the divergence between two sets of probability values.

For two discrete probability distributions $P = (P_0, P_1, \dots, P_{k-1})$ and $Q = (Q_0, Q_1, \dots, Q_{k-1})$, with $P_i \geq 0, Q_i \geq 0$ and

$$\sum_{i=0}^{k-1} p_i = \sum_{i=0}^{k-1} q_i = 1$$

The HD between current distribution P and prior distribution Q is defined as:

$$HD(P, Q) = \frac{1}{2} \sum_{i=0}^{k-1} (\sqrt{p_i} - \sqrt{q_i})^2 \quad (1)$$

Where HD satisfies the inequality $0 \leq HD(P, Q) \leq 1$, and $HD(P, Q) = 0$ if $P = Q$. HD is a symmetric distance (e.g. $HD(P, Q) = HD(Q, P)$), and induces two spikes, one at the beginning of change, and the second at the end of the change, [18].

4.2.2. Chi-square divergence

χ^2 divergence is used to measure distance between two discrete probability distributions (P and Q). For 2 probability sets $P = (p_1, p_2, p_3, \dots, p_n)$ and $Q = (q_1, q_2, q_3, \dots, q_n)$, with $P_i \geq 0, Q_i \geq 0$ & $\sum_{i=0}^{k-1} p_i = \sum_{i=0}^{k-1} q_i = 1$

The Pearson χ^2 divergence between P and Q is given by:

$$x^2(P||Q) = \sum_{i=1}^n \frac{(P_i - Q_i)^2}{Q_i} \quad (2)$$

Where Q is the estimated probability distribution and P is the measured probability distribution, and $\chi^2(P||Q)$ is the distance between distributions P and Q .

For hypothesis testing, such as H_0 (normal traffic hypothesis) and H_1 (traffic with anomalies), χ^2 values can run from zero into infinity. χ^2 will be zero if P and Q are identical ($P_i = Q_i$) under hypothesis H_0 , and χ^2 increases as the distributions become dissimilar, and eventually so high (infinity) when the two distributions are independent ($P \neq Q$) under hypothesis H_1 . It is important to note that χ^2 divergence is nonnegative and the division $0/0$ is treated as 0, and the division by zero is replaced by a very small value ϵ .

The χ^2 divergence between 2 probability distributions P and Q must be near zero under normal traffic, with a large deviation (one spike) when distributions change occurs. χ^2 is asymmetric ($\chi^2(P||Q) \neq \chi^2(Q||P)$), and its symmetric version raises two spikes. One spike at the beginning and the second at the end of the attack.

$$x^2(P||Q) + x^2(Q||P) = \sum_{i=1}^n \frac{(P_i - Q_i)^2 + (P_i + Q_i)}{P_i * Q_i} \quad (3)$$

We intend to use Pearson chi-square divergence (asymmetric) to detect anomaly through the detection of deviations from normal traffic profile, and we will modify the input time series to constrain χ^2 to raise alarms (spikes) for the whole duration of attack. In [30], authors prove that χ^2 divergence behaves better than all classical divergences (Hellinger distance, Kullback-Leibler, Likelihood, etc, [6].

5. PROPOSED APPROACH

The proposed approach for anomaly detection in Mobile Agent networks is based on Sketch and divergence measures (Hellinger Distance and Chi-square)

The detection system records the number of monitored point (e.g. #packets, #SYN, #flows, etc.) in the Sketch for each discrete time interval T . Random aggregation of traffic flows in Sketch is the first step of our processing, followed by time series forecasting with divergence measures (Figure. 2).

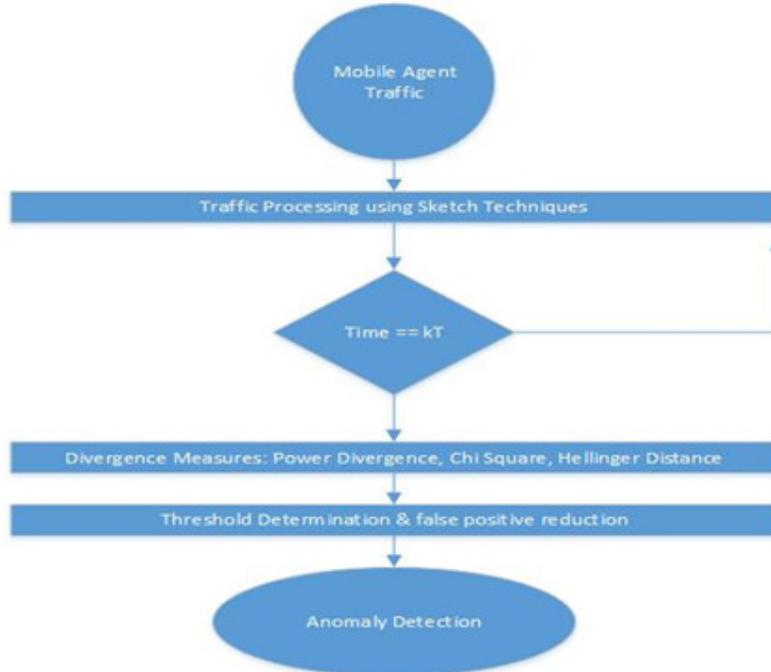


Figure 2 : Architecture of the proposed approach for network anomaly detection.

During each interval, the destination IP address (DIP), for each packet containing a SYN segment, is hashed by H hash functions. The resulted hash value by the i th function ($j = h_i(\text{DIP})$) is used as index of the associated counter $S_{i,j}$ with DIP. Each arriving SYN segment increments the associated counter.

Our analysis will be focused on TCP SYN flooding by counting the number of SYN.

At the end of each epoch T , we derive probability distributions from Sketch. First, we get the sum of the counter in each line, and the probability $p_{i,j}$ in each cell is calculated as the ratio of each counter to the total number of SYN:

$$P_{i,j} = \frac{S_{i,j}.Counter}{\sum_{j=0}^{k-1} S_{i,j}.Counter} \quad (4)$$

Each cell $S_{i,j}$ becomes a data structure, that contains: current counter, current and previous probabilities. Therefore, each line (or hash table) provides two probability distributions: the first one is from previous interval and used as reference distribution Q_i . The second one is from current interval P_i , and used to measure the divergence from the reference distribution, in order to detect anomalies. Divergence measures between the current (P_i) and reference probability (Q_i) distributions is calculated for each line in the Sketch, at the end of each time interval (i.e. at $n.T$).

During malicious activities, the divergence measure $D(P_i||Q_i)$ produces spikes, and when more than L ($L < H$) divergences resulted from different hash tables exceed a dynamic threshold, an alarm is raised.

To detect deviations in the time series resulted from divergence measures, we derive a subsequent time series containing the values of $D(P_i||Q_i)$ without spikes. In this last time series (without large values), we define a dynamic bound of $\mu_i + \alpha\sigma_i$. Significant deviations are larger than the dynamic bound:

$$D(P_i||Q_i) > \mu_i + \alpha\sigma_i \quad (5)$$

Where $D(P_i||Q_i)$ is the divergence measure in the time interval $n.T$ for the i th line in the Sketch, and μ_i & σ_i are the mean and the standard deviation respectively of smoothed time series that doesn't contain spikes ($D^\wedge(P_i||Q_i)$). μ_i and σ_i are updated dynamically using the Exponentially Weighted Moving Average (EWMA):

$$\mu_i = \beta\mu_{(i-1)} + (1-\beta) D^\wedge(P_i||Q_i) \quad (6)$$

$$\sigma_i^2 = \beta\sigma_{(i-1)}^2 + (1-\beta)(D^\wedge(P_i||Q_i) - \mu_i)^2 \quad (7)$$

The threshold is updated dynamically with the value of μ_i and σ_i as shown in above equations. α is a parameter used for calibrating the sensitivity of the detection algorithm to variations. It is also used to reduce the false alarm rate. Under normal traffic, divergence $D(P_i||Q_i)$ falls inside the bound of $\mu_i + 2\sigma_i$. When $D(P_i||Q_i)$ exceeds the dynamically updated threshold over L lines, an alarm is triggered.

6. EXPERIMENTAL WORKS

In this section, we present the performance analysis results for integrating divergence measures over Sketch, for detecting SYN flooding attacks in a mobile agent network. As we want to compare 2 divergence measures (HD & γ^2) over Sketch for the detection of flooding attacks, we will implement a mobile agent network.

For the sake of simplicity, we focus our analysis on the detection of TCP SYN flooding attacks, as it is the widely used attack for DDoS in these days.

6.1 Dataset

The following techniques and tools are used: Two workstations with 8 GB and 768 MB of RAM respectively, which run Windows Server 2003 and a number of Mobile Agents are used.

We have considered the above describe mobile agents will have to execute the similar path. To measure the capability of the proposal towards eavesdropping threat, a test environment is set up using the above mentioned computers as shown in Figure 3. Computer A is considered to act as trusted server (TS) and computer B runs many host nodes simulated through various port numbers as well as the home node in a virtualized mode. Ethereal will be running regularly over computer A. its job is to collect packets in the mobile agent network and store them for a period of 4h00 from 18/02/2017 07h30 to 11h30. These traces are used to test the efficiency of divergence measures. IP addresses in the traces are scrambled by a modified version of tcpdriv tool, but correlation between addresses are conserved. We analyze these 8h30 traces using Sketch data structure, with a key of the Sketch ($\kappa_n = DIP$), and a reward $v_n = 1$ for SYN request only, and $v_n = 0$ otherwise. We set the Sketch width K to 1024, and the number of hash H to 5.

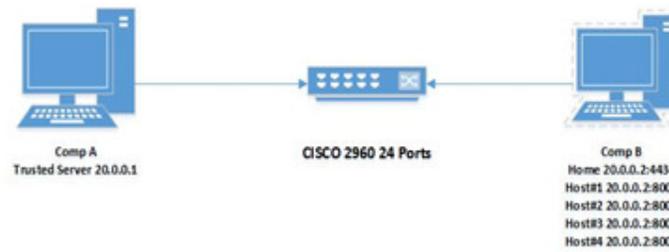


Figure 3: Experimental Lab

Afterward, we inject 12 real distributed SYN flooding attacks with different intensity inside this trace. These attacks are inserted each 30 minutes (on instants $t=30, 61, 90, 127, 157, 187$, etc.) and span for 10 minutes. These different intensity attacks are shown in Figure 4. The first attack begins with a value of 900 SYN/min and decreases until 280 SYN/min.

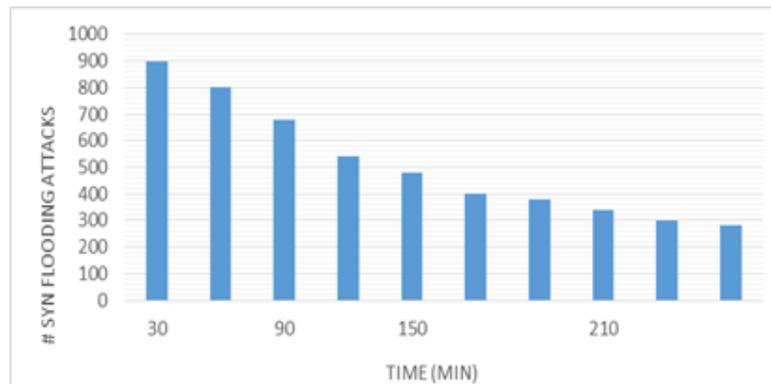


Figure 4 : SYN flooding Attacks

Figure 5 & Figure 6 show the variation of total number of mobile agents' packets before and after the injection of SYN flooding attacks. By comparing these variations, we might not notice the differences between both figures without deep inspection. Inserted attacks don't induce heavy deviations in the time series of the total number of SYN requests. This can be explained by the fact that the intensity of SYN flooding attacks is not large compared to the intensity of the total number of SYN segments. In such cases, the detection of attacks is very challenging, because no heavy changes in the time series describing the variations of the total number of SYN, and the intensity of the SYN flooding attacks is buried by the large number of SYN (as shown in Fig. 4) before attacks injection.

6.2 Evaluation Strategy

In this section, we present the evaluation results of the application of these divergence measures on the mobile agent IP traces.

First, we begin our analysis by applying HD & χ^2 divergence over the traces (before attacks injection). We set the dynamic threshold as given in Eq. 5. We will begin our analysis by applying the HD and Chi-Square over the mobile agent IP traces (before injection SYN flooding attacks). Figure 7 & Figure 8 show the variation of these 2 divergence algorithms as well as the dynamic threshold (dashed line) before the injection of attacks. When the value of divergence measures is larger than threshold in at least 3 hash tables in the Sketch, an alarm is triggered. We see that both algorithms were able to detect anomalies at different time ($t=90, 127, 157, 180$ etc.).

These anomalies are temporary and they don't persist more than many minutes. However, there are more anomalies that can be detected by using the source IP address as the key of the Sketch, but we will restrict our analysis to SYN flooding attacks. In fact, after the manual verification of traces, we found that HD triggers 4 false alarms, and the χ^2 divergence achieves very high detection accuracy with 1 false alarm.

Indeed, we continue our analysis by applying the HD and Chi-Square over the mobile agent IP traces (after injection SYN flooding attacks). We noticed that in case of Hellinger Distance using a dynamic threshold, we obtain 4 false alarms with a detection of 100% (Figure 9). However, in the case of Chi-Square, we did not obtain any false alarm (Figure 10). We found through our conducted experiments that Chi-square divergence performs better than HD in terms of reducing false alarm, with less effort for tuning the dynamic threshold. The intensity of raised spikes in Chi-square increases with the intensity of attacks and dynamic threshold becomes useless.

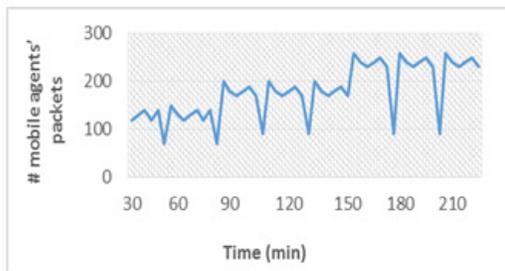


Figure 5 : Total number of mobile agents' packets

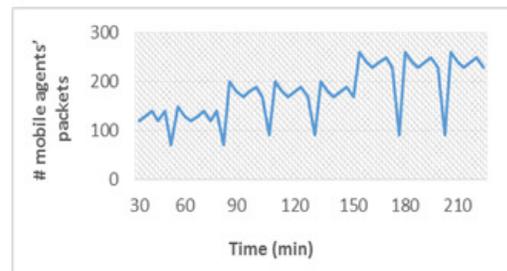


Figure 6 : Total number of mobile agents' packets after SYN flooding attacks injection

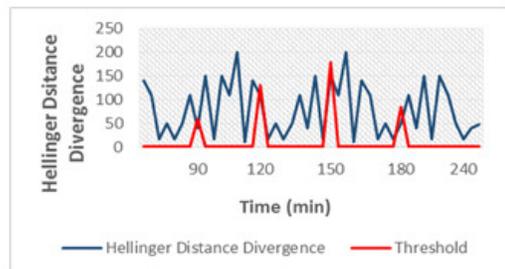


Figure 7 : Hellinger Distance before attacks

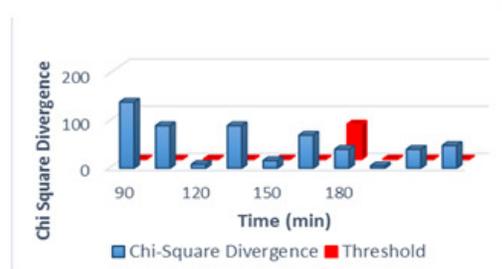


Figure 8 : Chi-square before attacks

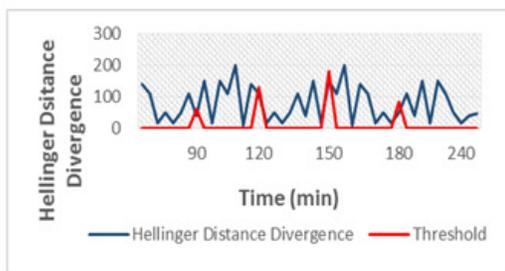


Figure 9 : Hellinger Distance after attacks

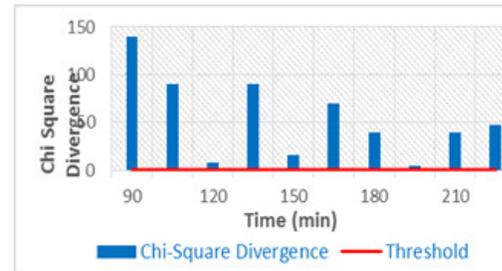


Figure 10 : Chi-square after attacks

7. CONCLUSIONS

In this paper, we analyzed the accuracy of 2 divergence measures (HD & Chi-square divergence) over Sketch data structure for network anomaly detection. We compared their performances in

terms of true positive and false alarm ratio, over real mobile agents IP traces with injected real distributed SYN flooding attacks at known instants.

Afterward, we used dynamic threshold for achieving the best tradeoff between false alarm and true detection.

We found that HD performs a good detection, but with higher false alarm ratio than Chi-square divergence. We can conclude that Chi-square conducts better detection than HD for mobile agents' network. Furthermore, the intensity of triggered spikes by Chi-square divergence increases significantly with the intensity of attacks. It is important to note that these divergence measures with Sketch are computationally efficient for handling traffic on mobile agents' traffic.

In our future work, we will introduce another divergence measure which is Power Divergence in order to compare its performance on the detection of flooding attacks over mobile agents with Chi-square and Hellinger Distance. In addition, we will focus on providing additional information to pinpoint malicious flows, in order to trigger automatic reaction against ongoing attacks. We also intend to provide a method for reducing the amount of monitoring data on mobile agents networks, and to analyze the impact of sampling on the precision of this divergence measure.

REFERENCES

- [1] D. Moore, G. M. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity" in Proceedings of USENIX Security Symposium (SSYM'01), 2001, pp. 9–22.
- [2] HU, Jiang-Ping, Zhi-Xin LIU, Jin-Huan WANG, Lin WANG, Xiao-Ming HU. "Estimation, Intervention and Interaction of Multi-agent Systems." *Acta Automatica Sinica* 39, no. 11 (2013): 1796-1804.
- [3] O. Salem, S. Vaton, and A. Gravey, "A novel approach for anomaly detection over high-speed networks," in Proceedings of the 3rd European Conference on Computer Network Defense (ECND'07), vol. 30, 2009, pp. 49–68.
- [4] G. Cormode and S. Muthukrishnan, "An improved data stream summary: The count-min sketch and its applications," *J. Algorithms*, vol. 55, pp. 29–38, 2004.
- [5] J. Tang, Y. Cheng, and C. Zhou, "Sketch-based sip flooding detection using hellinger distance," in Proceedings of the 28th IEEE conference on Global telecommunications (GLOBECOM'09), 2009, pp. 3380–3385.
- [6] M. Broniatowski and S. Leorato, "An estimation method for the neyman chi-square divergence with application to test of hypotheses," *J. Multivar. Anal.*, pp. 1409–1436, July 2006.
- [7] J. Havrda and F. Chavrat, "Quantification method of classification processes: The concept of structural α -entropy," *Kybernetika*, vol. 3, pp. 30–35, 1967.
- [8] P. N. Rathie and P. Kannappan, "A directed-divergence function of type β ," *Inform. Contr.*, vol. 20, pp. 38–45, 1972.
- [9] D. Haussler and M. Opper, "Mutual information, metric entropy, and cumulative relative entropy risk," *Ann. Statist.*, vol. 25, pp. 2451–2492, 1997.
- [10] "MAWI working group traffic archive," <http://mawi.wide.ad.jp/mawi/>.
- [11] M. Bishop, "Introduction to security network", Addison Wesley, 1 edition, 26 October 2004
- [12] VOIP Security and Privacy Threat Taxonomy, public release, 24 October 2005

- [13] Mohamed Nassar, Saverio Niccolini, Radu State, Holistic “VOIP Intrusion Detection and Prevention System”, ACM SIGCOMM, New York, July 2007.
- [14] Mohamed Nassar, Radu State, and Olivier Fester. “Voip Honeypot Architecture”. In: Integrated Network Management (IM 2007), pages 109-118. IEEE, Munich, May 2007
- [15] V. Jacobson, “Congestion avoidance and control,” SIGCOMM Comput. Commun. Rev., vol. 25, pp. 157–187, January 1995.
- [16] Tascos Dagiuklzd, Jiri Markl, Michal Rokos, low cost tools for secure and highly available voip communication services, snocer 2
- [17] <http://www.webbasedconferencing.org/blog/vishing-spiting-eavesdropping-security-threats-to-voip-primer>
- [18] Hemant Sengar, Duminda Wijesekera, Sushil Jjodia, “Detecting VOIP Floods Using the Hellinger Distance”, IEEE, Vol.19, June 2008
- [19] Danny B. Lange, Mitsuru Oshima. “Mobile Agents with Java: The Aglet API”, September 1998, Volume 1, Issue 3, pp 111–121
- [20] Sun: Java 2 SDK security documentation. (2003).
- [21] Guido J. van 't Noordende, Frances M. T. Brazier, Andrew S. Tanenbaum. “Security in a Mobile Agent System”, 2004, IEEE Symposium on Multi-Agent Security and Survivability
- [22] Michelle S. Wingham, Joni da Silva Fraga, Rafael R. Obelheiro. “A Security Scheme for Agent Platforms in Large-Scale Systems”, 2013, IFIP International Conference on Communications and Multimedia Security Mobile, pp 104-116
- [23] Gray, R., Kotz, D., Cybenko, G., Rus, “Security in a multiple language, mobile agent systems”. LNCS 1419. Springer-Verlag (1998)
- [24] Karnik, N. “Security in Mobile Agent Systems”. PhD thesis, University of Minnesota (1998)
- [25] Maria Zubair, Umar Manzoor. “Mobile Agent based Network Management Applications and Fault-Tolerance Mechanisms”, The Sixth International Conference on Innovative Computing Technology (INTECH 2016)
- [26] Mouhammed Alkasassbeh, Mo Add. “Network fault detection with Wiener filter-based agent”, Journal of Network and Computer Applications 32(4) (4):824-833 • July 2009
- [27] Talal Rahwan, Tarek Rahwan, Iyad Rahwan, and Ronald Ashri. “Agent-based Support for Mobile Users using AgentSpeak(L)”, Agent-Oriented Information Systems Volume 3030 of the series Lecture Notes in Computer Science pp 45-60
- [28] Tu, Griffel and Lamersdof. “Integration of intelligent and mobile agent for E-commerce”
- [29] Ryszard Kowalczyk, Mihaela Ulieru and Rainer Unland. “Integrating Mobile and Intelligent Agents in Advanced e-Commerce: A Survey”, Agent-Oriented Information Systems Volume 3030 of the series Lecture Notes in Computer Science pp 45-60
- [30] Jansen W. and Karygiannis “T. Mobile Agent Security”, National Institute of Standards and Technology, Gaithersburg, MD 220899.
- [31] HU, Jiang-Ping, Zhi-Xin LIU, Jin-Huan WANG "Estimation, Intervention and Interaction of Multi-agent Systems." Acta Automatica Sinica 39, no. 11 (2013): 1796-1804.

- [32] Umar Manzoor, Samia Nefti, Yacine Rezgui “Categorization of malicious behaviors using ontology-based cognitive agents”, *Data & Knowledge Engineering*, Volume 85, May 2013, Pages 40-56.
- [33] Umar Manzoor, Samia Nefti, “iDetect: Content Based Monitoring of Complex Networks using Mobile Agents”, *Applied Soft Computing*, Volume 12, Issue 5, May 2012, Pages 1607-1619.
- [34] Chen, Bo, Harry H. Cheng, and Joe Palen. "Integrating mobile agent technology with multi-agent systems for distributed traffic detection and management systems." *Transportation Research Part C: Emerging Technologies* 17, no. 1 (2009): 1-10.

AUTHORS

Jean TAJER is working as Estimation Unit Head – Low Current at Nesma Trading (KSA). He is a PHD student at University of Portsmouth (UK). My research interests are focused on areas related to security, detection of DDOS attacks over a mobile agents network, Sketch techniques, Divergence measures. He gained my MSC in Communication Network Planning and Management from University of Portsmouth in 2007. Another Master had been gained from University of Paris Sud in 2008. He worked previously at Spie Communication (France) as team leader in Network and Unified Collaboration. I gained several certificates from Cisco, HPE, Avaya, Juniper.



Mo ADDA is a Principal Lecturer at the University of Portsmouth since 2002. He obtained a PhD in distributed systems and parallel processing from the University of Surrey. As a Senior Lecturer, he taught programming, computer architecture and networking for 10 years at the University of Richmond. From 1999-2002, He worked as a senior software engineer developing software and managing projects on simulation and modelling. He have been researching parallel and distributed systems since 1987. His research interests include multithreaded architectures, mobile networks and business process modelling, parallel and distributed processing, wireless networks and sensor networks, network security, simulation and modeling, mobile intelligent agent technology.



Benjamin Aziz is a senior lecturer at the School of Computing, University of Portsmouth. He gained a PhD in Computer Science from Dublin City University in 2003 and since, He has held several post-doctoral research posts in University College Cork, Imperial College London and Rutherford Appleton Laboratory in Oxford. My research in the field of computer and information security spans more than 15 years. In particular, his research interests are focused on areas related to formal analysis of security properties, engineering secure large-scale distributed systems, security requirements at the engineering level, trust management and digital forensic analysis and formalisation. Over the years, he has published over 70 articles, papers, reports and book chapters in these areas. He is a member of several international working groups. He is also an Associate Editor-in-Chief of the *International Journal of Security (IJS)* and an Associate Editor of Wiley's *Security and Communications Networks*.



DYNAMIC CURATIVE MECHANISM FOR GEOGRAPHIC ROUTING IN WIRELESS MULTIMEDIA SENSOR NETWORKS

Mohamed Nacer Bouatit, Selma Boumerdassi

Centre d'Etude et de Recherche en Informatique et Communications /
Conservatoire National des Arts et Métiers - Paris

ABSTRACT

Maintaining network stability and extending network lifetime to cope with breaking links and topology changes remain nowadays a unsolved issues in Wireless Multimedia Sensor Networks (WMSNs), which aim to ensure flow delivery while guaranteeing QoS requirements, particularly, during data transmission phase. Therefore, in this paper, we jointly consider multipath transmission, load balancing and fault tolerance, to enhance the reliability of transmitted data. We propose a Dynamic Curative Mechanism for Geographic Routing in WMSNs. Theoretical results and those obtained from simulation study demonstrate the validity and efficiency of our proposed mechanism, and indicate that it is highly advised for multimedia transmission and network stability.

KEYWORDS

Wireless Multimedia Sensor Networks, Fault-Tolerance, Geographic Multipath Transmission, Load Balancing.

1. INTRODUCTION

Latest advances in wireless networks and MEMS technologies (Micro-Electro-Mechanical-Systems) and embedded systems have led to the creation of Wireless Sensor Networks (WSNs). With a small size and low cost, wireless sensors can be deployed in hundreds over a wide area, to take scalar measurements of physical phenomena, such as temperature, pressure, humidity or the location of objects. In general, these applications are tolerant of delay and do not require a high bandwidth.

The vulgarization of miniaturized technological devices, such as cameras and microphones, has favoured the development of Wireless Multimedia Sensors Networks (WMSNs). Equipped with an on-board camera, these new sensors provide a better description of the observed phenomena by collecting, in addition to scalar data, multimedia content (video and audio streams). On the other hand, they pose new challenges in terms of ensuring QoS parameters, such as bandwidth, data delivery rate, and end-to-end transmission delay, as well as the limited energy of sensor nodes, which consequently reduces network lifetime and causes instability during data transmission phase. This last one is usually subject to ruptures due the unreliability of wireless links and also to topology changes especially in case of nodes' mobility.

In addition to inherited constraints of classical sensor networks (energy, storage and computing), a common problem in such networks is information error, loss caused by components failure, wireless transmission error and external interference [1].

Fault tolerance is considered as an importance domain among various fields of research in wireless sensor networks, due to their constraints such as energy limitations, environment and deployment, where redeployment prohibitive cost, presents a handicap for reorganization network in case of failures of one or more of its components.

Various existing solutions proposed for fault tolerance in WSNs suffer from a poor trade-off between the scalability of the system and the level of tolerance offered by produced topology [2] and they do not cover the issues related to broken communication links and network stability by conserving operational paths during data transmission phase.

In this paper, we propose a Dynamic Curative Mechanism for geographic routing, in order to ensure this trade-off and to meet multimedia transmission constraints:

Our proposed Dynamic Curative Mechanism for geographic routing supports two features:

1. Dynamic load balancing strategy: which dynamically distributes the flow between operational paths without interrupting transmission, in order to prevent packets' early loss and provides better load balancing;
2. Local repair strategy: repair locally broken links by bypassing dynamic holes formed during transmission, due to energy depletion, physical destruction or topology changes.

The remainder of this paper is organized as follows. Section II, discusses fault tolerance procedure's and gives a classification of existing solutions. Section III, presents network model. Proposed DCM is described in section IV. Results of extensive simulations and experimentation are shown in section IV. Finally, section V, concludes the paper.

2. RELATED WORK

Energy limitation and hostile environment in which sensor nodes are deployed, as well as the unreliability of radio communications and loss of wireless connections due to the depletion of sensor node's battery or its physical destruction are factors that increase the risk of breakdowns and make the network vulnerable.

Since physical access to these nodes is often impossible given the intended applications, such as monitoring forest fires in which sensors are scattered by air, or monitoring of urban infrastructure like bridges where sensors are incorporated in the structure. Failures and malfunctions of sensor node, produce data loss and generate topologies changes, which consequently impact network connectivity and reduce its lifetime.

The design of fault-tolerance procedure depends on the architecture and functionality of the system. Nevertheless, some steps are common to most systems [3]:

- Error detection: allows to recognize that an unexpected event occurred. Techniques used are classified into two categories, the offline detection that runs when the system is inactive while the online detection enables the identification of real-time fault and is performed during system activity.

- Damage containment: determines failure impact limits on the system to prevent the spread to other regions.
- Error recovery: two techniques are used, in backward recovery, the system state is restored to an earlier state that is error-free, while in forward recovery, the goal is to reach a consistent state, which is error free.
- Fault treatment: after isolation of failed component, this phase repairs it according to the type of failure.

Fault-tolerant routing algorithms can be divided into two main classes depending on treatment phase: preventive algorithms use fault-tolerant techniques to delay or avoid any errors, while curative algorithms, do not trigger implemented mechanism only when failure is detected.

One can also see fault tolerance from an architectural point of view, which deals different types of component management, namely:

2.1. Battery management

Considered as preventive, it aims to set a uniform distribution of energy dissipation between sensors to better manage energy consumption and increase network lifetime. The Duty-Cycling technique is used to determine the percentage of activity of a node that periodically sleeps to conserve energy. In McTPGF [4], which is an extension of TPGF [5], sleeping-delay is taken into account in routing decision. Its weakness lies in the fact that authors use a single path in their studies and results proved that McTPGF reduces end-to-end delay with the cost of adding few hop counts compared with usual TPGF.

2.2. Flow management

This category includes data transfer management:

- Multipath routing : uses a preventive algorithm to establish multiple paths from the source to the sink. This ensures the presence of more reliable paths for transmission and offers rapid recovery of transfer in case of failures. In [6], a fault-tolerant routing protocol, that modifies conventional DSR [7] protocol, is proposed. It tries to find two routing paths from the source to the destination. Its disadvantage is the use of a single route for transmission, which causes an overload at nodes constituting this path. Additionally, in case of failure, protocol uses secondary route and therefore all packets that have borrowed the faulty path are lost.
- Route recovery: curative technique that creates an alternative new path to ensure the retransmission of data. In [8], to achieve reliability and fault tolerance, proposed protocol continuously updates routes status, while transmitting data along paths simultaneously, it uses a subsequent reliable path, if some unpredicted fault happens in the path. Its drawback is the flooding of network by updating requests and control messages, additionally, the overload of nodes involved in the transmission, which consequently depletes their residual energy.
- Channel allocation: implemented at the MAC layer, this solution performs allocation of transmission channel, in order to reduce interferences between neighboring nodes and avoids packet collisions during transfer. In [9], Channel Utilization and Delay Aware Routing protocol (CUDAR) is proposed that satisfy QoS parameters (throughput, delay and jitter) and reduce energy consumption by using adaptive channel utilization module in MAC layer.
- Mobility: this approach allows to choose a number of mobile nodes, with superior capacity (energy and calculation), which are responsible for collecting data by moving between network sensors. This allows a saving in energy consumption by reducing hops number of transmitted packets. In [10], a pragmatic approach to area coverage in hybrid WSNs is proposed to enhance and maintain the area coverage by moving mobile sensor nodes in the uncovered area. Its main drawback is that only the sink launches the hole detection and recovery process.

2.3. Data management

Solutions in this category offer better data management and processing. Two major subcategories are derived :

- Aggregation: considered as preventive approach, performs processing on the raw data collected from the environment and combines captured data from multiple nodes to reduce the amount of data transmitted across the network and thereby increases its lifetime. In [1], a combination of trust mechanism, data aggregation, and fault tolerance is proposed, to enhance data trustworthiness in WMSNs. It incurs high routing and computational overhead while exchanging trust information. This exchange of information leads to false reporting attack where a malicious node may propagate false information to decrease trust rating for well reputed node [11].
- Clustering: preventive technique (sometimes considered as curative approach) treats structure of sensor networks. It allows to form a virtual backbone for better use of resources such as bandwidth and energy. A cluster-based transmission mechanism with dynamic changes in the path has been proposed in [12]. However, its limitation is that authors assume a secure communication channel and they have not taken into account malicious attacks against trust models [13].

In this article, we focused on a curative technique that uses an optimistic approach, we have also combined two techniques of flow management: multipath routing and route recovery, in order to guarantee QoS and maintain network stability.

3. NETWORK MODEL

We consider a flat network architecture and the wireless sensor network is composed of N sensors, deployed in a static deterministic manner, each sensor node being aware of its geographic location and its 1-hop neighbor nodes' geographic locations. We assume that only source nodes defined for area of interest know the Base Station (BS) location and all other sensor nodes know BS by receiving packets from source nodes. All nodes have the same transmission range and are homogeneous and are endowed with identical physical capabilities (detection and communication). Only bidirectional links are used to build paths. Each sensor node may be in one of the following states:

- Valid: ready to build a path;
- Active: already used in a path (locked for specific path);
- Blocked: no valid next-hop except its predecessor;
- Failed: broken, damaged or exhausted battery.

Each path is composed of a finite set of links, each node can belong to only one path, except source nodes and sink node. In GMFT, all generated routing paths are node-disjoint routing paths.

Let $N = \{n_1, \dots, n_m\}$ be the finite set of nodes, $L = \{l_1, \dots, l_k\}$ be the finite set of links, and $P = \{p_1, \dots, p_n\}$ be the finite set of paths. MaxPathNum is the maximum path number used by source node for each flow transmission.

Static holes are the subset of failed nodes before transmission or blocked nodes, while dynamic holes are the subset of active nodes or failed nodes during data transmission phase, due to the depletion of their energy or their physical destruction.

4. DYNAMIC CURATIVE MECHANISM

4.1. Description

The design of our DCM is based on the different characteristics related to multimedia transmission in WMSNs and is also exposed to three following sub-problems:

- Maintain network stability when communication links break;
- Ensure load balancing, to extend network lifespan;
- Bypass dynamic holes formed during data transmission phase, to avoid packets' early loss.

Our main goal is to ensure reliability of transmitted data, extend network lifetime, satisfy QoS requirements by tolerating potential faults that may occur such as, energy depletion, hardware failure, communication link errors and interferences, in order to keep the system stable and without interruption, in case of failure of some of its components.

Based on some limits of offline geographic routing protocols, where the discovery of paths is performed before sending data. Thus, constructed paths may not reflect network reality at transmission time, especially in case of mobility or breakdown of nodes. Any disturbance impacts directly the quality of multimedia streams at the receiver. Furthermore, as multimedia data transmission is generally characterized by a long duration and a large size, sensor nodes are highly likely to fail due to the depletion of their energy. Therefore, we introduced our dynamic curative mechanism based on fault detection, dynamic load balancing and local repair, to conserve network stability during data transmission phase and solve the issue related to reliability of transmitted data, hence allowing to cope with topology changes.

4.2. Operation

In case of failure, the curative mechanism is used when a node notices that its successor is missing by not receiving successive acknowledgments and reacts by applying:

- Dynamic load balancing strategy: it sends a blocking message to source node, to prevent packet loss and increase delivery rate. Upon receiving a blocking message, source node blocks immediately the faulty path and distributes the flow on the remaining operational paths, to ensure continuity of transmitted data.
- Local repair strategy: It initiates exploration phase similar to the one previously described. Nevertheless, it stops this action after meeting a junction node belonging to the initial broken path, and which distance from the collector is closer than failed node distance. This allows a local reparation without having to rebuild a new path from the source to the collector. When the link is established, all nodes that are no longer part of the new path are released.

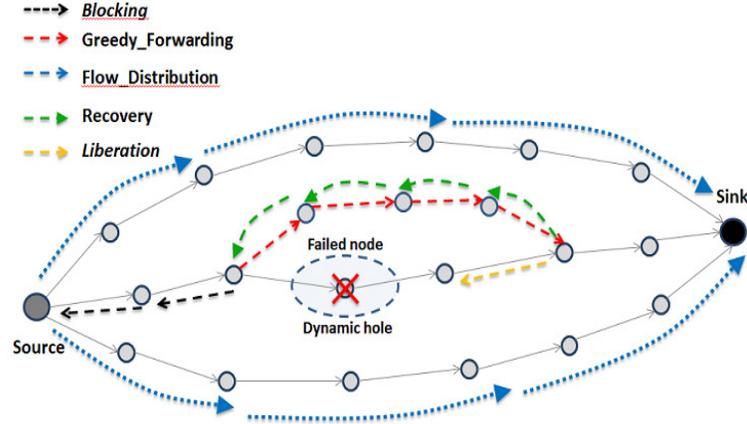


Figure 1. Dynamic curative mechanism

4.3. Theoretical modelling

To the best of our knowledge, all geographic routing protocols endowed with fault tolerance mechanism or not, do not support dynamic holes involving one or more nodes caused by broken links during data transmission phase.

If node $n_i \in N$, $n_i \in p_j$ with $p_j \in P$, fails at time t_{n_i} , all packets which used this faulty path after t_{n_i} are lost. We define the Number of Lost Packets (NLP) as follows:

$$NLP_{n_i} = \frac{NSP}{N_{path}} \left(1 - \frac{t_{n_i}}{T_S} \right) + \beta \quad (1)$$

NSP: Number of sent packets.

N_{path} : Number of paths.

t_{n_i} : Time when node n_i fails.

T_S : Total simulation time.

β : Number of lost packets due to transmission errors.

The first argument of formula (1) defines the packet loss due to path breaks, while the second argument is related to transmission errors.

In general, we obtain, with N_{FN} , Number of failed nodes during transmission phase

$$NLP_{(n_1, \dots, n_{FN})} = \frac{NSP}{N_{path}} \left(N_{FN} - \frac{\sum_{i=1}^{FN} t_{n_i}}{T_S} \right) + \beta \quad (2)$$

In our design, a node realizes the lack of its successor by not receiving successive acknowledgments after i_n iterations number, the NLP is independent of failure time and paths containing failed nodes. This can be modelled this way:

$$NLP = i_n N_{FN} + \alpha \quad (3)$$

i_n : Iterations number required to trigger curative mechanism;

α : Number of lost packets due to transmission errors.

5. EVALUATION

In order to demonstrate the strength of our proposed solution, we implemented our DCM on our previous protocol [14] and conducted several simulation, based on TinyOS [15] platform which conception defers from other OSs and relies on low-energy consumption operations. We also implemented both TPGF and AGEM [16] protocols, which are geographic routing protocols cited in several recent papers, according to their advantages. TPGF is effective in optimal path discovery, while AGEM is efficient in energy, also their common strengths, the bypassing of static holes during paths construction and flow management by multipath transmission.

Parameter	Value
Network size	500m x 250m
Number of sensors nodes	253
Bandwidth	250 Kbits/s
Transmission range	25 m
Packet size	128 Bytes
Data size	3.34 Mb
Number of paths	5

Table 1. Main configuration parameters

5.1. Theoretical results

Protocols	Number of failed nodes				
	1	2	3	4	5
TPGF/AGEM	2990	5981	8971	11962	14953
With DCM	5	10	15	20	25

Table 2. Theoretical number of lost packets

Table 2. shows the NLP for a total of 30000 sent packets and we assume that the time of failure of each node is the same: $t_{n1}=t_{n2}=t_{n3}=t_{n4}=t_{n5}=Ts/2$.

Our dynamic curative mechanism allows a better reliability as compared with other existing solution in WMSN. Therefore, it allows to minimise the NLP in case of path breaks during data transmission phase, which had never been solved in previous works.

5.2. Simulation results

5.1.1 Delivery Ratio (DR)

Is the ratio, for a given period of time, between the Number of Received Packets (NRP) and total Number of Sent Packets (NSP), it reflects the reliability of the protocol during packets' transfer from source node to destination.

$$DR = \frac{NRP}{NSP} = \frac{NSP - NLP}{NSP} \quad (4)$$

Having an optimal DR heavily relies on reducing the NLP. Figure.2(a) shows that the success rate of TPGF and AGEM registered a significant deterioration as compared to DCM and decreases from 99% to (48%, 81%) respectively for TPGF and AGEM, when five nodes belonging to different paths failed, and this is due to the non-maintenance of broken paths. One can note that

larger the dynamic hole (several failed nodes), the smaller the delivery rate. This confirms formulas (1 and 2) mentioned above and the effectiveness of our curative mechanism.

Delivery ratio

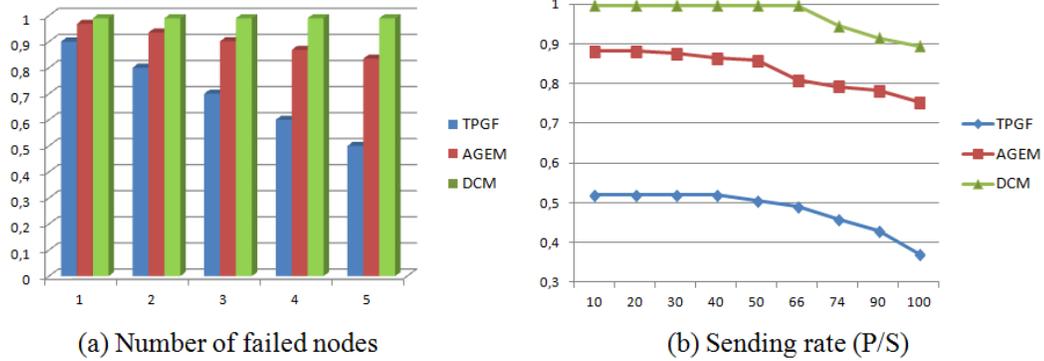


Figure 2. Impact of number of failed nodes and sending rate on delivery ratio

5.1.2 Receiving Rate (RR)

This metric allows to measure the continuity of the receiving flow at the collector (a high-receiving rate provides a good quality of multimedia stream). Otherwise, it is the average rate of received packets per unit of time.

$$RR = \frac{N_{rp}}{T_{rp} - T_1} \tag{5}$$

Nrp: Number of received packets.

Trp: Reception time of the last packet.

T1: Reception time of the first packet.

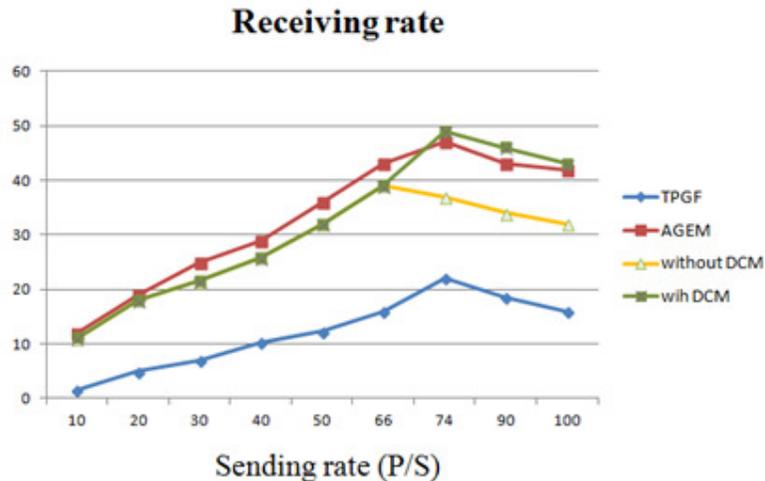


Figure 3. Impact of number of failed nodes on receiving rate curative

According to Figure.3, TPGF receiving rate is much lower, due to high rate of lost packets caused by number paths reduction, while AGEM receiving rate is lower than with DCM, due to overflow of waiting queues, when holes are detected. When the number of paths is reduced, sensor nodes

are more loaded and queues are faster overflowed and a more important packet loss is recorded. Simulation results shows that using DCM takes advantage by keeping higher the number of paths, which consequently improves the receiving rate.

6. CONCLUSION

In this work we presented a dynamic curative mechanism of fault tolerance for geographic routing dedicated to multimedia streams and faulty wireless sensor networks, which tolerate failures and adapts to topology changes (dynamic holes and node's mobility). Simulation results show that our solution provides better performance and satisfies QoS requirements for multimedia transmission especially in terms of delivery rate and receiving rate.

More importantly, our dynamic curative mechanism of fault tolerance, can be adapted to any geographic routing protocol, which aims to improve the reliability of transmitted data especially in case of communication links break.

REFERENCES

- [1] Sun, H. Luo, and S. K. Das. A trust-based framework for faulttolerant data aggregation in wireless multimedia sensor networks. *IEEE Transactions on Dependable and Secure Computing*, 9(6):785–797, Nov 2012.
- [2] A. Ouadjaout, Y. Challal, N. Lasla, and M. Bagaa. Seif: Secure and efficient intrusion-fault tolerant routing protocol for wireless sensor networks. In *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*, pages 503–508, March 2008.
- [3] S. Dave and A. Raghuvanshi. Fault tolerance techniques in distributed system. *International Journal of Engineering Innovation and Research*, 1(2):124–130, March 2012.
- [4] K. Wang, L. Wang, C. Ma, L. Shu, and J. Rodrigues. Geographic routing in random duty-cycled wireless multimedia sensor networks. In *2010 IEEE Globecom Workshops*, pages 230–234, Dec 2010.
- [5] L. Shu, Y. Zhang, L. T. Yang, Y. Wang, M. Hauswirth, and N. Xiong. Tpgf: geographic routing in wireless multimedia sensor networks. *Telecommunication Systems*, 44(1):79–95, 2010.
- [6] R. E. Ahmed. A fault-tolerant, energy-efficient routing protocol for wireless sensor networks. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on*, pages 175–178, May 2015.
- [7] D. B. Johnson and D. A. Maltz. *Dynamic source routing in ad hoc wireless networks*. In *Mobile computing*, pages 153–181. Springer, 1996.
- [8] C. Kavitha and K. V. Viswanatha. An energy efficient fault tolerant multipath (eeftm) routing protocol for wireless sensor networks. In *Advance Computing Conference, 2009. IACC 2009. IEEE International*, pages 746–751, March 2009.
- [9] Z. Hamid, F. B. Hussain, and J. Y. Pyun. Delay and link utilization aware routing protocol for wireless multimedia sensor networks. *Multimedia Tools and Applications*, pages 1–22, 2015.
- [10] N. Ahmed, S. S. Kanhere, and S. Jha. A pragmatic approach to area coverage in hybrid wireless sensor networks. *Wireless Communications and Mobile Computing*, pages 23–45, 2011.
- [11] A. Ahmed, K. A. Bakar, M. I. Channa, K. Haseeb, and A. W. Khan. Terp: A trust and energy aware routing protocol for wireless sensor network. *IEEE Sensors Journal*, 15(12):6962–6972, Dec 2015.

- [12] S. Ozdemir. Functional reputation based data aggregation for wireless sensor networks. In 2008 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pages 592–597, Oct 2008.
- [13] G. Han, J. Jiang, L. Shu, J. Niu., and H. C. Chao. Management and applications of trust in wireless sensor networks: A survey. *J. Comput. Syst. Sci.*, 80(3):602–617, 2014.
- [14] M. N. Bouatit, S. Boumerdassi, and A. Djama. A Geographic Multipath Routing Protocol for Wireless Multimedia Sensor Network. In International Conference on Mobile, Secure and Programmable Networking, pages 99-108, June 2016.
- [15] P. Levis and D. Gay. *TinyOS Programming*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.
- [16] S. Medjiah, T. Ahmed, and F. Krief. Agem: Adaptive greedy-compass energy-aware multipath routing protocol for wmsns. In 2010 7th IEEE Consumer Communications and Networking Conference, pages 1–6, Jan 2010.

AUTHORS

Mohamed Nacer Bouatit is currently working toward the PhD degree in computer science at CEDRIC-CNAM laboratory in Paris. He received a Master in web technologies from Telecom Bretagne School in 2011 and the engineer degree in computer science from the Polytechnic School of Algiers (EMP) in 2006. His research interests include wireless sensor networks, multimedia transmission, fault tolerance and energy efficiency.



Selma Boumerdassi is an Associate Professor at Conservatoire National des Arts et Métiers, Paris. She received a PhD in Computer Science from University of Versailles in 1998, where she also served as an Assistant Professor from 1998 to 2000. Her research interests include wireless and mobile networks, with a special focus on the impact and use of social networks. She worked on several national projects and served as an expert for the evaluation of French national projects (ANR). She is the author of more than 50 articles and serves as a TPC member for various international journals and conferences.



SECRET IMAGE TRANSMISSION THROUGH MOSAIC IMAGE

Shahanaz N and Greeshma R

Department of Computer Science and Engineering,
M Dasan Institute of Technology, Kozhikode, India

ABSTRACT

A secret image hiding scheme is proposed with new security features. This scheme utilizes the mosaic images, which is created from the secret and target images. A mosaic image is similar to that of the target image. The secret image fragments are hidden in the target image by performing appropriate color transformations. The inverse color transformation is performed for the lossless recovery of secret image. The color transformation is controlled by the proper overflow /underflow methods. The relevant information for recovering the secret image is embedded in the mosaic image by a lossless data hiding with the help of a key. Only with the proper key, the secret image is retrieved from the mosaic image

KEYWORDS

Image hiding, Mosaic images, Color transformation, Data hiding, Image encryption

1. INTRODUCTION

In the present world, enormous data are transmitted over networks around the clock and the security of the data is the major concern. By the boom of web, a large sort of data are being exchanged between devices, which may fall into different categories like images, audio, video, hypertext, graphics etc. Among these images play an important role. Various applications which are dealing with image data are business archives, medical images, space and research related graphics, confidential enterprise archives, document storage systems, and military image databases and so on. These images usually contain private or confidential information so that they should be protected from fraudulent access during transmissions. Recently, many methods have been proposed for securing image transmission, for which two common approaches are image encryption and data hiding.

Image Encryption makes use the Shannon diffusion confusion process. Here the natural properties like spatial correlation and data redundancy are utilized.[1][7]. These are mainly Chaos based algorithms which controls the encryption with variety of parameters like ergodicity , initial condition. When using the 3D cat maps [2] for encryption the statistical and gray code attacks are avoided due to the large key space. But it cannot resist the brute force attack. The pseudo random substitution and permutation [6] using standard map overcomes the brute force attack and chaos specific attacks. The computational complexity is the main issue. The encryption using Henon

Chaotic map [7] provides a lossless recovery and easy implementation. Randomness is the result of all these Chaos based methods. But this makes the attraction of the eavesdroppers while in the transmission. An alternative is used for avoiding this problem that is data hiding [8][14]. In these types of approach a cover image is used for hiding the secret data. Any one cannot realize the existence of secret data in this cover image. Existing data hiding methods mainly utilize the techniques of LSB substitution [], histogram shifting [9], difference expansion [10], prediction-error expansion [13]. When using LSB substitution [8] the quality of stego image may loss depending bits substituted. The histogram shifting [9] method cannot provide security for the data embedding. Thus, a main problem of the methods for hiding data in images is the difficulty to embed a large amount of message data into a single image. Specifically, if one wants to hide a secret image into a cover image with the same size, the secret image must be highly compressed in advance.

Some of the methods that are similar to the proposed method are [15][19]. The mosaic images play an important role in these methods. Mosaic [21] is a kind of art in which small pieces of material such as glass, stone are composed together to form a single image. In digital form small fragments of images called as tiles are arranged to form a single image called as mosaic. Creation of mosaic by computer is a new research area now a day. Different mosaics can be created from a single image depending on their choice of tiles and their placement in resulting image. There are different types of mosaic that includes crystallization mosaic, ancient mosaic that are created by dividing the secret image into tiles and then reconstructing the image by properly painting the tiles so these types of mosaic can also be called as tile mosaic.

Other types of mosaic includes photo-mosaic, and puzzle image mosaic that are formed by painting or covering the given sources image by fitting different images from the database hence they can also be called as multi-picture mosaic image. A special kind of mosaic include secret fragment visible mosaic image that are created by dividing secret image into small fragments called as tiles and then arranging these tiles in a random or puzzled sequence with the help of another image called as carrier image. The resultant mosaic is such that all fragments of secret image are visible to user but as they are arranged in puzzled form no one will be able to guess or read the contents of secret image.

In [15] create a mosaic image using the preselected target from the database and secret image. But here keeping the large database makes the process very complex and the user is not free to select the target image as his/ her own wish. A genetic, algorithm [16] based method is proposed and use the pseudorandom permutation [20] to create the mosaic image. The fitting information is embedded in the mosaic image then extracted the for the recovery of the secret image. Here the user is free to select the target image, but the genetic algorithm feature makes the process complex. An Enhanced Image Steganography Technique in Art Images [17] is used for creating mosaic images. In this approach first create cubism like image from the target image and then divide each secret and target to form the mosaic image by computing the matching score of secret tiles with that of target.

A new scheme is introduced in which the user is free to select the target image of his/her own wish. The created mosaic image looks similar to that of target image selected. The secret image retrieval is done perfectly.

In the proposed method two phases are there. In the first phase the secret and target image are divided in to tiles and blocks respectively. The fitting is done based on the standard deviation of

each tile and block, while performing the appropriate colour transformation and rotating each tile with respect to lowest RMSE value. Thus mosaic image is created. Then embed the relevant information. In the second phase the recovery information is extracted and inverse colour transformation is done. Then the secret image is recovered.

The following sections, In Section 2 introduce proposed method and presents necessary algorithms used by the proposed method. Section 3 explains the implementation details. In section 4 relates the results and discussions. Conclusion are summarised in section 5. Last section contains the papers, books, referred during the preparation of this paper.

2. PROPOSED METHOD

The overview of the proposed system is shown in Fig.1. The proposed method includes two main phases.1) mosaic image generation and 2) secret image recovery.

In the first phase a mosaic image is generated, which consists of the fragments of an input secret image with modified color properties that of target selected. In this phase 1) fitting of the secret tiles into target blocks, 2) transforming color characteristics of each tile image in the secret image to become that of the corresponding target block in the target image;3) rotating extracted to recover the secret image losslessly. The phase includes two stages: 1) extracting the embedded each tile image to find the maximum match with target with respect to smallest RMSE value.; and 4) embedding relevant information in to the mosaic image for the future recovery of the secret image.

In the second phase the embedded information is information for secret image recovery, and 2) recovering the secret image using the extracted information.

For the first phase we give two input images secret and target. A key can also be used to assure the security purpose. If the key is correctly decrypted then only the secret image is recovered. The output of this phase is a mosaic image with embedded information

For the next phase the input is a mosaic image and a key. Performing the decryption and decoding step by step the information needed to recover secret image is gathered. Thus the output of this phase is the secret image.

Algorithm 1 Mosaic image generation

Input: a secret image S, a target image T and a secret key K

Output: a secret-fragment-visible mosaic image F.

Stage 1.Fitting the tile images into the target blocks.

1. Divide the secret image in to tile image fragments and target in to blocks up to n.
2. Compute the mean and standard deviation of each tile and blocks.
3. Sort the tiles images and the target blocks according to the computed standard deviation and create a hash map.

4. Create a blank image and fit the first block from the target in that and again fit the secret tile up to n with respect to the mean and standard deviation calculated.

Stage 2. Performing color conversions between tile images and the target blocks.

5. Based on the mean, standard deviation, standard deviation quotient transform the color of each pixel in the tile image.

Stage 3. Rotate tile images.

6. Compute the RMSE values of each color transformed tile image with respect to its corresponding target block after rotating in to each of directions $\Theta = 0^\circ, 90^\circ, 180^\circ$ and 270° with smallest RMSE value.

Stage 4. Embedding the secret image recovery information.

7. Construct Huffman table using the mapping sequence constructed previously.
8. Construct a bit stream with; 1) index of target blocks, 2) optimal rotation angle, 3) the mean, 3) standard deviation quotients of three color channels, 4) the bit sequence for overflows/underflows.
9. Embed the bit streams with reversible contrast mapping [14]
10. Construct the bit stream I including 1) number of iterations for embedding; 2) the number of pixel pairs in the last iteration; 3) The Huffman Table constructed above.

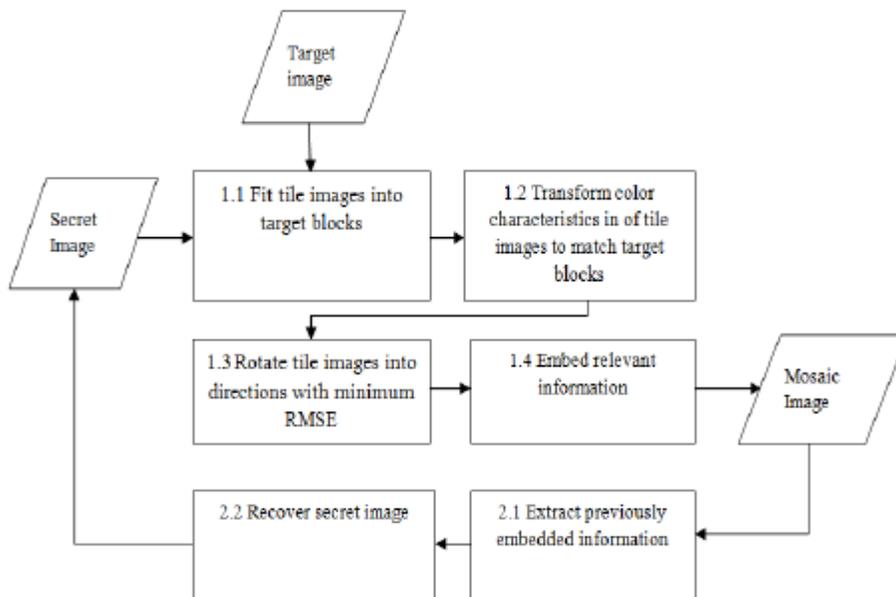


Fig. 1 Overview of Proposed System

Algorithm 2 Secret image Recovery

Input: a mosaic image F with n tile images and the key.

Output: the secret image S.

Stage 1. Extracting the secret image recovery information.

1. Extract from F using reversible contrast mapping [14] and decode them to obtain the recovery information and extract the bit sequences.
2. Decrypt the bit stream.
3. Decompose the bit stream to get n tile image information.
4. Decode the tile image information to get the following data items: 1) the index of target blocks, 2) the optimal rotation angle, 3) the means, standard deviation quotients of all color channels, and 4) the overflow /underflow residual values.

Stage 2. Recovering the secret image.

1. Recover in a raster scan order as one by one of the tile images from $i=1$ to n to get the secret image S by following steps: 1) rotate the tiles in a reverse angle, 2) use the mean, standard deviation quotients to recover the original pixel values, 3) use mean, standard deviation quotients to compute the parameters to balance the overflow/under flow problems, 4) take the results as the final tile image.
2. Compose all the final tile images to form the secret image S as output.

3. IMPLEMENTATION DETAILS

The proposed method can be implemented in two modules like mosaic image generation and secret image recovery. For the first we select two images as secret and target from the web. Each selected images can be saved in a folder that it can be viewed when the particular user wishes. For the purpose of finding the similarity a score can be calculated. It helps the user to select the appropriate target for the secret image. Then each image is divided in to tiles and blocks respectively. Mean, standard deviation and standard deviation quotient of three color channel(R, G, B) are calculated. This can be mapped into a hash map and sorted.

Thus we get two maps sorted. Then a blank image is created and each target block is placed in that and corresponding secret tile. This can be continued up to the divided units (that is if 8×8 ; 64 tiles, 16×16 ; 256 tiles and so on. For each tile image perform the color conversion of each pixel using the mean and standard deviation quotient. Then rotate each tile image in different angles and fix when lowest RMSE value is reached. This can be done for all n tiles. Next we have to embed the all information that we used to generate the mosaic image in the mosaic image itself and it can be controlled using a key. The hash map, rotation angles, bits that control under flow /over flow are embedded. The bit stream is encoded to particular stream.

Secret image is recovered only when the key is decrypted correctly. This is the main security feature that the proposed system offers. When the key is decrypted all the embedded information is decoded. The reverse rotation and inverse color transformation are performed. From the map we get the original positions of the secret tiles. A blank image is created and each tile is fitted as in the previous step. Thus the secret image is created.

4. RESULT AND DISCUSSIONS

4.1 Results

A series of experiments have been conducted to test the proposed method using many secret and target images with different sizes. To show that the created mosaic images looks the preselected target image, the quality metric of root mean square error (RMSE) is used.

An example is shown in the Fig. (2); Fig.2(c) shows the mosaic image created from Fig.2(a) as secret image and Fig.2(b) as target image. The tile image size is 16 x 16. The recovered secret image using the correct key is shown in Fig.2 (d) which is similar to that of the secret image in Fig.2 (a) with an RMSE value of 0.91 to that of the original one. Another example is shown in Fig. (3); Fig.3(c) shows the mosaic image created from Fig.3(a) as secret image and Fig.3(b) as target image. The tile image size is 8 x 8. The recovered secret image using the correct key is shown in Fig.3 (d) which is similar to that of the secret image in Fig.3 (a) with an RMSE value of 0.948 to that of the original one. Fig.3 (e) and Fig.3 (f) shows the mosaic image created from with tile image size 16 x 16 and 32 x 32 respectively. This means that when the tile size increases the clarity of the mosaic image get reduced.

4.2 Performance Evaluation

In this section, present several performance evaluation metrics that have been used for the quality of mosaic images and secret images. The various parameters are: 1) RMSE values of created mosaic images with respect to target image; 2) RMSE values of recovered secret image with respect to original secret image; 3) MSSIM values of created mosaic images with respect to target images. RMSE values find the mean square errors between the images, while the MSSIM values shows the similarity of images. Both can be used in alternatively.

It is recommended that we can use the PSNR values also for finding the difference in mosaic and target images. It will give how much data is hidden in the mosaic image when the ratio is calculated. The signal can be the mosaic and the secret hidden is considered to be the noise.

It is recommended that we can use the PSNR values also for finding the difference in mosaic and target images. It will give how much data is hidden in the mosaic image when the ratio is calculated. The signal can be the mosaic and the secret hidden is considered to be the noise.

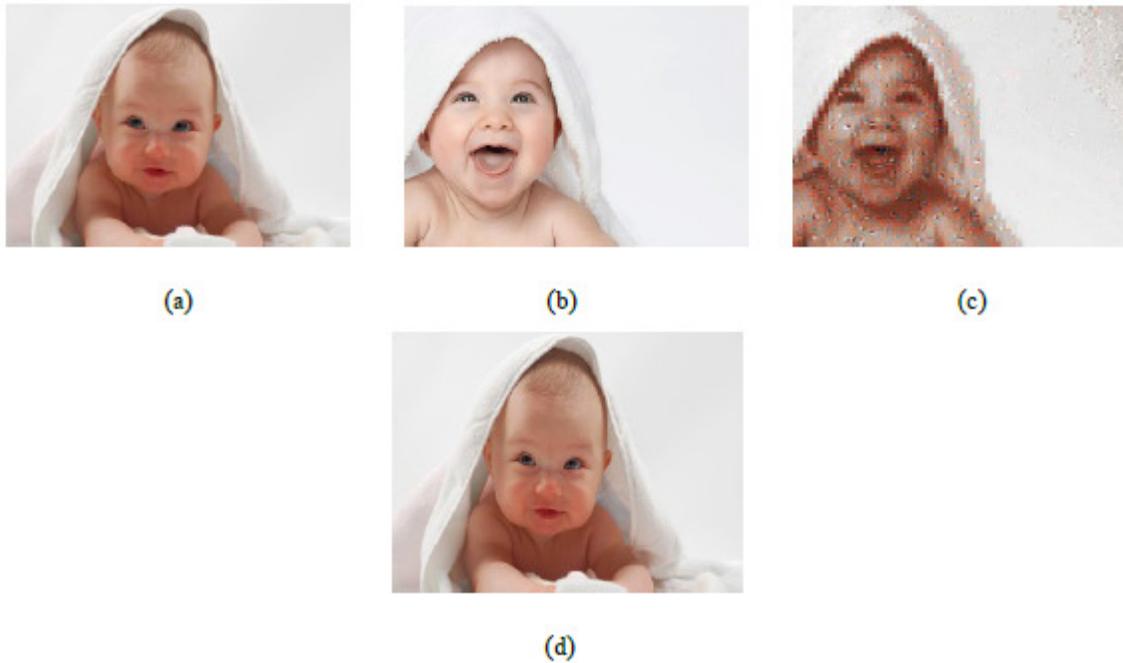


Fig. 2. Result yielded by the proposed method. (a) Secret image. (b) Target image. (c) Mosaic image created with tile image size 16×16 from (a) and (b) by the proposed method. (d) Secret image recovered from the mosaic image.



Fig. 3. Experimental result of mosaic image creation. (a) Secret image. (b) Target image. (c) Mosaic image created with tile image size 8×8 . (d) Recovered secret image using a correct key with $RMSE = 0.948$ with respect to secret image (a). (e) And (f) Mosaic images created with different tile image sizes: 16×16 , 32×32

The following graphs shows variations of RMSE values when different tile image sizes are used. The Fig. (4) Shows the RMSE value lies between 25 and 45, that is the mosaic image contains the target and secret images. So when calculating RMSE value of mosaic image with respect to target image it shows high variation when comparing with the next graph values, it lies between 0.5 and 2. From the graph Fig.(5), we can infer that the RMSE value quite increases when the tile size increases for the Fig. (2) and also for Fig.(3). We can achieve high similarity and low RMSE value when using the reduced tile image size. So it is preferred to use small sized tile images and high quantity of tiles.

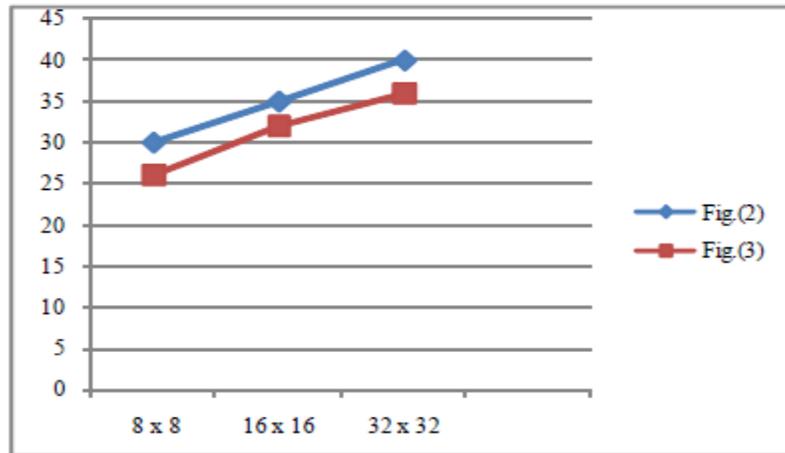


Fig 4. Graph showing the RMSE values of created mosaic images w.r.t. target images

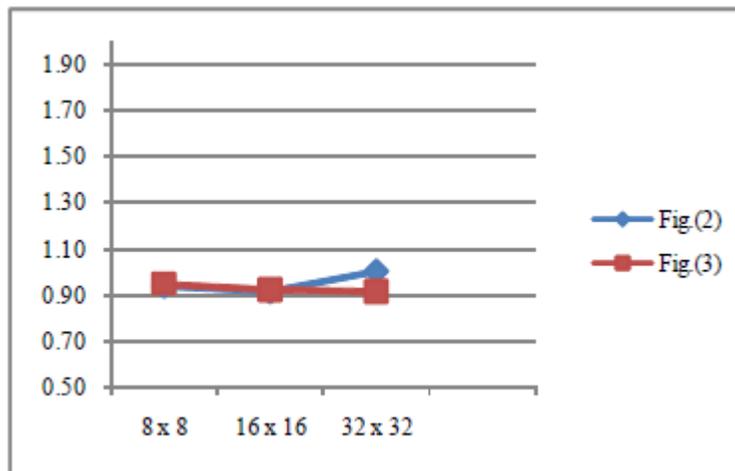


Fig 5. Graph showing the RMSE values of recovered secret images w.r.t original secret images

5. CONCLUSION

A secret image hiding scheme has been proposed with new security features. A meaning full mosaic image is created from the secret image and the selected target image. The fitting is done properly for generating the mosaic image. By using the proper color transformation to each pixel of secret image, we can achieve high visual similarities to the selected target image. These similarities are checked using the quality metric RMSE value. The main advantage of this scheme is that the secret image is recovered nearly without any loss. The security can be achieved by using a key that controls the recovery information embedding process.

ACKNOWLEDGEMENT

The authors would like to thank everyone, in the Department of Computer Science and Engineering for their support and help

REFERENCES

- [1] J. Fridrich, "Symmetric ciphers based on 2D chaotic maps," *Int. J. Bifurcat. Chaos* vol. 8, no. 6, pp. 1259–1284, 1998.
- [2] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos Solit. Fract.*, vol. 21, no. 3, pp. 749–761, 2004.
- [3] H. S. Kwok and W. K. S. Tang, "A fast image encryption system based on chaotic maps with finite precision representation," *Chaos Solit. Fract.*, vol. 32, no. 4, pp. 1518–1529, 2007.
- [4] "Arnold's Cat Map Gabriel Peterson" *Math 45 – "Linear Algebra" Fall 1997*
- [5] Yen JC, Guo JI. "A new chaotic key-based design for image encryption and decryption". In: *Proc IEEE Int Conference Circuits and Systems*, vol. 4, 2000. p.49– 52.
- [6] V. Patidar, N. K. Pareek, G. Purohit, and K. K. Sud, "A robust and secure chaotic standard map based pseudorandom permutation substitution scheme for image encryption," *Opt. Commun.*, vol. 284, no. 19, pp. 4331–4339, 2011.
- [7] Asia Mahdi Naser Alzubaidi, "Color Image Encryption and Decryption using Pixel Shuffling with Henon Chaotic System", "International Journal of Engineering Research & Technology", ISSN:2278-0181, vol. 3, issue 3, March – 2014
- [8] C. K. Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," *Pattern Recognit.*, vol. 37, pp. 469–474, Mar. 2004.
- [9] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, Mar. 2006.
- [10] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.
- [11] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Reversible data hiding," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Sep. 2002, pp. 157–160.

- [12] Ran-Zan Wang, Chi-Fang Lin, Ja-Chen Lin, “Image hiding by optimal LSB substitution and genetic algorithm”, Pattern Recognition 34 (3) (2001) 671– 683.
- [13] X. Li, B. Yang, and T. Zeng, “Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection,” IEEE Trans. Image Process., vol. 20, no. 12, pp. 3524–3533, Dec. 2011.
- [14] D. Coltuc and J.-M. Chassery, “Very fast Watermarking by Reversible Contrast Mapping,” IEEE Signal Process. Lett., vol. 14, no. 4, pp. 255–258, Apr. 2007.
- [15] I. J. Lai and W. H. Tsai, “Secret-fragment-visible mosaic image—A new computer art and its application to information hiding,” IEEE Trans. Inf. Forens. Secur., vol. 6, no. 3, pp. 936–945, Sep. 2011.

AUTHORS

Mrs. Shahanaz N is an Assistant Professor in the Information Technology Department of College of Engineering Vadakara, Kerala, India. She did her B.Tech in 2006 from College of Engineering, Vadakara, Kerala, India under the Cohin University of Science and Technology, followed by her M.Tech Post Graduation at M.Dasan Institute of Technology, Ulliyeri, Kerala, India, Calicut University, in 2016. Her interested areas are Digital Image Processing, Bioinformatics.



Mrs. Greeshma R is an Assistant Professor in Department of Computer Science and Engineering, M.Dasan Institute of Technology, Kozhikode, Kerala, India. Her interested area is Image Processing.

MANAGING SECURITY AND COMPLIANCE RISKS OF OUTSOURCED IT PROJECTS

Moneef Almutairi and Stephen Riddle

School of Computing Science, Newcastle University,
Newcastle Upon Tyne, UK

ABSTRACT

Several constraints, such as business, financial, and legal can lead organizations to outsource some of their IT services. Consequently, this might introduce different security risks to major security services such as confidentiality, integrity and availability. Analysing and managing the potential security risks in the early stages of project execution allows organizations to avoid or minimize such security risks. In this paper, we propose an approach that is capable of managing the security and compliance risks of outsourced IT projects. Such an approach aims to allow organizations to minimize, mitigate, or eliminate security risks in the early stages of project execution. It is designed to manage variation in security requirements, as well as provide a methodology to guide organizations for the purpose of security management and implementation.

KEYWORDS

Security and compliance management, outsourced IT projects, security management approach.

1. INTRODUCTION

As a result of globalization and heightened competition, many organizations today are confronted with significant challenges in developing services that satisfy customer requirements, and these needs must be met on time despite limited resources [1]. With the emergence of cloud computing and advances in web technology, such challenges and customer demands continue to expand. In response, many organizations have considered outsourcing to deliver and improve their IT services. Over the last two decades, Information Systems (IS) outsourcing has grown rapidly. This growth has prompted academia and industry to investigate the benefits that organizations may gain from outsourcing, and to determine the reduction in risk that might be achieved when adopting such a choice [2],[3].

Information Systems outsourcing has been defined in [2],[1],[4] as:

a business practice in which an organization subcontracts with a preferred third party to develop, operate, manage, or maintain its information system functions partially or totally for a specific period of time.

Outsourcing is an attractive option for organizations, offering benefits including cost reduction and the opportunity to concentrate on core business activities [5], [6], [7]. However, it is an option which must be managed properly as it brings risk, such as to security, contract violations,

and the loss of technology skills for the organization [6],[8],[9]. Failure to manage these risks could lead to major issues not only in a particular project, but also for the entire organization or business [10].

The rest of this paper is structured as follows: section 2 gives information about the background and related work. In section 3, we present our framework. Section 4 is used to give an example. In section 5, we conclude this paper.

2. BACKGROUND AND RELATED WORK

Securing information systems should follow a systematic approach which does not necessarily rely only on technical aspects, but also takes into account other aspects such as people and environment [11],[12],[13]. Such a systematic approach can help organizations achieve business continuity and minimize security risks [14],[15]. The most common systematic approaches are Information Security Management System (ISMS) standards and frameworks such as the ISO/IEC 2700x family [16], OCTAVE [17], and COBIT [18]. These standards and frameworks represent general security best practice guidance of IT processes and procedures, and can be adopted by organizations to achieve information systems confidentiality, integrity, and availability, and reduce associated security risks [16],[19],[20]. Ensuring compliance with ISMS standards and frameworks is an essential part of information systems security, as unenforced ISMS will not achieve the expected value of such practices [21]. The status of the compliance with the ISMS standards and frameworks is normally achieved via audit or self-assessment. However, although audits can provide good outcomes, they suffer from a lack of broad assessment, and are time consuming. Moreover, while self-assessment can provide broader assessment, it may also suffer from a lack of depth assessment [22],[23].

Although many organizations have adopted ISMS standards and frameworks to secure their information systems, these represent general best practice of ISMS and do not consider that security requirements differ from one organization to another [24]. Moreover, there is no adequate guidance for implementing or complying with such standards and frameworks, and nor are they designed to manage the security and compliance risks of outsourced IT project [25]. Updating these standards and frameworks to fit the outsourced IT project context might make them more complicated and increase time and resource consumption. Instead, our proposal is designed to overcome these weaknesses.

When outsourcing IT services, two main parties are involved: a client and a provider. Security requirements need to be documented in project documents such as the project contract and Service Level Agreement (SLA). The provider has to comply with these security requirements to deliver these IT services to the client correctly.

IT organizations today deal with diverse security risks such as terrorist attacks and natural disasters [14]. Such security risks force organizations to take action to minimize, mitigate, or eliminate issues as early as possible before they are exploited by attackers or their systems are damaged. To manage the security and compliance risks of outsourced IT project effectively, specific requirements need to be met:

- Security requirements management: the security program or framework should be comprehensive and systematic as well as establishing a complete methodology that is capable of adequately managing the security of outsourced IT projects. This includes security policies, access controls, plans, and procedures.

- Risk Management: Security risks are not only technical, and so the security program should manage risks from different perspectives such as technical, human, and environmental and physical risks.
- Compliance management: The security program should establish a method to enforce compliance properly.
- Usability: The security program should be usable from different perspectives such as cost effectiveness, time efficiency, and simplicity.

3. MANAGING SECURITY AND COMPLIANCE RISKS OF OUTSOURCED IT PROJECTS

We propose a framework for managing the security and compliance risks of outsourced IT projects. The framework utilizes project phases (initiating, execution, monitoring and controlling, and closing) and the Plan-Do-Check-Act (PDCA) model [26], as shown in Fig 1. Each project phase has its own security activities. During the planning, execution, and monitoring and controlling phases, the PDCA model is applied. Managing project security should be aligned with the project phases in consideration of improvements during the project execution. This improves flexibility, simplicity, and ease of use, regardless of the project size, cost, or any other constraints. The framework uses a hybrid threat modelling approach that is designed for the outsourcing context, in which environments are less stable and more systems are integrated. The threat modelling approach in such environments needs to achieve some desired properties. It should be exclusive, exhaustive, unambiguous, repeatable, comprehensive and useful to capture the largest possibility of potential security threats [27], [28], [29]. The hybrid threat modelling is designed to overcome the limitations of existing threat modelling approaches that use only two or three criteria, and the lack of consideration of the desired properties [30],[31]. It combines different threat modelling criteria and considers threats from different perspectives such as external threats, provider threats, client threats, and physical and environmental threats. It is designed to be capable of capturing the largest possibilities of potential security threats that might occur during the project execution.

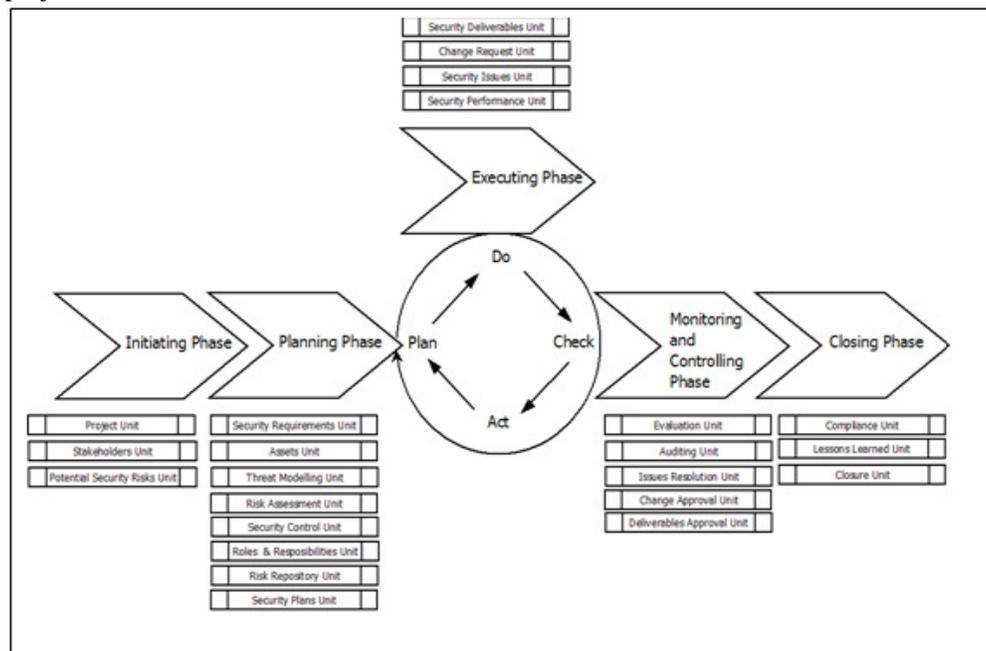


Figure 1 : Security and compliance risks management framework

3.1. Initiating Phase

The aim of this phase is to establish the project's unique details, and identify major project stakeholders and general security risks to which the project might be exposed. The client is responsible for this phase, as no provider has yet been selected. This phase achieves its objectives through the following units:

- Project unit: designed to handle (create, update) essential project data such as project Id, name, code...etc.
- Stakeholders unit: the aim of this unit is to identify all project stakeholders and their roles and responsibilities in the project.
- Potential security risks unit: the aim of this unit is to identify potential security risks that might take place while executing the project. This allows decision makers to take the right decision whether to outsource or not.

3.2. Planning Phase

The planning phase is the core part of this framework. It corresponds to the Plan stage of the PDCA model. The main parties involved in the project execution carry out all core security activities at this phase. The primary parties are the client and the provider. If other stakeholders are involved in executing security activities, then they participate with the primary parties in preparing the required security analysis and plans. The planning phase aims are achieved in units as follows:

- Security requirements unit: the project security requirements are documented and signed off by the client and the provider through this unit.
- Assets unit: all project assets are identified and categorized (hardware, software, information, or network) through this unit. It answers what should be protected. If the project affects any organization's assets, then they should be identified and categorized as well.
- Threat modelling unit: a hybrid threat modelling approach is designed to be capable of capturing the largest number of potential security threats that might occur during the project execution. In another words, it identifies what to protect from. This threat modelling approach uses six criteria:
 - Threat Source: it represents the origin of the threat, which can be external threats, client threats, provider threats, or environmental and physical threats.
 - Threat agent: the agent that causes the threat. This can be technical, human, or organizational.
 - Asset type: the type of the asset impacted by this threat, such as networks, software, hardware, or information.
 - Threat intention: the type of human behaviour who caused the threat. It can be accidental or intentional.
 - Environmental and physical threat type: the type of environmental and physical threat. It can be controlled or uncontrolled.
 - Threat impact: the result of the threat when it occurs. This can be any breach in confidentiality, integrity, or availability.

- Risk assessment unit: the ultimate aim of this unit is to estimate and prioritize the impact of the potential security risk on project assets. The risk assessment unit has five steps: vulnerabilities identification, risk likelihood determination, risk magnitude determination, risk estimation, and risk prioritization.
- Security controls unit: through this unit, security controls or countermeasures that can mitigate identified security risks are selected. Countermeasures determine how to protect project or organization assets. Countermeasures are categorized in this framework to technical, human, or organizational.
- Roles and responsibilities unit: The aim of this unit is to assign security activities to project teams using a clear method that helps prevent any ambiguities between the project teams, especially if there is another client or stakeholder team involved. In this framework, we propose a role based on the responsible, accountable, consult, and inform method (RACI) for assigning the roles and responsibilities of project security activities.
- Risk repository unit: this unit contains all thus-far identified project security risks. Any risk that has been logged into the risk repository unit should have sufficient information about the risk such as risk Id, description, impact, asset name, and so on.
- Security plans unit: This unit is responsible for developing security plans that will be used to achieve project security goals and contribute to building a secure and protected environment such as an incident management plan, business continuity management plan, and so on.

3.3. Executing phase

In the previous phase, the project teams engaged in planning security activities and controls that mitigate and minimize potential security risks associated with the project. In this phase, which represents the Do stage of the PDCA model, the security plans and controls proposed in the previous phase are implemented. Any security issues that might be experienced at this phase are documented and monitored. If there is any need for improvements or changes, the project team will record them. This phase has four units:

- Performance unit: prepare and submit security performance reports. These reports are reviewed and signed off by the project steering committee in the next phase.
- Security issues unit: record any security issues that might be experienced during the project execution.
- Change requests unit: if there is any need to change any security plan or control, the change is raised through this unit.
- Security deliverables unit: ensure that security deliverables are submitted on time to avoid any delay, which might lead to penalties.

3.4. Monitoring and controlling phase

The project execution needs to be monitored and controlled not only by the project manager, but also by the project steering committee to ensure that it meets its requirements and provide all the support that contributes to the achievement of the security and non-security goals while executing the project. The aim of this phase, which represents the Check and Act stages of the PDCA model, is to review the execution performance reports, and assess if there is any need for

improvement. Moreover, the project steering committee supports the project manager in resolving security issues that require intervention. Security change requests and security deliverables are reviewed and approved during this phase too. The monitoring and controlling phase has five units:

- Evaluation unit: the security performance reports are reviewed and assessed. Based on this review, the project steering committee may propose improvements that help achieve project security goals in an effective and efficient way. If the performance is good and there is no need for any improvement, then the steering committee signs off existing performance reports.
- Auditing unit: this unit is designed to enforce compliance with the security requirements to reduce any security violations. Security countermeasures and plans are audited to assess their conformance to what have been planned and agreed.
- Issues resolution unit: security issues that might be experienced during the project execution are resolved. Security issues resolution might be beyond the ability of the project manager, and therefore intervention by the project steering committee might help in their resolution.
- Change approval unit: the project steering committee analyses security changes and assesses their potential impact on different aspects such as the project budget and schedule. If these changes can be tolerated by both parties, then they approve them.
- Deliverables approval unit: the security deliverables that have been achieved so far are reviewed and approved, or rejected.

3.5. Closing Phase

When the project is completed, it will be handed over to the client. Before control is taken by the client, the project requirements including security requirements need to be verified to ensure that the project has achieved its security and non-security goals. The aim of this phase is to audit and verify the project requirements to close the project officially and issue the provider with a closure certificate. Moreover, the lessons learned during the project phases are documented at this phase for future use. The closing phase has three units:

- Compliance unit: demonstrate that the applications or the products being delivered by the project are secure and work according to the requirements agreed in the project scope of the work.
- Lessons learned unit: document all security lessons learned for future use.
- Closure unit: issue the provider with the project closure certificate and officially close the project.

4. EXAMPLE SCENARIO

Let us assume that a government agent, who runs major IT systems for a government, has contracted with a provider to develop e-services using their existing systems. After completing the project, the agent discovers that the confidentiality of their watch list data has been breached by the provider staff while they were integrating the e-services with the watch list systems, and by external attackers when an attack on e-services took place. Although this example is very simple, it illustrates some of the security issues that might arise when outsourcing in the absence of a

comprehensive and systematic approach to the management of security risks. Breaches of confidentiality can be avoided or mitigated by using the proposed framework as follows:

- In this scenario, the agent identifies confidentiality breaches of private data by the provider staff or attackers, as these e-services are provided over the internet. In the absence of this step, the agent may enter a contract without knowing the risk level involved. This step allows decision makers to take the right decision in advance concerning whether to outsource, after considering appropriate security countermeasures that mitigate such a security risk, or consider alternatives such as in-house development.
- If the agent has taken the decision to outsource, then there should be a comprehensive and systematic method of effectively identifying and managing potential security risks. The proposed framework provides that method. In our example, the security requirement is to protect the agent's data confidentiality. The asset under impact is the agent data, which is categorized in our framework as information. By using the proposed hybrid thread modelling approach, the asset is exposed in this scenario to some threats, which include external threats by attackers and provider threats such as information disclosure. The risk of exposure to these threats should be estimated and prioritized based on the proposed semi-quantitative approach that the proposed framework provides. This risk may be mitigated by technical countermeasures such as cryptography and firewalls for external threats and organizational countermeasures, such as a non-disclosure agreement for provider threats. Finally, countermeasures implementation is assigned to the correct teams, and the required security plans are developed.
- Having analysed and assessed security risks in addition to planning security countermeasures that help mitigate potential security risks, the project execution starts by following what has been planned. This prevents security changes that might violate the security requirements being made without approval, and help in achieving security activities on time, as they will be part of the project master schedule. In our example, this includes executing technical and organizational countermeasures and security plans.
- To enforce compliance, the proposed framework allows the project steering committee to review, evaluate, and audit security requirements and controls as well as approve or reject security changes while the project team engage in executing the project. This allows the agent and provider to minimize compliance violations as much as possible. In our scenario, this includes auditing the cryptography and security plans implementation.
- At the end of the project execution, the proposed framework provides a way of demonstrating that the project complies with the project security requirements as well as officially closing the project. In our scenario, this includes demonstrating that cryptography and security plans work as claimed.

4. CONCLUSION

In this paper, we propose a framework for the management of security and compliance risks of outsourced IT projects. It is designed to meet all identified requirements and overcome any weaknesses in existing information security system standards and frameworks. Risks associated with all parties involved in project execution are analysed and managed in a systematic way. It is a structured approach, which uses project phases to manage and control project security risks. The framework is flexible as it follows the PDCA model, which allows the project teams involved in managing security risks to monitor and evaluate security controls continuously, and implement any improvements or changes. Simplicity and ease of use are other features of this framework as

it utilizes project phases for the management of security risks, which allows the separate management of security risks during each phase. Utilizing project phases makes the framework applicable to any project regardless of size, time, or other constraints. The risks analysis and threat modelling of the current project can be applied to new projects that have similarities, making reusability another feature of this framework. We aim to apply this framework to a real case study in the near future, and also to use a focus group to provide independent validation evidence.

REFERENCES

- [1] I. Oshri, J. Kotlarsky, and L. P. Willcocks, *The Handbook of Global Outsourcing and Offshoring* 3rd Edition: Springer, 2015.
- [2] J. Dibbern, T. Goles, R. Hirschheim, and B. Jayatilaka, "Information systems outsourcing: a survey and analysis of the literature," *ACM Sigmis Database*, vol. 35, pp. 6-102, 2004.
- [3] C. Brandas, "Risks and audit objectives for IT outsourcing," *Informatica Economica Journal*, vol. 14, pp. 113-118, 2010.
- [4] Q. Hu, C. Saunders, and M. Gebelt, "Research report: Diffusion of information systems outsourcing: A reevaluation of influence sources," *Information Systems Research*, vol. 8, pp. 288-301, 1997.
- [5] K. Han and S. Mithas, "Information Technology Outsourcing and Non-IT Operating Costs: An Empirical Investigation," *MIS Quarterly*, vol. 37, pp. 315-331, 2013.
- [6] R. Gonzalez, J. Gasco, and J. Llopis, "Information systems outsourcing reasons and risks: an empirical study," *International Journal of Human and Social Sciences*, vol. 4, pp. 181-192, 2009.
- [7] C. Schwarz, "Toward an understanding of the nature and conceptualization of outsourcing success," *Information & Management*, vol. 51, pp. 152-164, 2014.
- [8] R. Gonzalez, J. Gasco, and J. Llopis, "Information systems outsourcing risks: a study of large firms," *Industrial management & Data systems*, vol. 105, pp. 45-62, 2005.
- [9] R. Gonzalez, J. Gasco, and J. Llopis, "Information systems outsourcing reasons and risks: a new assessment," *Industrial Management & Data Systems*, vol. 110, pp. 284-303, 2010.
- [10] J. Iqbal, R. Binti Ahmad, and M. A. Noor, "Frequently occurring risks for IT outsourcing projects," In *Proceedings of the International Conference on Computer and Communication Engineering (ICCCCE)*, pp. 957-960, 2012.
- [11] J. Holappa and T. Wiander, "Practical Implementation of ISO 17799. Compliant Information Security Management System Using Novel ASD Method," In *Technical Report*, 2006.
- [12] D. Tse, "Security in modern business: Security assessment model for information security practices," In *Proceeding of the Pacific Asia Conference of Information Systems*, pp. 1509-1519, 2004.
- [13] P. Ifinedo, "Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory," *Computers & Security*, vol. 31, pp. 83-95, 2012.
- [14] R. Saint-Germain, "Information security management best practice based on ISO/IEC 17799," *Information Management*, vol. 39, pp. 60-66, 2005.
- [15] J. Eloff and M. Eloff, "Information security architecture," *Computer Fraud & Security* 2005, pp. 10-16, 2005.

- [16] E. Humphreys, "Information security management system standards," *Datenschutz und Datensicherheit-DuD*, vol. 35, pp. 7-11, 2011.
- [17] C. Alberts, A. Dorofee, J. Stevens, and C. Woody, "Introduction to the OCTAVE Approach," Pittsburgh, PA, Carnegie Mellon University, 2003.
- [18] G. Ridley, J. Young, and P. Carroll, "COBIT and its Utilization: A framework from the literature," In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 2004.
- [19] E. Humphreys, "Information security management standards: Compliance, governance and risk management," *information security technical report*, vol. 13, pp. 247-255, 2008.
- [20] J. S. Broderick, "ISMS, security standards and security regulations," *information security technical report*, vol. 11, pp. 26-31, 2006.
- [21] S. B. von Solms, "Information Security Governance—compliance management vs operational management," *Computers & Security*, vol. 24, pp. 443-447, 2005.
- [22] M. Vogel and V. Broer, "Security Compliance Monitoring—The next Evolution of Information Security Management," *ISSE 2013 Securing Electronic Business Processes*, pp. 183-194, 2013.
- [23] A.-M. Ghirana and V. P. Bresfelean, "Compliance Requirements for Dealing with Risks and Governance," *Procedia Economics and Finance*, vol. 3, pp. 752-756, 2012.
- [24] M. Siponen and R. Willison, "Information security management standards: Problems and solutions," *Information & Management*, vol. 46, pp. 267-270, 2009.
- [25] T. Wiander, "Positive and negative findings of the ISO/IEC 17799 framework," In *Proceedings of 18th Australian Conference on Information Systems (ACIS 2007)*, 2007.
- [26] R. Moen and C. Norman, "Evolution of the PDCA cycle," In *Proceedings of the Asian Network for Quality Congress*, pp. 15-19, 2009.
- [27] M. Jouini, L. B. A. Rabai, and A. B. Aissa, "Classification of security threats in information systems," *Procedia Computer Science*, vol. 32, pp. 489-496, 2014.
- [28] U. Lindqvist and E. Jonsson, "How to systematically classify computer security intrusions," In *Proceedings of the 1997 IEEE Symposium on Security & Privacy*, pp. 154-163, 1997.
- [29] F. Farahmand, S. B. Navathe, G. P. Sharp, and P. H. Enslow, "A management perspective on risk of security threats to information systems," *Information Technology and Management*, vol. 6, pp. 203-225, 2005.
- [30] M. Alhabeeb, A. Almuhaideb, P. D. Le, and B. Srinivasan, "Information security threats classification pyramid," In *24th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 208-213, 2010.
- [31] S. Gerić and Ž. Hutinski, "Information system security threats classifications," *Journal of Information and Organizational Sciences*, vol. 31, pp. 51-61, 2007.

AUTHORS**Moneef Almutairi**

Mr. Almutairi received his bachelor degree in information systems from King Saud University (KSA) in 2001, and his master degree from Newcastle University (UK) in 2009. He is currently doing his PhD in computer science at Newcastle University (UK).

**Stephen Riddle**

Dr Riddle obtained his BSc in Computer Software Technology at the University of Bath in 1991, and completed a PhD at Bath in 1997 on the use of partial specifications and refinement theory to aid the process of explaining complex systems. Dr Riddle currently works at Newcastle University and delivers courses in formal specification (VDM-SL) and software development techniques at undergraduate level; software engineering, dependable systems, Java programming and high-integrity software development (SPARK) at postgraduate level. His research interests include: software engineering, security risk management, and requirements specifications for systems of the systems.



LARGE-SCALE MULTI-USER MIMO APPROACH FOR WIRELESS BACKHAUL BASED HETNETS

Mostafa Hefnawi

Department of Electrical and Computer Engineering,
Royal Military College of Canada, Kingston, Canada

ABSTRACT

In this paper, we consider the optimization of wireless capacity-limited backhaul links in future heterogeneous networks (HetNets). We assume that the HetNet is formed with one macro-cell base station (MBS), which is associated with multiple small-cell base stations (SBSs). It is also assumed both the MBS and the SBSs are equipped with massive arrays, while all mobile users (macro-cell and small-cell users) have single antenna. For the backhaul links, we propose to use a capacity-aware beamforming scheme at the SBSs and MRC at the MBS. Using particle swarm optimization (PSO), each SBS seeks the optimal transmit weight vectors that maximize the backhaul uplink capacity and the access uplinks signal-to-interference plus noise ratio (SINR). The performance evaluation in terms of the symbol error rate (SER) and the ergodic system capacity shows that the proposed capacity-aware backhaul link scheme achieves similar or better performance than traditional wireless backhaul links and requires considerably less computational complexity.

KEYWORDS

HetNets, wireless backhaul, cognitive radio, Massive MIMO, multiuser MIMO, PSO.

1. INTRODUCTION

Recently, deploying small cell networks over existing macro-cellular networks, also known as heterogeneous networks (HetNets), has emerged as a promising solution to deal with the increasing wireless traffic demands in next generation 5G cellular networks [1]-[6]. The users in these HetNets are offloaded from the congested macro-cell base stations (MBSs) to the small-cell base stations (SBSs), which enhanced their quality of service (QoS) and increase the overall system capacity. These HetNets are supported by Gigahertz bandwidth backhaul links that connect MBSs and the associated SBSs. Such Gigahertz bandwidth can be achieved by conventional optical fiber or millimeter-waves (mmWaves) based wireless backhails. Optical fiber backhails while reliable, they might be expensive and difficult to deploy in HetNets where several small cells are unplanned and installed quite arbitrarily. Wireless backhails, on the other hand, are more attractive to overcome the restriction of deployment and installation and can provide a cheap and scalable solution. However, to achieve high spectral efficiency in HetNets with wireless backhauling, frequency reuse across the coexisting network tiers (backhaul and access links) is essential and interference management is critical. Cognitive radio based HetNets (CR-HetNets) has emerged as a promising solution that provides a more energy efficient and dynamic way to use the spectrum by enabling small-cell to share licensed bands in opportunistic manner [5]-[6]. In CR-HetNets, macro-cell users, which are considered as primary users (PUs)

take the priority to access the channels, whereas small-cell users, which are considered as secondary users (SUs), can access the channels as long as the corresponding PUs do not use them. However, most of these proposed CR-HetNets have assumed opportunistic spectrum sharing which may not be reliable and may limit the system capacity since it suffers from the interruptions imposed by the primary network (PN) on the SUs who must leave the licensed channel when PUs emerge. Also, with opportunistic spectrum sharing, SUs can still cause interference to PUs due to their imperfect spectrum sensing. In cellular systems, one way to overcome these limitations is to incorporate multiuser multi-input multi-output (MU-MIMO) approach into cognitive radio networks (CRNs) to achieve higher spectral efficiency by multiplexing multiple SUs on the same time-frequency resources and protecting PNs from SUs' interferences. MU-MIMO techniques have been successfully deployed in 4G cellular systems for traditional fixed spectrum assignment (FSA) approaches [7]-[15] and a vast number of multiuser detection algorithms are presently being tailored towards solving the MU-MIMO processing in cognitive networks [16]-[22], by imposing additional constraints to protect licensed users' QoS. More specifically, capacity-aware MU-MIMO schemes have been proposed for both FSA [13]-[15] and CR networks [16]-[17] using different multiuser detection schemes such as maximum ratio combining (MRC) and minimum mean-squared error (MMSE), and have shown the potential to exhibit better system capacity and provide better SER enhancement than traditional singular value decomposition (SVD)-based MU-MIMO systems. On the other hand, it was shown that the use of large-scale antenna arrays (also called massive MIMO) could achieve tremendous boost of MU-MIMO systems system performance [23]-[26]. In this paper, therefore, we will be applying the concept of MU massive MIMO and CR in HetNets. We assume that the MBSs and SBSs are equipped with massive arrays, while all mobile users have single antenna. We deploy two MU-MIMO schemes, namely, MRC at the access link (SUs to SBSs) and capacity-aware/MRC at the backhaul link (SBSs to MBS). Such a system can significantly improve the system performance in terms of link reliability, spectral efficiency, and energy efficiency. It can also achieve optimal performances with the simplest forms of user detection techniques, i.e., MRC [12]. On the other hand, most of the proposed capacity-aware MU-MIMO schemes require the use of gradient search algorithms in order to solve the constrained optimization problem in CRNs [16]-[17]. These techniques become very computationally expensive in large-scale MIMO systems because of the vast amounts of baseband data that are generated and require the constrained optimization problem to be differentiable. Thus, in our capacity-aware backhaul link scheme we will be exploring free-derivative population-based training algorithms such as the particle swarm optimization (PSO) that are well known by their simple/fast hardware implementation. PSO was initially introduced by Kennedy and Eberhart in [27] and has received a lot of attention in recent years. It is an evolutionary computation technique inspired by swarm intelligence such as fish schooling and bird flocking looking for the best food spot (exploring the optimal solution) in the search space where a quality measure, fitness, can be evaluated without any a priori knowledge. The PSO algorithm in this paper will be used at the backhaul link to seek iteratively the transmit beamforming weights of each SBS that maximize the uplink (UL) MIMO backhaul channel capacity. Under the assumption of very large number of antennas at the SBSs and the MBS, we derive semi-analytic expressions for the symbol error rate (SER) and the ergodic channel capacity, which quantify the reliability and spectral efficiency of the MU-MIMO based HetNet. The derived expressions are then used to evaluate the performance of the proposed PSO-based capacity-aware (PSO-CA) backhaul link. The contribution of this paper includes the extension of the cognitive capacity-aware massive MU-MIMO schemes to wireless backhaul links and the development of semi-analytical model for the SER and channel capacity analyses in HetNets.

2. SYSTEM MODEL

We consider the UL access scenario shown in Fig. 1 of a HetNet with K small cells and one macro cell that share the same frequency band. Each small cell includes one SBS equipped with massive N -element antenna array and L_s single-antenna secondary users (SUs). Each SBS and its users act as a cognitive network that coexist, via concurrent spectrum access, with L_p macro-cell primary users (PUs) and their primary MBS, which is also equipped with massive M -element antenna array. It is also assumed that both the SBS and the MBS detect independent OFDM data streams from their mobile users simultaneously on the same time-frequency resources.

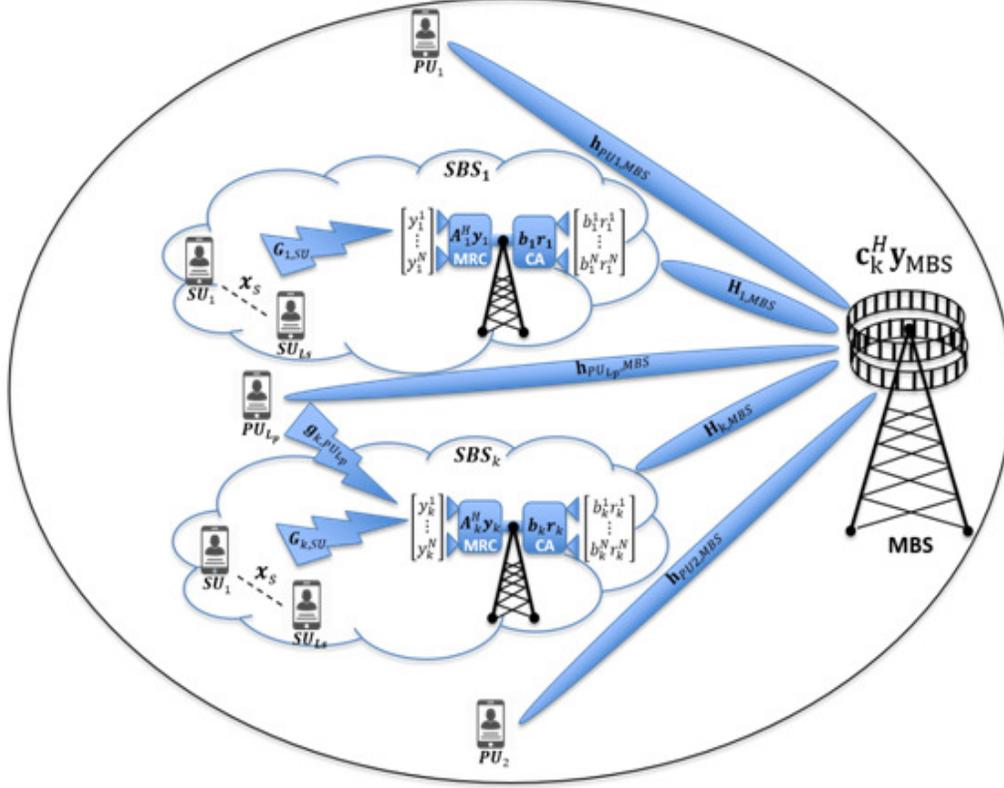


Figure 1. System Model: HetNet consisting of one macro-cell and K small-cells and their corresponding users.

Let $\mathbf{x}_s[f_i] = \{x_1^s, x_2^s, \dots, x_{L_s}^s\}$ and $\mathbf{x}_p[f_i] = \{x_1^p, x_2^p, \dots, x_{L_p}^p\}$ denote, respectively, the set of L_s SUs signals and L_p PUs signals transmitted on each subcarrier, f_i , $i = 1, \dots, N_c$, where N_c denotes the number of subcarriers per OFDM symbol in the system. The analysis is done separately on each subcarrier. For brevity therefore, we drop the frequency index f_i .

2.1. Access link:

The $N \times 1$ received signal vector at the k^{th} SBS is given by

$$\mathbf{y}_k = \sqrt{p_u} \mathbf{G}_{k,SU} \mathbf{x}_s + \mathbf{n}_{k,SBS} + \mathbf{I}_{PU,SBS} \quad (1)$$

where $\mathbf{G}_{k,SU} \in \mathbb{C}^{N \times L_s}$ is the channel matrix between the k^{th} SBS and its L_s users, $\mathbf{x}_s \in \mathbb{C}^{L_s \times 1}$ is the transmitted signal vector of L_s users in the k^{th} small-cell, p_u is the average power transmitted

by each user (Here we assume equal power allocation for all users) , $\mathbf{n}_{k,SBS} \in \mathbb{C}^{N \times 1}$ is the received AWGN vector at the SBS, and $\mathbf{I}_{PU,SBS}$ represents the interference introduced by macro-cell users (PUs) at the SBS, and is given by

$$\mathbf{I}_{PU,SBS} = \sqrt{p_p} \mathbf{G}_{k,PU} \mathbf{x}_p, \quad (2)$$

where $\mathbf{G}_{k,PU} \in \mathbb{C}^{N \times L_p}$ is the channel matrix between the k^{th} SBS and L_p users, p_p is the average power transmitted by each PU, and $\mathbf{x}_p \in \mathbb{C}^{L_p \times 1}$ is the transmitted signal vector of L_p users in the HetNet.

For the uplink access link, we consider MRC detection scheme at each SBS. The k^{th} SBS processes its received signal \mathbf{y}_k by multiplying it by the $N \times L_s$ receive beamforming weight matrix \mathbf{A}_k^H as follows

$$\mathbf{r}_k = \mathbf{A}_k^H \mathbf{y}_k = \sqrt{p_u} \mathbf{A}_k^H \mathbf{G}_{k,SU} \mathbf{x}_s + \mathbf{A}_k^H \mathbf{n}_{k,SBS} + \mathbf{A}_k^H \mathbf{I}_{PU,SBS} \quad (3)$$

The detection of user l_s by its k^{th} SBS can then be expressed as

$$r_{k,l_s} = \mathbf{a}_{k,l_s}^H \mathbf{y}_k = \sqrt{p_u} \mathbf{a}_{k,l_s}^H \mathbf{G}_{k,SU} \mathbf{x}_s + \mathbf{a}_{k,l_s}^H \mathbf{n}_{k,SBS} + \mathbf{a}_{k,l_s}^H \mathbf{I}_{PU,SBS} \quad (4)$$

where r_{k,l_s} is the l_s^{th} element of \mathbf{r}_k and \mathbf{a}_{k,l_s} is the l_s^{th} column of \mathbf{A}_k .

2.2. Backhaul link:

For the backhaul link, the expression for the array output of the MBS in Fig. 1 can be written for each subcarrier as

$$\mathbf{y}_{MBS} = \sum_{k=1}^K \mathbf{H}_{k,MBS} \mathbf{b}_k \mathbf{r}_k + \mathbf{n}_{MBS} + \mathbf{I}_{PU,MBS}, \quad (5)$$

where \mathbf{y}_{MBS} is the $M \times 1$ vector containing the outputs of the M -element array at the MBS, $\mathbf{H}_{k,MBS}$ is the $M \times N$ frequency-domain channel matrix representing the transfer functions from the N -element antenna array of the k^{th} SBS to the M -element antenna array of the MBS, $\mathbf{b}_k = [b_1, b_2, \dots, b_N]^T$ is the $N \times 1$ complex transmit weight vector of the k^{th} SBS, \mathbf{n}_{MBS} is the received $M \times 1$ complex additive white Gaussian noise vector at the MBS, and $\mathbf{I}_{PU,MBS}[k]$ represents the interference introduced by PUs to SUs at the MBS and is given by

$$\mathbf{I}_{PU,MBS} = \sqrt{p_p} \mathbf{H}_{PU,MBS} \mathbf{x}_p \quad (6)$$

where $\mathbf{H}_{PU,MBS}$ is the $M \times L_p$ channel matrix from the L_p PUs to the MBS's M -element antenna array.

The MBS detects the k^{th} SBS data by multiplying the output of the array \mathbf{y}_{MBS} with the $M \times 1$ receiving weight vector, \mathbf{c}_k^H as follows

$$\hat{x}_k = \mathbf{c}_k^H \mathbf{y}_{MBS} = \mathbf{S}_k + \mathbf{S}_{I_s} + \mathbf{S}_{I_p} + \mathbf{N} \quad (7)$$

where

$\mathbf{S}_k = \mathbf{c}_k^H \mathbf{H}_k \mathbf{b}_k \mathbf{r}_k$ is the signal detected from the k^{th} SBS,

$\mathbf{S}_{I_s} = \mathbf{c}_k^H \sum_{i=1, i \neq k}^K \mathbf{H}_i \mathbf{b}_i \mathbf{r}_i$ is the multiple-access interference (MAI) from the $K - 1$ other SBSs, $\mathbf{S}_{I_p} = \sqrt{p_p} \mathbf{c}_k^H \mathbf{H}_{PU, MBS} \mathbf{x}_p$ is the MAI from L_p PUs, and $\mathbf{N} = \mathbf{c}_k^H \mathbf{n}_{MBS}$ is the noise signal at the array output of the MBS,

For the backhaul link, it is assumed that each SBS is transmitting with a capacity-aware beamforming scheme that will be discussed in section XX and that the MBS is detecting SBSs' signals using MRC scheme.

3. SYMBOL ERROR RATE AND ERGODIC CHANNEL CAPACITY

The symbol error rate, $SE_{R_{k,l_s}}$, associated with l_s^{th} user of the k^{th} SBS can be expressed as

$$SE_{R_{k,l_s}} = E_{\gamma_{k,l_s}} [aQ(\sqrt{2b\gamma_{k,l_s}})], \quad (8)$$

where $E[\cdot]$ denotes the expectation operator, $Q(\cdot)$ denotes the Gaussian Q-function, γ_{k,l_s} is the signal-to-interference-plus-noise ratio (SINR) associated with the l_s^{th} user of the k^{th} SBS, and a and b , are modulation-specific constants. For binary phase shift keying (BPSK), $a = 1$ and $b = 1$, for binary frequency shift keying (BFSK) with orthogonal signaling $a = 1$ and $b = 0.5$, while for M-ary phase shift keying (M-PSK) $a = 2$ and $b = \sin^2(\pi/M)$.

Using (7), the signal detected from the l_s^{th} user of the k^{th} SBS can be expressed by (9) and the signal detected by the MBS from the l_s^{th} user of the k^{th} SBS can be expressed by (10).

$$\begin{aligned} \mathbf{S}_{k,l_s} &= \mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k r_{k,l_s} = \mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{y}_k \\ &= \mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{G}_{k, SU} \mathbf{x}_s + \mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{n}_{k, SBS} + \mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{I}_{PU, SBS} \end{aligned} \quad (9)$$

$$\hat{\mathbf{x}}_{k,l_s} = \mathbf{c}_k^H \mathbf{y}_{MBS} = \mathbf{S}_{k,l_s} + \mathbf{S}_{I_s} + \mathbf{S}_{I_p} + \mathbf{N} \quad (10)$$

The SINR at the MBS for user l_s of the k^{th} SBS can thus be depicted as

$$\gamma_{k,l_s} = \frac{\mathbf{c}_k^H \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{G}_{k, SU} \mathbf{G}_{k, SU}^H \mathbf{a}_{k,l_s} \mathbf{b}_k^H \mathbf{H}_{k, MBS}^H \mathbf{c}_k}{\mathbf{c}_k^H \mathbf{B}_k \mathbf{c}_k} \quad (11)$$

and the ergodic channel capacity, per subcarrier, for each SBS k is given by [13]

$$C(\mathbf{H}_{k, MBS}, \mathbf{b}_k) = E \left(\log_2 \left\{ \mathbf{I} + \frac{p_u}{N} \left| \mathbf{B}_k^{-\frac{1}{2}} \mathbf{H}_{k, MBS} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{G}_{k, SU} \right|^2 \right\} \right) \quad (12)$$

where $E[\cdot]$ denotes the expectation operator, \mathbf{B}_k is the covariance matrix of the interference-plus-noise, and is given by

$$\mathbf{B}_k = \mathbf{B}_{SBS} + \mathbf{B}_{PU, MBS} + \mathbf{B}_{PU, SBS} + \mathbf{B}_n, \quad (13)$$

where

$$\mathbf{B}_{SBS} = \sum_{i=1, i \neq k}^K \mathbf{H}_{i, MBS} \mathbf{b}_i \mathbf{r}_i \mathbf{r}_i^H \mathbf{b}_i^H \mathbf{H}_{i, MBS}^H$$

$$\mathbf{B}_{\text{PU,MBS}} = p_p \mathbf{H}_{\text{MBS},l_p} \mathbf{H}_{\text{MBS},l_p}^H$$

$$\mathbf{B}_{\text{PU,SBS}} = p_p \mathbf{H}_{k,\text{MBS}} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{G}_{k,\text{PU}} \mathbf{G}_{k,\text{PU}}^H \mathbf{a}_{k,l_s} \mathbf{b}_k^H \mathbf{H}_{k,\text{MBS}}^H$$

$$\mathbf{B}_n = \mathbf{c}_k^H \mathbf{c}_k + \mathbf{H}_{k,\text{MBS}} \mathbf{b}_k \mathbf{a}_{k,l_s}^H \mathbf{a}_{k,l_s} \mathbf{b}_k^H \mathbf{H}_{k,\text{MBS}}^H$$

4. CAPACITY-AWARE BACKHAUL LINK

In the proposed capacity-aware backhaul link, the weight vector for the k^{th} SBS is updated at each iteration n until it reaches the optimal beamforming vector, $(\mathbf{b}_i)_{\text{opt}}$, that maximizes the ergodic backhaul channel capacity for each SBS k of the HetNet. This channel capacity can be expressed, at each iteration n , by:

$$C(n) = E \left(\log_2 \left\{ \mathbf{I} + \frac{p_u}{N} \left| \mathbf{B}_k^{-\frac{1}{2}} \mathbf{H}_{k,\text{MBS}} \mathbf{b}_k(n) \mathbf{a}_{k,l_s}^H \mathbf{G}_{k,\text{SU}} \right|^2 \right\} \right) \quad (14)$$

To maximize the backhaul link capacity we propose to employ particle swarm optimization algorithm where particles are mapped to the transmit beamforming and fly in the search space, aiming to maximize the fitness function given by the channel capacity of (14). First, the PSO generates Z random particles for each SBS (i.e., random weight vector $\mathbf{b}_k^{(z)}$, $z = 1, \dots, Z$ of length $N \times 1$) to form an initial population set S (swarm). The algorithm computes the channel capacity according to (14) for all particles $\mathbf{b}_k^{(z)}$ and then finds the particle that provides the global optimal channel capacity for this iteration, denoted $\mathbf{b}_k^{(z,\text{gbest})}$. In addition, each particle z memorizes the position of its previous best performance, denoted $\mathbf{b}_k^{(z,\text{pbest})}$. After finding these two best values, PSO updates its velocity $\mathbf{v}_k^{(z)}$ and its particle positions $\mathbf{w}_k^{(z)}$ at each iteration n using (15) and (16), respectively, where c_1 and c_2 are acceleration coefficients towards the personal best position (pbest) and/or global best position (gbest), respectively, φ_1 and φ_2 are two random positive numbers in the range of $[0, 1]$, and ω is the inertia weight which is employed to control the exploration abilities of the swarm.

$$\mathbf{v}_k^{(z)}(n+1) = \omega \mathbf{v}_k^{(z)}(n) + c_1 \varphi_1 (\mathbf{b}_k^{(z,\text{pbest})}(n) - \mathbf{b}_k^{(z)}(n)) + c_2 \varphi_2 (\mathbf{b}_k^{(z,\text{gbest})}(n) - \mathbf{b}_k^{(z)}(n)) \quad (15)$$

$$\mathbf{b}_k^{(z)}(n+1) = \mathbf{b}_k^{(z)}(n) + \mathbf{v}_k^{(z)}(n+1) \quad (16)$$

Large inertia weights will allow the algorithm to explore the design space globally. Similarly, small inertia values will force the algorithms to concentrate in the nearby regions of the design space. This procedure is repeated until convergence (i.e., channel capacity remains constant for a several number of iterations or reaching maximum number of iterations). An optimum number of iterations is tuned and refined iteratively by evaluating the average number of iterations required for PSO convergence as a function of the target MSE for algorithm termination and as a function of the population size. Since random initialization does not guarantee a fast convergence, in our optimization procedure we consider that the initial value of $\mathbf{b}_k^{(z)}(n)$ at iteration index $n = 0$ is given by the eigen-beamforming (EBF) weight, i.e., $\mathbf{b}_k^{(z)}(0) = \sqrt{P_u} \mathbf{u}_{\text{max},k}$, where $\mathbf{u}_{\text{max},k}$ denotes the eigenvector corresponding to $\lambda_{\text{max},k}$, the maximum eigenvalue of $\mathbf{H}_{k,\text{MBS}}^H \mathbf{H}_{k,\text{MBS}}$. This initial guess enables the algorithm to reach a more refined solution iteratively by ensuring fast convergence and allows to compute the initial value of the received beamforming vector at iteration index $n=0$. In our case we assume MRC at the receiving MBS, i.e:

$$\mathbf{c}_k^H(0) = (\mathbf{B}_k)^{-1} \mathbf{H}_{k,MBS} \mathbf{b}_k(0) \quad (17)$$

5. SIMULATION RESULTS

In our simulation setups we consider a HetNet organized into K SBSs ($K=20$) and one macro-cell. The number of antennas at the SBSs and at the MBS is the same, $N = M$, and is varying from 25 to 100. Each SBS is serving $L_s = 10$ users and the macro-cell is serving $L_p = 10$ users, each transmitting with a single antenna. We assume QPSK modulation. For the OFDM configurations, we assume the 256-OFDM system ($N_c = 256$), which is widely deployed in broadband wireless access services. For the backhaul link we assume an MU-MIMO system with capacity-aware beamforming at each SBS and MRC detection at the MBS. For the access link we assume MRC detection at each SBS. For the PSO parameters, the swarm size is 30, the maximum iteration number is 25 and the acceleration coefficients are $c_1 = c_2 = 2$. The inertia weight ω ranges from 0.9 to 0.4 and varies as the iteration goes on.

Fig. 2 shows the system capacity of the proposed PSO-CA and the traditional eigen-beamforming schemes for $M = N = 25$, and 100. It is observed that for both cases PSO-CA is outperforming eigen-beamforming. It is also noted that as we increase the number of antennas the performance gap between the two schemes is reduced. This means that when the number of base station antennas becomes large, PSO-CA is able to achieve the same level or better performance than eigen-beamforming with less computational complexity. Fig. 3, on the other hand, compares the SER performance of both schemes for the same scenario as in Fig. 2. It is observed PSO-CA is outperforming eigen-beamforming in both cases.

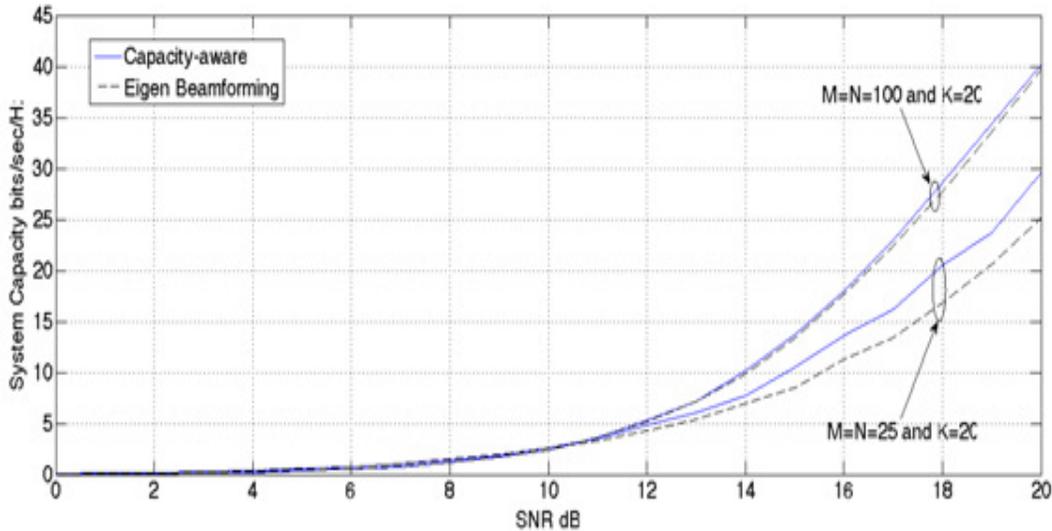


Figure 2. Ergodic channel capacity of HetNet using PSO-CA and eigen-beamforming schemes for $K=20$ SBSs and $M=N=25$ and 100 antennas.

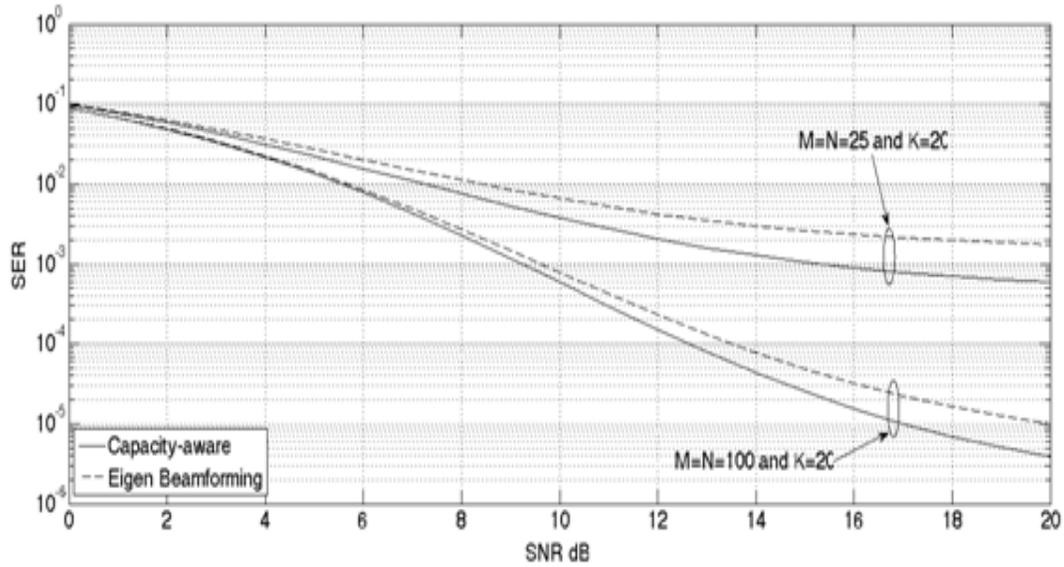


Figure 3. SER performance of HetNet using PSO-CA and eigen-beamforming schemes for $K=20$ SBSs and $M=N=25$ and 100 antennas.

6. CONCLUSION

This paper proposes a capacity-aware wireless backhaul link where cognitive small cells communicate with a MBS using a PSO-based large-scale multiple-input multiple-output (LS-MIMO) beamforming scheme. The proposed algorithm iteratively seeks the optimal transmit weight vectors that maximize the channel capacity of each SBS in the HetNet. It was shown that the proposed system is able to achieve a low computational complexity (without requiring an inverse of the covariance matrix) with the same level or better performance than the conventional eigen-beamforming.

ACKNOWLEDGEMENTS

The author would like to thank the Canadian Microelectronics Corporation (CMC) for providing the Heterogeneous Parallel Platform to run the computationally-intensive Monte-Carlo Simulations

REFERENCES

- [1] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: Challenges and solution approaches," *IEEE Wireless Communications*, Special Issue on "Smart Backhauling and Fronthauling for 5G Networks", vol. 22, no. 5, Oct. 2015, pp. 22-31.
- [2] Zhen Gao, Linglong Dai, De Mi, Zhaocheng Wang, Muhammad Ali Imran, and Muhammad Zeeshan Shakir, "MmWave Massive MIMO Based Wireless Backhaul for 5G Ultra-Dense Network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13-21, Oct. 2015.
- [3] H. Tabassum, A. Hamdi Sakr, and E. Hossain, "Analysis of massive MIMO-enabled downlink wireless backhauling for full-duplex small cells," *IEEE Transactions on Communications*, vol. 64, no. 6, June 2016, pp. 2354-2369.

- [4] Mehrdad Shariat, Emmanouil Pateromichelakis, Atta ul Quddus, and Rahim Tafazolli, "Joint TDD Backhaul and Access Optimization in Dense Small-Cell Networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 11, November 2015, pp. 5288-5299.
- [5] H. ElSawy, E. Hossain, and D. I. Kim, "HetNets with cognitive small cells: User offloading and Distributed channel allocation techniques," *IEEE Communications Magazine*, Special Issue on "Heterogeneous and Small Cell Networks (HetSNets)", vol. 51, no. 6, June 2013.
- [6] Zhi Yan, Wentao Zhou, Shuang Chen, and Hongli Liu, "Modeling and Analysis of Two-Tier HetNets with Cognitive Small Cells," *IEEE Access*, 2016.
- [7] M. Y. Alias, A. K. Samingan, S. Chen, and L. Hanzo, "Multiple antenna aided OFDM employing minimum bit error rate multiuser detection," *IEE Electron. Lett.*, Vol. 39, No 24, pp. 1769–1770, Nov. 2003.
- [8] P. Vandenameele, L. Van Der Perre, M. G. E. Engels, B. Gyselinckx, and H. J. De Man, "A combined OFDM/SDMA approach," *IEEE J. Select. Areas Commun.*, Vol. 18, No 11, pp. 2312–2321, Nov. 2000.
- [9] M. Jiang, S. Ng, and L. Hanzo, "Hybrid Iterative Multiuser Detection for Channel Coded Space Division Multiple Access OFDM Systems," *IEEE Trans. Veh. Technol.*, Vol.55, No1, Jan. 2006.
- [10] M. Munster and L. Hanzo, "Performance of SDMA Multi-User Detection Techniques for Walsh-Hadamard-Spread OFDM Schemes," *IEEE-VTC'01*, Vol. 4, pp. 2319-2323, Oct. 2001.
- [11] K.-K. Wong, R. Cheng, K. B. Letaief, and R. D. Murch, "Adaptive Antennas at the Mobile and Base Stations in an OFDM/TDMA System," *IEEE Trans. Commun.*, Vol. 49, No 1, Jan. 2001.
- [12] M. Kang, "A comparative study on the performance of MIMO MRC systems with and without cochannel interference," *IEEE Transactions on Communications*, Vol. 52, Iss. 8, pp. 1417 – 1425, 2004.
- [13] A. I. Sulyman and M. Hefnawi, "Adaptive MIMO Beamforming Algorithm Based on Gradient Search of the Channel Capacity in OFDMSDMA System," *IEEE Commun. Letters*, Vol. 12, No. 9, pp. 642-644, Sept. 2008.
- [14] A. I. Sulyman, and M. Hefnawi, "Performance Evaluation of Capacity-Aware MIMO Beamforming Schemes in OFDM-SDMA Systems," *IEEE Trans. Commun.*, Vol. 58, No. 1, Jan. 2010.
- [15] A. I. Sulyman, and M. Hefnawi, "Capacity-Aware Linear MMSE Detector for OFDM-SDMA Systems," *IET Communications*, Vol. 4, Iss. 9, June 2010.
- [16] M. Hefnawi and A. Abubaker, "Channel Capacity Maximization in Multiuser Large Scale MIMO-Based Cognitive Networks", *International Journal of Microwave and Optical Technology*, Nov. 2014.
- [17] M. Hefnawi, "SDMA Aided Cognitive Radio Networks," *IEEE 26th Biennial Symposium on Communications*, pp. 10 - 14, 2012.
- [18] L.-L. Yang and L.-C. Wang, "Zero-Forcing and Minimum Mean-Square Error Multiuser Detection in Generalized Multicarrier DS-CDMA Systems for Cognitive Radio," *EURASIP Journal on Wireless Communications and Networking*, pp. 1-13, 2008.
- [19] K. Hamdi, W. Zhang, and K. B. Letaief, "Opportunistic spectrum sharing in cognitive MIMO wireless networks," *IEEE Transactions on Wireless Communications*, Vol. 8, No 8, pp. 4098–4109, August 2009.

- [20] R. Zhang and Y.-C. Liang. Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 2, No. 1, pp. 88–102, February 2008.
- [21] S. Yiu, M. Vu, and V. Tarokh, “Interference Reduction by Beamforming in Cognitive Networks,” *IEEE GLOBECOM Telecom. Conf.*, pp. 1-6, 2008.
- [22] L. Bixio, G. Oliveri, M. Ottonello, M. Raffetto, and C. Regazzoni, “Cognitive radios with multiple antennas exploiting spatial opportunities,” *IEEE Transaction on Signal Processing*, Vol. 58, No 8, pp. 4453–4459, August 2010.
- [23] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Tran. Wirel. Comm.*, vol. 9, no. 11, pp. 3590 – 3600, Nov. 2010.
- [24] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.
- [25] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [26] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [27] J. Kennedy, R.C. Eberhart, “Particle swarm optimization,” *Proceedings of the IEEE Conference on Neural Networks IV*, , pp. 1942–1948, 1995.
- [28] E. Mijangos, “Approximate Subgradient Methods for Lagrangian Relaxation on Networks,” in *System Modeling and Optimization*,. IFIP International Federation for Information Processing, Vol. 312, pp. 370–381, 2007.

AUTHORS

Dr. Hefnawi is currently a professor and the Chair of Graduate Studies Committee in the Department of Electrical and Computer Engineering at the Royal Military College of Canada. Dr. Hefnawi is a licensed professional engineer in the province of Ontario. He is a contributing author of a number of refereed journal, book chapters, and proceeding papers in the areas of wireless communications. His research interest includes cognitive radio, wireless sensor network, massive MIMO, cooperative MIMO, multiuser MIMO, and smart grid communications.



JCWAEED: JOINT CHANNEL ASSIGNMENT AND WEIGHTED AVERAGE EXPECTED END-TO-END DELAY ROUTING PROTOCOL IN WIRELESS MESH NETWORKS

Wang Yi-rong¹, Wang Yan-ru², Zhang Hao³, Liu Kai-ming and Li Nan

^{1,2}Beijing Guodiantong Network Technology Co.Ltd., Beijing, China

³State Grid Shandong Electric Power Corporation,
Economic and Technical Research Institute, Shandong, China

ABSTRACT

In recent years, multi-channel multi-radio Wireless Mesh network has become one of the most important technologies in the evolution of next-generation networks. Its multi-hop, self-organization, self-healing and simple deployment is an effective way to solve the bottleneck problem of last mile. In this paper, we propose a new routing metric called WAEED, deployed in JCWAEED protocol, a joint channel assignment and weighted average expected end-to-end delay routing protocol which considers both interference suppression with factor IF and end-to-end delay. Additionally, we give the exact calculation formula of transmission delay and queuing delay. Simulations results demonstrate that JCWAEED outperforms other joint design routing protocols in terms of throughput, end-to-end delay and packet loss rate.

KEYWORDS

Multi-channel Multi-radio, Wireless Mesh Networks, End-to-end Delay Control, Joint Design, Channel Assignment, Routing Protocols

1. INTRODUCTION

Multi-radio multi-channel Wireless Mesh Network is one of the most important technologies in the evolution of next-generation networks. Its multi-hop, self-organization, self-healing and simple deployment is an effective way to solve the bottleneck problem of last mile. However, with the increase of the node density, it will cause the interference situation to deterioration, the channel utilization rate is low and other rigid environment. Considering the interaction between channel allocation and routing, it needs joint channel allocation and routing to optimize the design to improve the channel utilization and network performance, and play the advantages of multi-radio multi-channel technology. Since it is a challenging interdependent task to find routing paths and allocate channels to links in the WMN, its focus is joint schemes about channel assignment and routing in multi-channel multi-radio wireless mesh networks.

For the moment, there are two groups for joint channel assignment and routing schemes in multi-channel multi-radio wireless mesh networks [1]. One is called sequential channel assignment and

routing protocol design [2-5], and the strategy is that designers treat channel assignment and routing as two independent processes. People usually do channel assignment optimization and routing optimization separately in sequence. Nevertheless, we cannot acquire network request for communications information before channel assignment in reality. The other one is joint channel assignment and routing protocol design [6-8], and people treat channel assignment as a whole procedure for optimization. Specifically, this NP-hard problem can be turned into linear programming with some certain objective by setting constraint conditions for joint optimization. This strategy solution, nonetheless, is of both very high algorithm complexity and extraordinarily high time complexity. In addition, the approximate solutions are generally lack of strict theoretical support [9].

The paper further analyzes and compares the research status of joint optimal design of channel allocation and routing in multi-radio multi-channel wireless Mesh networks. In this paper, the necessity and feasibility of joint optimal design of channel allocation and routing in multi-radio multi-channel wireless Mesh networks are expounded in detail by referring to a lot of literatures. Based on the network model and protocol interference model, the mathematical model is carried out and the calculation method of average end-to-end delay is deduced out of a lot of theoretical research and simulation analysis. The HELLO packet is designed as the detection packet of network information, used to obtain the relevant network parameters in the network, and finally introduce the interference penalty factor IF, used to characterize the node using the interference of channel i . In this paper, the appropriate channel is selected by the size of the IF, and the routing metric is designed in combination with the average end-to-end delay. Based on the above work, this paper proposes an efficient joint channel allocation and routing optimization method (JCWAEED), considering delay control, interference suppression, channel allocation and routing establishment. The JCWAEED algorithm utilizes different channel combinations to obtain the set of single-hop routing metric WAEED from the node to the next hop based on the single hop metric. And the channel with the minimum single-hop routing metric in the set is taken as the best channel for the link from the node to the next hop. The performance of JCWAEED algorithm, including the number of streams and the rate of flow, is analyzed and compared with that of the existing protocols.

The paper is organized as follows. Section 1 introduce basic concepts and infrastructure of multi-radio multi-channel wireless Mesh networks, and describe the design idea and key technologies of multi-radio multi-channel Mesh networks in detail. Section 2 gives a exact description of some relative works about joint design of channel assignment and routing protocols. Section 3 demonstrates system model used in this paper. Section 4 proposes joint channel assignment and weighted average expected end-to-end delay routing protocol. Section 5 is our simulation results, and gives analysis of our protocol advantages in terms of network throughput, end-to-end delay, packet loss and channel utilization rate. Section 6 gives final conclusion of this paper.

Our major contributions can be summarized as follows.

1. We give average end-to-end delay in exact and strict theoretical calculation including transmission delay and queuing delay.
2. We provide more efficient way of interference factor IF between channels calculation during channel assignment period.
3. Considering two points above, we propose a new routing metric WAEED, deployed in JCWAEED protocol to conduct routing in MRMC-WMN.

2. RELATIVE WORKS

Channel assignment and routing design depend on each other can remedy weakness of each other .Hence the joint design can significantly improve network performance.

The first group of optimization is to treat channel assignment and routing as two independent processes optimizing them in sequence, and then select the appropriate channel according to the flow distribution in the network [10].

JCEED algorithm [2] assign channels after routing, considering end-to-end delay and queuing delay as routing metric and adjust channels allocation along with routing recursively with feedback information of EED until making total network interference deduction to minimum. Within every recursion, choose the path with the minimum delay, and choose the optimum channel according to the channel assignment algorithm. It decreases interference between channels by joint routing and channel assignment optimization under the limits of time delay. However, it does not give an exact and accurate computing method about end-to-end delay derivation. It is not comprehensive only using numbers of successfully received packets under channel i as index to represent the interference. Channel assignment initialization is not thorough enough, if co-channel were randomly assigned to different radios of one node, it will not only increase the channel switching times, but also increase the interference of other nodes assigned to the channel. Channel allocation process can be further optimized by reducing final establishing routing time.

The second group is joint channel assignment and routing protocol design [6][7][8], and people treat channel assignment as a whole procedure for optimization. Specifically, this NP-hard problem can be turned into liner programming with some certain objective by setting constraint conditions for joint optimization. This strategy solution, nonetheless, is of both very high algorithm complexity and extraordinarily high time complexity.

The key of channel allocation and routing joint optimization design is: network connectivity, network traffic and link-state information, technical limitations and the design of the routing metric [6]. Plus, routing metric design considers some key factors: the path hop, channel interference, radio interface switch [7], computational complexity decrement, convergence time decrement and local constraints [8].

3. SYSTEM MODEL

3.1. Network model

We focus on the study of the multi-radio multi-channel wireless Mesh network, and MRMC - WMN network topology can be expressed as a directed graph $G(N, L)$, N denotes the set of network nodes, L denotes set of physical wireless links. $C_k = \{1, 2, 3\}$ denotes available channels set of multi-channel multi-radio wireless Mesh network. Wireless link l_{uv} denotes the assigned radios from the sending node u to the receiving node v , assigning available channel k for the link can be expressed as l_{uv}^k . In order to facilitate comprehension, the table below gives the key symbols meaning used in the paper.

Table 1. Key symbols in MRMC WMN

C_k	Available channels set	R_t	Node transmission range
L	Physical wireless links set	R_u	Radio number configured for the node u
G	Network directed graph	$C(u)$	Assigning available channels for node u
I_{uv}^k	available channel k for the link between node u and node v	N	mesh nodes set
R_s	Node carrier sense range	R_i	node interference range

Because the number of radio for each wireless Mesh node configuration is limited, assigning channel number for the wireless Mesh nodes does not exceed the number of available radios wireless Mesh nodes, at the same time in order to improve the utilization rate of spectrum, different radios should be assigned with different channels, which means all of the radios of a wireless mesh node will not be assigned with the same channel. In addition, theoretically the radios of node interface can't be more than available orthogonal channels, there is a relationship as follows:

$$|C(u)| \leq R_u \leq k, C(u) \subseteq C_k \quad (1)$$

3.2. Interference model

This paper adopts the Protocol interference Model which is commonly used. In MRMC - WMN, effective communication range of the wireless Mesh network nodes can be classified into three parts: node Transmission Range R_t , node Carrier Sensing Range R_s and node Interference Range R_i . Transmission range R_t denotes maximum range of mesh node successfully receiving signal without any interference, in general, the mesh node's transmission range is related to its transmission power, frequency. And node transmission power is proportional to the node transmission range; Carrier sensing range R_s represents maximum range of nodes detecting other carrier signal sent by other nodes; node interference range R_i denotes conflict range caused by nodes interference. In general, relationship is as follows:

$$R_t < R_i = 1.78R_t < R_s = 2.2R_t \quad (2)$$

When there is transmission requirement, link a-b and link f-g, since the mesh node b is within interference range of link f-g, but not within the communication range of node f, when link a-b and link f-g are assigned with the same channel or orthogonal channel.

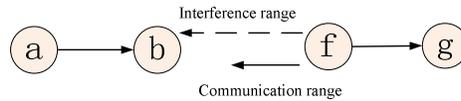


Figure 1. Relationship between interference and communication range

As shown, interference in WMNs is classified into two types: intra-flow and inter-flow interference. If different simultaneous transmissions of the same data flow interfere with each other, we call it intra flow interference. On the other hand, for simultaneous transmissions of different data flows, inter- flow interference instead. If the mutual interference of link can be assigned with different channels, then the two kind of interference can be reduced appropriately.

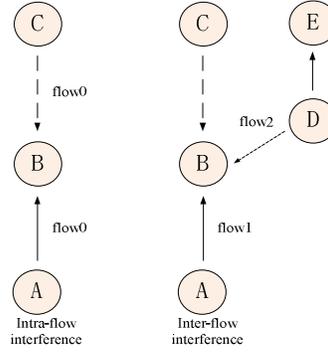


Figure 2. Intra-flow interference and inter-flow interference

4. JOINT CHANNEL ASSIGNMENT AND WEIGHTED AVERAGE EXPECTED END-TO-END DELAY ROUTING PROTOCOL

In the part, we will introduce a joint channel assignment and delay routing protocol with weight average expected end-to-end delay.

$$WAEED_{n,m}^c = (1 - \beta)EED_i + \beta IF_i \quad (3)$$

$WAEED_{n,m}^c$ denotes weighted average expected end-to-end delay from node n to node m by channel c, where EED_i denotes average end-to-end delay by channel i. As to IF_i denotes interference by channel i, plus $IF_i \geq 1$, that IF_i is closer to 1 means interference by channel i is smaller. $M_{n,i}$ denotes packets from node n to node m in total. $\sum p k_{m,i}$ denotes packets successfully received by node m. β is a weighted factor, $0 \leq \beta \leq 1$, usually β is set as 0.5.

JCWAEEED algorithm adopts contrast of the IF interference and the available channel end-to-end delay EED to choose the most appropriate channels, and in the process of channel allocation we need to follow these two principles: (1) for different radio of one node should be assigned with different orthogonal channels. (2) one radio unbounded with channel should be assigned appropriate channel in priority. While assigning channels, calculate routing metrics WAEED, then get the routing metrics collection, and select the result of channel allocation and routing with the minimum routing metric value .

4.1. Expected End-to-end delay

The end-to-end delay of a routing path is the sum of total hops delay along this path, including transmission delay along the links and queuing delay. For convenience, we define EED to represent the average end-to-end delay on the link.

To calculate routing metrics EED with the available channel i on the wireless link, each node needs to monitor the number of cache packages over the network layer waiting for MAC layer service, as well as the measure MAC layer transmission failure probability, the measurement of packet loss rate. Transmission failure may be caused by the MAC layer interference or the poor quality of channel. Link i, therefore, the end-to-end delay on the average represented as:

$$EED_i = E[T_i + Y_i] \quad (4)$$

T_i denotes transmission delay over the link i , Y_i denotes queuing delay, EED_i the average delay of a data packet over the link i . Assume that a new packet on the link i has entered a packet buffer queue, we define the average end-to-end delay as:

$$EED_i = (Q_i + 1)E[T_i] \quad (5)$$

EED_i consists of transmission delay and queuing delay. For EED_i calculation, as long as we derive the value of $E[T_i]$, then we can get the value EED_i . In particular, for CSMA protocol, transmission delay includes not only backoff time delay but also channel busy time due to the other node transmission. We define packet loss rate over the link i as p_i , define the service time as T_i , define maximum retransmission times as K , W_j denotes the j th backoff stage competition window, L denotes packet size, B denotes link bandwidth. If we ignore the constraints of backoff period, then $E[W_j] = (w_j - 1) / 2$. Therefore, average transmission delay $E[T_i]$ can be represented as follows:

$$E[T_i] = E[\text{transmissiontime}] + E[\text{backofftime}] \quad (6)$$

Plus,

$$\begin{aligned} E[\text{backofftime}] &= \sum_{k=1}^{K+1} p_i^{k-1} (1-p_i)^{I\{k < K+1\}} \sum_{j=1}^k E[W_j] \\ &= \frac{W_{\min}[1-(2p_i)^{K+1}]}{2(1-2p_i)} - \frac{1-p_i^K}{2(1-2p_i)} \end{aligned} \quad (7)$$

Hence, average transmission delay $E[T_i]$ can be denoted as follows.

$$\begin{aligned} E[T_i] &= E[\text{transmissiontime}] + E[\text{backofftime}] \\ &= \frac{L}{B} \sum_{k=1}^{\infty} k \cdot p_i^{k-1} \cdot (1-p_i) + \sum_{k=1}^{\infty} p_i^{k-1} (1-p_i) \sum_{j=1}^k E[W_j] \\ &= \frac{L}{B(1-p_i)} + \frac{W_{\min}}{2(1-2p_i)} - \frac{1}{2(1-p_i)} \end{aligned} \quad (8)$$

Multi-radio multi-channel wireless Mesh network has the characteristics of multi-hops network, considering there might be H hops along the end-to-end path, so the average end-to-end delay EED can be defined as:

$$EED = \sum_{i=1}^H EED_i \quad (9)$$

4.2. Interference calculation

We define channel interference factor as IF , which can represent interference between co-channels and different channels, therefore, for the channel i , its channel interference can be defined as:

$$IF_i = \frac{M_{n,i}}{\sum p k_{m,i}} \quad (10)$$

IF_i denotes interference using channel i , and $IF_i \geq 1$, IF_i the closer to 1 indicates the smaller interference using channel i , which can be negligible. $M_{n,i}$ denotes the total number of packages which node n sends to node m . $\sum p k_{m,i}$ denotes the total number of packets node m successfully

received from node n . We define the IF by proportional relations, which can represent interference more accurately, thus select more appropriate channel to reduce the co-channel interference.

4.3. Routing metric

From part 4.1 and 4.2 above, we can conclude the formula of the average end-to-end delay EED and channel interference factor IF, joint routing metric is expressed as:

$$WAEED_{n,m}^c = (1 - \beta)(M_i + 1) \left[\frac{L}{B(1 - p_i)} + \frac{W_{\min}}{2(1 - 2p_i)} - \frac{1}{2(1 - p_i)} \right] + \beta \frac{M_{n,i}}{\sum p k_{m,i}} \quad (11)$$

Nodes use different channels' combination, by sending HELLO packets to the neighbor node to get the information such as the queue length, link bandwidth, the total number of sent and received packets. According to one hop routing metrics WAEED collection, we will set the channel with the minimum one hop routing metric value as the best channel from this node to the next hop node. And establish routing according to the minimum joint routing metrics.

The performance of JCWAEED is compared with the average throughput, average delay and average packet loss rate. The average end-to-end delay and inter-channel interference are used as the metric of routing. The channel with the lowest inter-channel interference is selected as the optimal channel by polling, and the throughput and delay of the network are improved through delay control.

5. SIMULATION RESULTS

The simulation parameters are shown in the table below:

TABLE.2 Simulation parameter

Simulation time	100s	Orthogonal channels number	3
Simulator	NS2.35	flow number	4
Routing protocols	MRMC-AODV/JCEED/JCWAEED	transmission/interference range	250m/550m
Antenna	Omni-directional	channel bandwidth	11Mbps
Mesh nodes number	8	queue length	1000byte
Simulation area	1000m*1000m	packet loss rate	0.01
node radios number	3	flow rate	0.1 ~ 2 Mbps

5.1. Network throughput

When data flow rate is less than 0.3Mbps, the paper proposed JCWAEED algorithms, JCEED algorithm and the AODV-MRMC algorithm are similar on the average throughput performance.

When data flow rate is over 0.4Mbps, the average JCWAEED algorithm throughput outperforms JCEED algorithm at the same flow rate. JCWAEED works in high data flow rates because of the improving routing metrics, and it can undertake higher data flow rates.

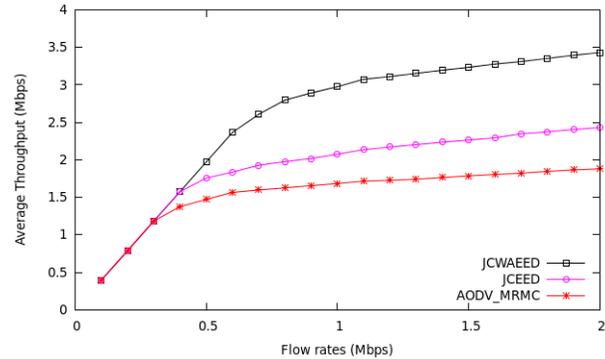


Figure 3. Comparison of average throughput performance in different data flow rates

5.2. End-to-end delay

When data flow rate is less than 0.6Mbps, the paper proposed JCWAEED algorithm, JCEED algorithm and the AODV - MRMC algorithm are similar on end-to-end delay performance which is negligible. When data flow rate is around 0.7Mbps, the average end-to-end delay of three algorithms increases rapidly. When data flow rate is over 0.7Mbps, the end-to-end delay of JCWAEED algorithm outperforms JCEED algorithm at the same flow rate. JCWAEED algorithm has the minimum average end-to-end delay.

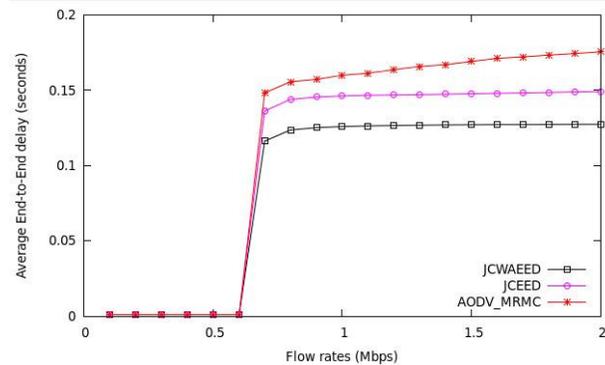


Figure 4. Comparison of end-to-end delay performance in different data flow rates

5.3. Packet loss

When data flow rate is less than 0.6Mbps, the paper proposed JCWAEED algorithm, JCEED algorithm and the AODV - MRMC algorithm are similar on packet loss performance, which is negligible. When data flow rate is over 0.6Mbps, the packet loss rate of JCWAEED algorithm outperforms JCEED algorithm at the same flow rate. JCWAEED algorithm has the minimum packet loss.

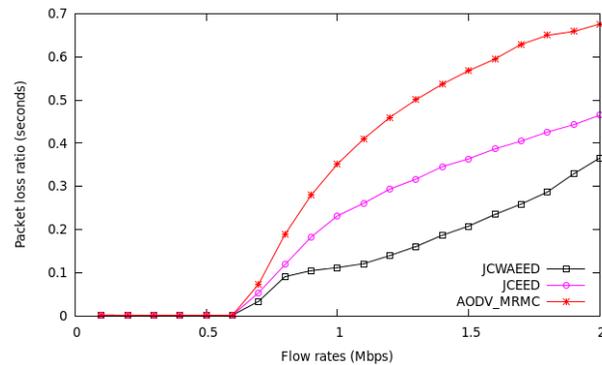


Figure 5. Comparison of packet loss ratio performance in different data flow rates

6. CONCLUSION

In this paper, we propose an effective joint channel assignment and routing optimization protocol called JCWAEED based on the network model and interference model, considering end-to-end delay control, interference suppression with factor IF . Additionally, we give the exact calculation formula of transmission delay and queuing delay. Simulation results demonstrate that JCWAEED outperforms other joint design routing protocols in terms of throughput, end-to-end delay and packet loss rate under the influence of different data flow rates.

REFERENCES

- [1] Wang J, Shi W, Jin F. On channel assignment for multicast in multi-radio multi-channel wireless mesh networks: A survey[J]. 2015, 12(1):122-135.
- [2] D. S. De Couto, D. Aguayo, J. Bicket, and R. Morris. A highthroughput path metric for multi-hop wireless routing[J]. *Wireless Networks*, vol. 11, no. 4, pp. 419–434, 2005.
- [3] R. Draves, J. Padhye, and B. Zill. Routing in multi-radio, multi-hop wireless mesh networks[C]. *Conf. on Mobile Computing and Networking*, 2004, pp. 114–128.
- [4] L. Ma and M. K. Denko, “A routing metric for load-balancing in wireless mesh networks,” in *Proc. of 21st Int. Conf. on Advanced Information Networking and Applications Workshops*, vol. 2, 2007, pp.409–414.
- [5] R.F. Ali, A.K. Kiani and A.A. Pirzada. Load Dependent Dynamic Path Selection in Multi-Radio Hybrid Wireless Mesh Networks[C]. *IEEE Wireless Communications and Networking Conference WCNC*, pp. 2044-2049, 2014.
- [6] Wu H, Yang F, Tan K. Distributed channel assignment and routing in multiradio multichannel multihop wireless networks *Selected Areas in Communications*, IEEE Journal on 24(11) 2006 pp. 1972-83.
- [7] Pham N T, Hwang W J, Joint Disjoint Path Routing and Channel Assignment in Multi-Radio Multi-Channel Wireless Mesh Networks, in *IEEE Vehicular Technology Conference, 2008 VTC 2008-Fall IEEE 68th*, 2008, pp. 1-5.
- [8] Gardellin V, Das S K, Lenzini L, et al. G-PaMeLA: A divide-and-conquer approach for joint channel assignment and routing in multi-radio multi-channel wireless mesh networks[J]. *Journal of Parallel & Distributed Computing*, 2011, 71(3):381-396.

- [9] Capone A, Carello G, Filippini I, et al. Routing, scheduling and channel assignment in Wireless Mesh Networks: Optimization models and algorithms[J]. *Ad Hoc Networks*, 2010, 8(6):545-563.
- [10] Sun W, Cong R, Xia F. R-CA: A Routing-Based Dynamic Channel Assignment Algorithm in Wireless Mesh Networks, in *IEEE Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing (UIC/ATC)*, 2010 7th International Conference on, 2010, pp. 228-32.

REDUCING FREQUENCY OF GROUP REKEYING OPERATION

YunSuk Yeo, Sangwon Hyun and Tai-Myoung Chung

Computer Engineering,
Sungkyunkwan University, Suwon, South Korea

ABSTRACT

In the past, Ad-hoc networks were used in limited areas which require secure group communication without Internet access, such as the army or emergencies. However, Ad-hoc networks currently are widely used in variety applications like group chat, smart applications, research testbed etc. Ad-hoc network is basically group based network in the absence of access point so it is prevalent to provide group key approach to prevent information leakage. When we use group key approach, we need to consider which group key management method is the most suitable for the architecture because the cost and frequency of the rekeying operation remain as an unresolved issue. In this paper, we present analysis about existing group key management solutions for Ad-hoc network and suggest a new approach to reduce frequency of the rekeying operation.

KEYWORDS

Rekeying operation, Group key management, ad-hoc networks, Frequency of rekeying, Time-driven method

1. INTRODUCTION

There exist many group key management (GKM) methods to provide secure group communication. In general, two essential keys are used in the methods: a *traffic encryption key* (TEK) and a *key encryption key* (KEK). As illustrated in Fig. 1, all the group members share the same TEK, while each of them has a private KEK that is only shared with the group controller (GC) or key server (KS). When a group member wants to send a message M to the entire group, it broadcasts M encrypted with TEK, denoted by $\{M\}_{\text{TEK}}$.

Whenever a member leaves or joins the group, a new TEK must be generated and securely distributed to all the group members. Such process to update the TEK is commonly referred to as *rekeying*. A simple approach to rekeying is that the GC/KS unicasts each member the new TEK encrypted with the member's KEK so that the member can retrieve the new TEK by decrypting the received message with its KEK. However, this approach requires $O(n)$ cost in terms of the number of messages, where n denotes the number of group members, and it is well known that this incurs "*broadcast storm*" problem; A broadcast storm occurs when a network system is overwhelmed by continuous multicast or broadcast traffic so it causes the whole network to melt down and lead to the failure of network communication. In addition, frequent rekeying in highly dynamic groups can also incur the broadcast storm problem.

Extensive research has been done to reduce the overhead of rekeying in two major research directions. The first category of research focuses on reducing the cost of each re-keying operation [1, 3, 5, 7, 9, 13, 14]. Especially, the local key hierarchy (LKH) method [1] reduces the cost into $O(\log n)$ by utilizing a binary key tree. However, in these approaches, the authors considered only the cost of rekeying without care about the frequency of rekeying would make their solution

incomplete so the solution cannot cope with dynamic and scalable groups. In ad-hoc network, very frequent rekeying operation will eventually cause a broadcast storm.

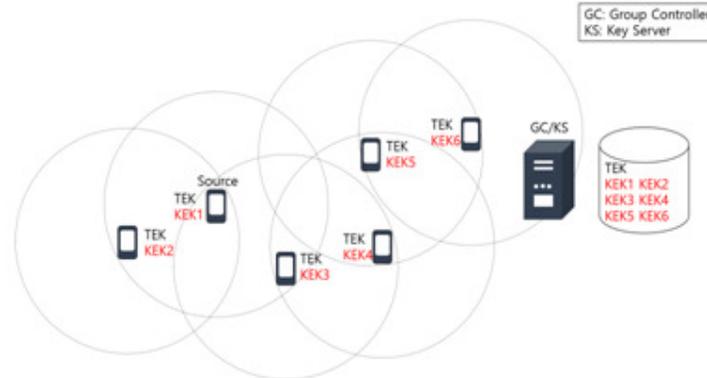


Figure 1. Ad-hoc Group Key Management

The second category of research focuses on reducing the frequency of re-keying [2, 4]. Especially, Kronos and Iolus adopt a time-driven method, which performs the rekeying operation periodically instead of performing it whenever member changes occur [2, 4]. The authors demonstrated that the time-driven method could considerably reduce the overhead of re-keying even though such limitations as no guarantee of forward secrecy and delays in adding new members. However, for highly dynamic groups, the cost of the time-driven re-keying could be $O(n)$ as well [2]. In ad-hoc networks, $O(n)$ cost usually causes “broadcast storm” situation due to centralized architecture (i.e., a GC/KS should generate n messages in worst-case). For these reasons, we cannot apply this method directly to GKM in ad-hoc networks.

In this paper, we suggest a composition of the time-driven method and GKMPAN method [3], which is one of the main GKM approaches for ad-hoc networks, to eliminate $O(n)$ problem in the time-driven method. GKMPAN is novel method of distributed GKM in ad-hoc network so it can remove bottleneck and $O(n)$ problems of the time-driven method. We modify the rekeying process of GKMPAN to facilitate the installation of the time-driven model in ad-hoc environment and to solve key exhaustion problem of GKMPAN. With this approach, we can achieve following works:

- Reduce the frequency and the cost of rekeying operation with reliable security strength
- Conduct rekeying operation without “broadcast storm” with distributed architecture
- Solve key exhaustion problem of GKMPAN

The organization of the remainder of this paper is as follows. In Section 2, we explain basic Local Key Hierarchy (LKH) [1], GKMPAN [3] and the cost analysis of the time-driven LKH method. In Section 3, we propose the expanded GKMPAN. Section 4 evaluates expanded GKMPAN and Section 5 present related works. Finally, Section 6 presents our conclusions and future works.

2. BACKGROUND

2.1. Local Key Hierarchy (LKH)

In the LKH method [1], a GC/KS constructs and manages a binary key tree T illustrated in Fig. 2. The root of T corresponds to the TEK. All the remaining nodes hold KEKs ($= K_n$) and a leaf node is associated with a group member ($= U_i$). A member U_i holds the TEK and all the KEKs associated with the nodes from K_i to the root. For instance, U_8 holds K_8, K_{78}, K_{5678} , and TEK.

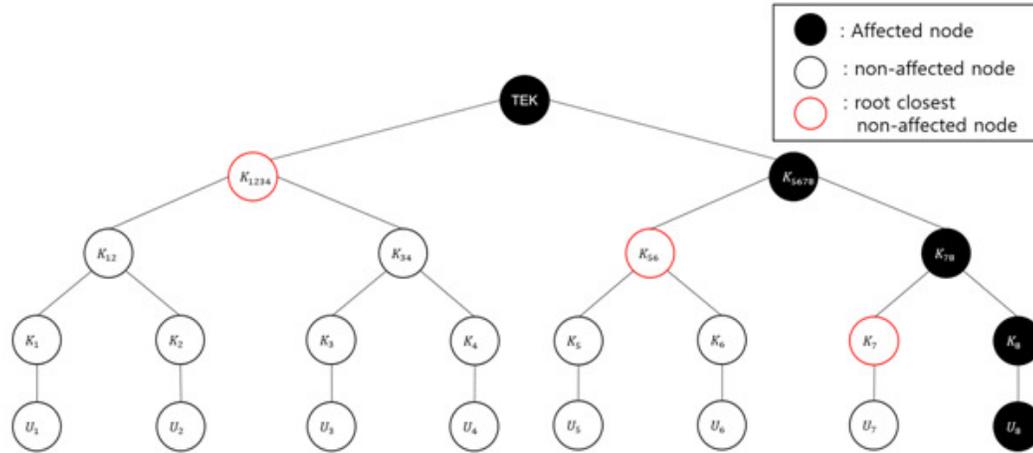


Figure 2. LKH Leaving Example

If U_8 leaves the group, then K_8 , K_{78} , K_{5678} , and TEK should be changed to ensure forward secrecy. We call the nodes whose key must be changed as “affected nodes.” To change these keys, the GC/KS only needs to send three different types of rekeying messages to the groups; $\{TEK\}_{K_{1234}}$ for the members (U_1 , U_2 , U_3 and U_4), $\{TEK\}_{K_{56}}$ for the members (U_5 and U_6), and $\{TEK\}_{K_7}$ for the member (U_7).

If U_8 joins the group, then K_8 , K_{78} , K_{5678} , and new TEK should be generated to ensure backward secrecy. The GC constructs and sends rekey messages to the users. A rekey message contains one or more encrypted new key(s), and a user needs to decrypt it with appropriate keys in order to get the new keys.

By using the binary key tree, the LKH scheme effectively reduces the cost of rekeying from $O(n)$ to $O(\log(n))$. We have applied cost-effective LKH scheme to effectively apply time-driven to ad-hoc networks and evaluated the cost.

2.1.1. Cost Evaluation of Time-Driven LKH

In Fig. 2, we count K_{1234} , K_{56} , and K_7 to calculate the cost of rekeying process because each group member can securely obtain a new group key by encrypting/decrypting message with the key of the root nodes of sub-trees composed of non-affected nodes; K_{1234} , K_{56} , and K_7 . Similarly, we can calculate rekeying cost of time-driven method with LKH by counting the root nodes of sub-trees composed of non-affected nodes with breadth-first search (BFS). These root nodes are referred to as root of subtree (RS). Considering the time-driven case, we denote the number of members whose state changed within a period as m and the size of the group as n . For the simplicity, we assume that leaving and entering events are occurred in the same frequency, besides m and n are in the form of the power of two: 2, 8, and 64.

The Fig. 3 shows how to calculate the maximum number of rekeying messages in time-driven method. In time-driven method, there are many leaving and entering events during a cycle, so the tree has multiple affected branches. If a member U_i leaves the group, $(\log n)$ nodes are affected, because a member U_i has all the keys along the path from its leaf to the root. To get the maximum cost of rekeying, we treat the case that leaving nodes do not overlap, as is possible in the tree. In this situation, the rows up to a height of $\log m$ from the root have fewer nodes than m in that rows; therefore, they do not have any unaffected nodes, because m nodes leave the group and a

leaving node follows one path from the leaf to the root in the tree T . In other words, all the nodes in the rows below a height of $\log m$ are affected by leaving or entering events.

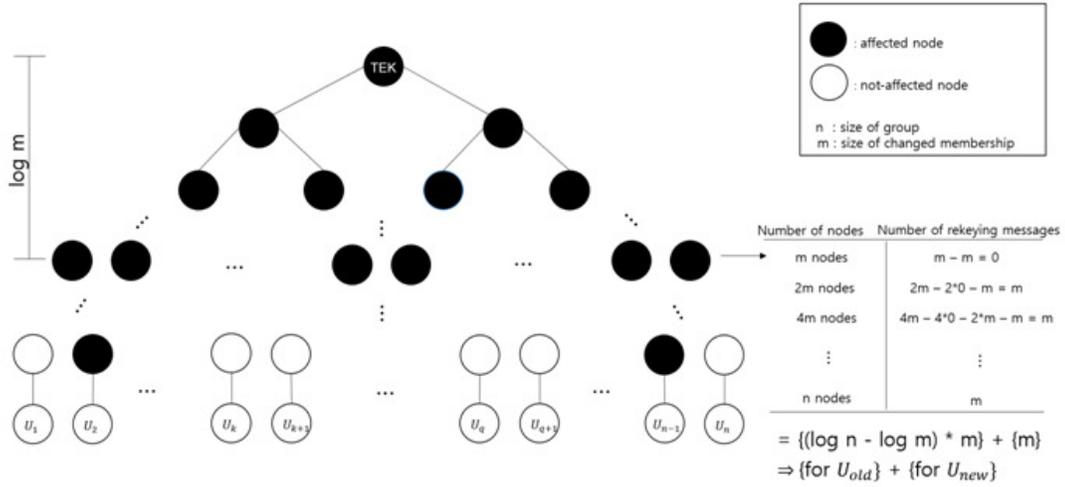


Figure 3. Maximum number of rekeying messages in the time-driven approach

Now, it is possible to find the RS nodes in each row by subtracting both the affected nodes and those already covered by the elements from parent RS nodes.

$$RS = E_{row} - E_{affected} - E_{covered_list} \quad (E: \text{node Element})$$

Once the RS nodes in a row have been found, the child nodes of them are inserted into the covered list, which is used to find the RS nodes of the next row. This sequence of operations is described right-below edge in the Fig. 3. For entering events, additional m messages are needed because each newbie should respectively receive its secure KEK. Hence, the total number of rekeying messages is

$$\text{Cost} = (\log n - \log m) \times m + m.$$

With this formula, if $n = 1024$ and $m = 16$, the number of rekeying messages exceeds 100 during a cycle, which means that a change of only 2% will cause a “broadcast storm” due to centralized model. The general equation for any form of m , which is not exactly a power of 2, is

$$\text{Cost} = (\log n - \text{floor}(\log m)) * m + m + 2 * (2^{\text{floor}(\log m)} - m).$$

2.2. GKMPAN

In GKMPAN [3], all node obtains a distinct subset of keys out of a large key pool from the Key Server (KS) and these keys are used as Key Encryption Keys (KEKs) for delivering group keys (TEK). This method has the advantage in that we can ignore the process of securely sharing KEKs between a GC and new member because it is already shared in offline mechanism. Moreover, this model uses hop-by-hop approach to eliminate bottleneck problem.

When a leaving event occurs, the GC broadcasts the list of revoked keys possessed by a leaving node, so that this revoked keys will not be used anymore in this group. The GC generates and sends a new TEK encrypted with K_m , the key is possessed by the maximum number of remaining nodes, to ensure forward security. With this broadcast message, however, some of the nodes cannot decrypt it because they do not contain K_m in their private key list. To fully deliver new TEK to its child nodes, the GC should resend the new TEK encrypted with other keys which are shared between GC and some of the excluded nodes based on a delivery tree.

When a node receives new TEK encrypted with K_m or other KEKs, then, it continuously sends new TEK to its child nodes in a manner like the GC along the delivery tree. Each node only cares the child nodes like an independent key server. Then, we can infer that the maximum overhead of each node is the number of child nodes of itself on the delivery tree.

We can apply the time-driven method to this approach because it eliminates the bottleneck problem, which could be found in the centralized model [5]; however, there is a problem that the number of available keys (KEKs) decreases as the number of removed nodes increases. GKMPAN [3] asserts that the problem can be solved using big l and small m . However, for performance, we should choose small l and large m , because if a small l and a large m are used, a key is shared with many nodes so that the node can share the new TEK with a very small number of messages

The biggest difference between GKMPAN and our method is that GKMAPN uses the pre-allocated KEKs to share the new TEK and we use the old TEK that was used in the previous rekeying stage. The advantage of our approach is that the dependency between performance and key exhaustion is eliminated. We excluded the keys held by the revoked nodes only during the rekeying stage of one cycle. When node i leaves the group, we can use the remaining keys in the rekeying stage, except for the keys that node i had in the key pool. However, in the next rekeying stage, the keys that were excluded can also be used.

3. OUR APPROACH (EXTENDED GKMPAN)

Table 1. Terminology

GC/KS	Group Controller/Key Server
KEK	Key Encryption Key
TEK	Transmission Encryption Key (group key)
l	Size of a key pool
K_m	The key is possessed by the maximum number of remaining nodes in network
m	The number of keys in the possession of a group member (= remaining node)
K_b	The key owned by node b
L_{leave}	List of leaving node at current stage
L_{join}	List of joining node at current stage

To solve the $O(n)$ and bottleneck problem in the time-driven method, we propose a new rekeying strategy based on GKMPAN [3], which uses a pre-distributed set of keys. The main idea of our approach is to securely pass a new TEK encrypted with the old TEK “hop-by-hop”; $\{TEK\}_{old_TEK}$. Before we explain details, the Table 1 describes terminologies which are used in last of the paper.

3.1. Model Description (Extended GKMPAN)

As described above, our model uses time-driven enabled GKMPAN approach to complement weak point of time-driven method. Our model is also hop-by-hop process so we assume that each node can easily acquire the information about the nodes within its one-hop distance. We also assume all group members have pre-distributed set of keys from key pool l , which will be used as KEKs. With the GKMPAN method, it is possible to know which keys a member possesses with only the id of the member. Because of this, a node can figure out which keys the surrounding nodes have. Moreover, asymmetric key mechanism is used, so the sender only needs to know the

public key of a key of the receiver. In this approach, we do not need to worry about which keys overlap between the sender and receiver because the all members have capability to encrypt a message with any key in l (i.e., a sender knows all the public keys of all the KEKs).

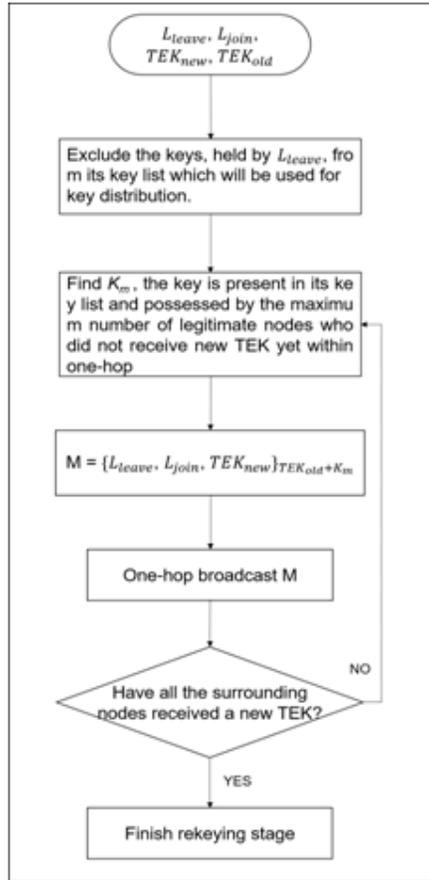


Figure 4. Flow Chart of a Rekeying Stage in Extended GKMPAN

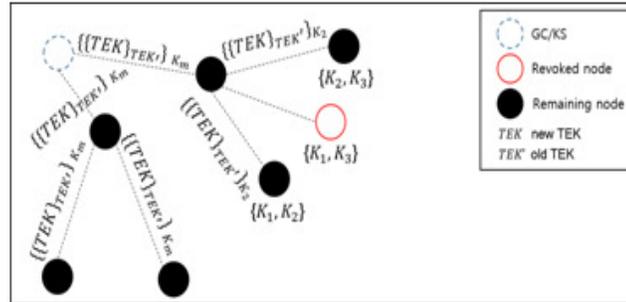


Figure 5. Hop-by-hop Encryption

After completing one cycle of the time-driven method, the GC obtains both a list of revoked members and newly joined members list. Then, the rekeying process starts and the GC securely sends a new group key (TEK), the revoked list, and the newly joined list to its surrounding nodes. The node receiving the rekeying message becomes the distributor and starts the process represented in Fig. 4. The legitimate nodes contain remaining nodes and the members in L_{join} .

As depicted in the Fig. 4, the first step of secure rekeying process is to figure out which keys can be used to distribute new group key by excluding the keys held by L_{leave} . The second step is to find K_m that the key is possessed by the maximum number of legitimate nodes who did not receive new TEK yet within one-hop. After finding K_m , the distributor sends a rekeying message encrypted with both the old TEK and K_m to the all nodes within one hop, $\{L_{leave}, L_{join}, TEK_{new}\}_{TEK_{old} + K_m}$. If all the neighboring nodes have received a new TEK, the rekeying stage ends. If not, the distributor starts again from the second step by finding K_m , which is possessed by the maximum number of legitimate nodes who did not receive new TEK yet within one-hop.

The role of K_m is to hide the new TEK from the nodes that are revoked at current stage (i.e., L_{leave}), because they know the old TEK. K_m is selected from the keys possessed by legitimate nodes except for the revoked nodes so it can hide new TEK from the revoked nodes.

The reason for encrypting the new TEK with the old TEK is to solve the key exhaustion problem of GKMPAN. The nodes without knowledge of the old TEK are unable to retrieve the new TEK; thus, it is unnecessary to consider whether previously eliminated nodes have some keys in their key pool because they do not have knowledge about the old TEK. So, we only need to care about the nodes which are revoked at current cycle (i.e., The KEKs that were excluded from the rekeying stage can be used in the next cycle).

For instance, in the rekeying stage, node v uses the old TEK and K_b to share a new group key. K_b is owned by node b which was revoked previously. Then, although revoked node b knows the K_b , it cannot decrypt the new group key encrypted with both the K_b and the old TEK (i.e., $\{\text{TEK}_{\text{new}}\}_{\text{TEK}_{\text{old}} + K_b}$) because it does not know the old TEK which is used in current time-driven cycle. From this point of view, we do not need to delete the K_b from the key pool even though the node b leaved the group. It means that the lifetime of the keys in the key pool is semi-permanent.

This approach has several advantages. First, in the best case each node of a group needs to send only one multicast message to propagate the new TEK and in the worst case each node is required to send a message to each of the surrounding legitimate nodes (i.e., remaining node + L_{join}). Second, this approach enables us to select a large l and small m , because we have already solved the problem of key exhaustion. By encrypting the new TEK with the old TEK, the keys owned by each group member are simply used to hide the new TEK from the nodes who have left current rekeying stage.

The diagram in the Fig. 5 illustrates a simple case. In short, the GC/KS sends a new TEK encrypted with the old TEK to surrounding nodes, then they subsequently forward the new TEK to their neighbors securely by encrypting it with K_m , which is K_2 in this example. Here, we do not consider other processes, such as key distribution or message verification, because these properties can be referenced from GKMPAN [3].

4. PERFORMANCE EVALUATION

We evaluated our proposed method by calculating the number of messages a node v should send to share a new group key in four different cases. In the Fig. 6, n , l , and m , respectively, denote the number of surrounding nodes of a node v , the size of the key pool, and the number of keys possessed by a node. In evaluation, we assume that there is no collision. Each node evenly has random leaving and joining probability (0.1 to 0.9). We count all messages in the group and compute average message counts from a node v . Moreover, in our and GKMPAN approach, some nodes may not get a new TEK because the all keys possessed by a node could be excluded from current rekeying stage due to the revoked nodes. So, we also evaluate failure ratio according to n , l and m .

In Fig. 6, The red line shows the failure ratio based on the right-side y axis, and the black line shows the number of required messages based on the left-side y axis. As shown in *a* and *b*, they depict the number of messages and failure ratio when the number of neighbor nodes (n) and the size of the key pool (l) are fixed. When m is small, the number of necessary messages decreases as the value of m increases. However, as the value of m increases from a specific value, the number of required messages increases. In *c* and *d*, they depict the number of messages and failure ratio when the number of neighbor nodes (n) and the key set size of each node (m) are

fixed. The results of c and d are similar to those of a and b . As the value of l increases before the specific value, the number of necessary messages decreases, but it increases thereafter.

For a large m and small l , the failure ratio is small, and for a small m and large l , the failure ratio is large. In addition, a high probability of failure ratio means that legitimated nodes are less likely to have the same keys, and this result can be seen in the above experiment.

The efficiency of the algorithm depends on how many legitimate nodes have the same keys. If m is sufficiently small compared to l , we can see that the number of keys needed is reduced because the number of keys that overlap each other increases. If m is large enough for l , the number of keys that a node has is large, so there are more keys to be excluded for revoked nodes in the key pool. This reduces the probability that the legitimated nodes will have the same key, which makes the efficiency worse.

In the same simulation environment, when the population is 100, it is shown that a central node of the time-driven LKH sends an average of 72 ~ 75 messages, and in case of 200, it sends close to 150 messages. It means ($n * 3/4$) messages are required to share a TEK in time-driven LKH in normal cases. LKH is affected by the total number of nodes, but our approach is much more effective than LKH because it is only affected by the number of neighbor nodes, not the whole.

5. RELATED WORKS

In Ad-hoc networks, we can categorize the GKM into three sub groups; Centralized approach, Decentralized approach, and Distributed approach. In Centralized approach, there is only one GC/KS. This server is responsible for the generation, the distribution and the renewal of the group key (TEK). Centralized approach is suffered from the $O(n)$ problem and bottleneck problem. GKMPAN [3], CKDS [12], Kaya et al. solution [13] belong to this group. In Decentralized approach, it divides the group into several sub groups to relieve $O(n)$ problem and bottleneck problem. However, this approach requires several decryption and re-encryption operations of multicast message, when it passes from a sub-group to another. ILOUS [4], AKMP [9], and BLADE [5] belong to this group.

The one of Decentralized approaches is BLADE [5]. Its basic idea is to divide the multicast group dynamically into clusters. Each cluster is managed by a local controller which shares with its local members a local cluster key. The multicast flow is encrypted by the source with the traffic encryption key TEK and sent in multicast to all the group members. The source of the group and the local controllers from a multicast GLC (Group of Local Controllers) share beforehand a session key called KEK_{CCL} . Each new local controller has to join this group and receive the session key KEK_{CCL} from the source of the group, encrypted with its public key. It could support mobility and insure energy efficiency.

In Distributed approach, all group members cooperate and generate TEK, to establish secure communications between them. This approach eliminates the bottleneck problem, but generally it suffered from $O(n)$ problem. Our approach is also included in this category without $O(n)$ problem with time-driven method. Several papers propose the solution for this category [9, 10, 14, 15, 16]. Chiang et al. [16] proposes a GKM protocol for MANETs based on GPS measures and on the GDH (Group Diffie Hellman) [11]. In this protocol, during protocol initialization, each node in the ad hoc network, floods its GPS information and its public key to all the others nodes, although the authors assume that the protocol does not rely on any certification authority. Using the GPS information received from others nodes, each group member can build the network topology. When a source wants to multicast the data flow to the group members, it computes the minimum multicast tree, based on the Prüfer algorithm and then sends message with GDH keys.

6. CONCLUSIONS

The Ubiquitous Network Society suggests a world in which many ad-hoc environments can be applied. Moreover, the number of applications, such as Vehicular ad hoc networks (VANETs) or Smartphone ad hoc networks (SPANs), relying on the ad-hoc multicast technique is increasing. However, no efficient GKM method has been suggested yet so, in this paper, we propose an improved GKM scheme with the aim of enhancing the efficiency GKM in Ad-hoc network. The dominant factor of our scheme is reduced frequency of rekeying operation, whereas other schemes were more concerned with the cost. Considering the feasibility of GKM, we believe our method is more practical than others. In the future, we devise modified version of extended GKMPAN to apply this to wired networks as distributed GKM.

ACKNOWLEDGEMENTS

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2016-0-00078, Cloud based Security Intelligence Technology Development for the Customized Security Service Provisioning).

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0020210).

REFERENCES

- [1] C.K. Wong, M. Gouda, and S. S. Lam, Secure Group Communications Using Key Graphs, ACM SIGCOMM (1998).
- [2] S. Setia, S. Koussih, S. Jajodia, and E. Harder, Kronos: A Scalable Group Re-Keying Approach for Secure Multicast, IEEE Symposium on Security and Privacy, 2000.
- [3] Sencun Zhu, Sanjeev Setia, Shouhuai Xu, Sushil Jajodia, GKMPAN: An Efficient Group Rekeying Scheme for Secure Multicast in Ad-Hoc Networks, IEEE, 2004.
- [4] Suvo Mittra, Iolus: A Framework for Scalable Secure Multicasting, ACM Sigcomm, 1997.
- [5] Mohaned-Salah Bouassida, Isabelle Chrisment, and Olivier Festor, Group Key Management in MANETs, International Journal of Network Security, Vol.6, No.1, PP.67-79, Jan. 2008.
- [6] Yacine Challal, Hamida Seba, Group Key Management Protocols: A Novel Taxonomy, International Journal of Information Technology, Vol.2, 2005.
- [7] Yan Sun, Wade Trappe, and K. J. Ray Liu, A Scalable Multicast Key Management Scheme for Heterogeneous Wireless Networks, IEEE/ACM, Vol. 12, No 4, 2004
- [8] W.T. Li, C.H. Ling, M.S. Hwang, Group Rekeying in Wireless Sensor Networks, International Journal of Network Security, Vol, 16, No.6, 2014
- [9] H. Bettahar, A. Bouabdallah, and Y. Challal, "An adaptive key management protocol for secure multicast," in 11th International Conference on Computer Communications and Networks ICCCN, Florida USA, Oct. 2002.
- [10] W. Diffie and M.E. Hellman, "New directions in cryptography," IEEE Transactions on Information Theory, vol. 22, no. 6, pp. 644-654, 1976.

- [11] I. Ingemarson, D. Tang, and C. Wong, "A conference key distribution system," *IEEE Transactions on Information Theory*, pp. 714-720, Sep. 1982.
- [12] M. Moharrun, R. Mulkalamala, and M. Eltoweissy, "Ckds: An efficient combinatorial key distribution scheme for wireless Ad Hoc networks," in *IEEE International Conference on Performance, Computing and Communications (IPCCC'04)*, pp. 631-636, Apr. 2004.
- [13] T. Kaya, G. Lin, G. Noubir, and A. Yilmaz, "Secure multicast groups on Ad Hoc networks," in *Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 94-102, 2003.
- [14] S. Rahul, and Hansdah. "An Efficient Distributed Group Key Management Algorithm". In *Proceedings Tenth International Conference on Parallel and Distributed Systems*, 2004. ICPADS 2004, pages 230 -237, California.
- [15] L.R. Dondeti, S. Mukherjee, and A. Samal. "Comparison of Hierarchical Key Distribution Schemes". *IEEE Globcom Global Internet Symposium*, 1999.
- [16] T. Chiang and Y. Huang, "Group keys and the multicast security in Ad Hoc networks," in *Proceedings of the 2003 International Conference on Parallel Processing Workshops*, pp. 385, 2003.

CLUSTERING HYPERSPECTRAL DATA

Arwa Alturki and Ouiem Bchir

Department of Computer Science,
King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Spectroscopy or hyperspectral imaging consists in the acquisition, analysis, and extraction of the spectral information measured on a specific region or object using an airborne or satellite device. Hyperspectral imaging has become an active field of research recently. One way of analysing such data is through clustering. However, due to the high dimensionality of the data and the small distance between the different material signatures, clustering such a data is a challenging task. In this paper, we empirically compared five clustering techniques in different hyperspectral data sets. The considered clustering techniques are K-means, K-medoids, fuzzy C-means, hierarchical, and density-based spatial clustering of applications with noise. Four data sets are used to achieve this purpose which is Botswana, Kennedy space centre, Pavia, and Pavia University. Beside the accuracy, we adopted four more similarity measures: Rand statistics, Jaccard coefficient, Fowlkes-Mallows index, and Hubert index. According to accuracy, we found that fuzzy C-means clustering is doing better on Botswana and Pavia data sets, K-means and K-medoids are giving better results on Kennedy space centre data set, and for Pavia University the hierarchical clustering is better.

KEYWORDS

Image Processing, Hyperspectral Imaging, Imaging Spectroscopy, Clustering, FCM, K-means, K-medoids, hierarchical, DBSCAN

1. INTRODUCTION

Hyperspectral imagery (HSI) is a main tool of remote sensing applications. As a signal is transmitted, reflected, and scattered again by interacting with the various components of the atmosphere and surface; the reflectance spectra analysis may allow recognition or quantification of the materials. It can be used in several applications such as environmental, geographic, and military applications [1, 2].

HSI is a three-dimensional data cube that containing the values of the radiation that has been collected over an area in a wide range of wavelengths. It is perceived as a high resolution image with hundreds of wavelength (spectral) bands[1, 3].

HSI provides more accurate and detailed information than normal colored image. When HSI captures hundreds or thousands of spectral channels, the normal colored image captures only three: red, blue, and green [3]. In fact, each pixel in the HSI is a spectrum with continuous values to precisely represent an object in a natural scene. Thus, each pixel in a hyperspectral image is a high dimensional vector that represents the radiance of a specific material. It can be considered as its signature.

Clustering is a powerful machine learning tools that allows analysing the content of HSI. Although the signature of each material is unique, two adjacent materials can have very similar signatures. Due to the characteristics of the hyperspectral data, clustering this data is challenging. In this work, we aim to compare empirically several existing clustering techniques in four hyperspectral data sets.

2. CLUSTERING TECHNIQUES

Number of clustering techniques have been applied on HIS [4]. Depending on the spectral content information, clustering methods such as k-means[5, 6], K-medoids [7, 8], fuzzy C-Means (FCM) [9, 10], hierarchical [11, 12], and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [13] are highly exploited [14]. To further enhance the clustering performance, the spatial context information is incorporated with the spectral information to enhance the clustering performance [15]. In the following we describe the considered clustering algorithms in some details.

2.1. K-means Clustering

K-means is one of the simplest and well known clustering techniques that aims to partition N observations into K clusters [5]. The algorithm starts by randomly choosing K centroids, where one centre for each cluster. Let $X = \{x_1, x_2, \dots, x_N\}$ be the set of observations that we want to clusters, and $C = \{c_1, c_2, \dots, c_K\}$ be the initial centroids set, the following two steps are repeatedly visited [6]:

- Assignment, and this can be accomplished by: (1) calculate the mean (distance) from the observation and each cluster centre, (2) assign the observation to the cluster with minimum mean among all clusters. The algorithm tries to minimize the mean (total squared error) depending on Equation (1):

$$I = \sum_{i=1}^K \sum_{j=1}^N \|x_j - c_i\|^2 \quad (1)$$

- Update, by recalculating the clustercentres according to Equation (2):

$$c_i = \left(\frac{1}{m_i}\right) \sum_{j=1}^{m_i} x_j \quad (2)$$

Where m_i is the number of observations in the i^{th} cluster. The algorithm stops when the results converge and the observations are no longer reassigned to other clusters. K-means clustering is a fast one with time consumption equals to $\theta(NKdt)$, where d is the dimension of each observation and t is the number of iteration needed to converge [5]. It is important to note that, K-means gives the best results when the data is well separated.

2.2. K-medoids Clustering

A clustering algorithm that is very similar to K-means, but with small differences. Unlike K-means, K-medoids tries to minimize the total dissimilarities between the observations in a cluster and its medoid, which make K-medoids more robust to noise [7]. A medoid of cluster is an observation that have the minimal dissimilarity average to all other observations [8]. Similar to K-means, the initial medoids are randomly selected then the assignment and updating step are processed repeatedly. $X = \{x_1, x_2, \dots, x_N\}$ and $M = \{m_1, m_2, \dots, m_K\}$ be the set of observations and

medoids respectively, the difference between K-means and K-medoids appears in the update step as the following:

- Swap between medoid m_i and observation x_j in the i^{th} cluster.
- Recalculate the dissimilarity average between x_j and other observation in the cluster.
- If the resulted dissimilarity is larger than before swapping, return to the previous medoid.

2.3. Fuzzy C-Means (FCM)

Unlike other clustering techniques, an observation can belong to more than one cluster with FCM clustering [9, 10]. At the end of the clustering process, each observation is given a degree of membership to each cluster, where the degree is computed based on specific similarity measure [16]. Let $X = \{x_1, x_2, \dots, x_N\}$ be the observations that we want to be partitioned into K cluster and m is the clusters overlapping scaler, the algorithm assigns a random degree for each observation with each cluster. After that, the following steps are repeatedly processed:

- For each cluster i , find its centroid by calculating the mean of all observations weighted by the degree of membership U to the i^{th} cluster. This can be summarized in Equation (3):

$$c_i = \frac{\sum_{j=1}^N u_{ji}^m \cdot x_j}{\sum_{j=1}^N u_{ji}^m} \quad (3)$$

- Re-compute the degree of membership for all observations of all clusters as in Equation (4):

$$u_{ji} = \frac{1}{\sum_{k=1}^K \left(\frac{\|x_j - c_i\|}{\|x_j - c_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

The above two steps are repeated until the algorithm is converging, and the change of membership degrees U in two consecutive iterations is no more than a predefined threshold. FCM clustering is a good choice when the data set is overlapped and not well separated.

2.4. Hierarchical Clustering

A clustering algorithm that builds a hierarchy of clusters based on two approaches: agglomerative and divisive [11, 12]. Agglomerative approach (also known as bottom-up) treats each observation as a cluster and recursively merge (agglomerate) pairs of clusters until all of them are merged into one cluster. On the other hand, divisive hierarchical clustering (also known as top-down) starts by putting all observations in one cluster and recursively split (divide) the clusters until each observation represent one cluster. The hierarchical clustering results usually presented in a dendrogram graph.

In order to determine which clusters to merge (or split); we need to measure the dissimilarity between observations by calculating the distance between them. There are several metrics of distance, which are: Euclidean, squared Euclidean, Manhattan, maximum, and Mahalanobis. Then, the distance metric itself is used by the linkage criteria to measure the dissimilarity between two clusters. Here are some of the well-known linkage criteria: maximum (complete) linkage, minimum (single) linkage, and average (mean) linkage.

2.5. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN discovers the clusters (of arbitrary shapes) and noise in spatial database (points) based on the density [13]. It requires determining two parameters, which are: the maximum distance between two points to be neighbours (*Eps*) and the minimum number of points to form a cluster (*MinPts*). Then, DBSCAN starts with arbitrary un-visited point P by retrieving its neighbours, i.e. points where the distance between P and them is less than or equals to *Eps*. Then, if the number of neighbour is equal to or larger than *MinPts*, a new cluster is created; otherwise P labelled as a noise point. The process repeated until all points are visited.

3. EXPERIMENTS

In this paper, we investigated the behaviour of clustering techniques that explained in Section 1 on hyperspectral data (see Section 2). The followed approach starts by applying two pre-processing steps which are (1) noise reduction by removing zero labelled data, and (2) normalization to reduce the variance between data and exclude the outliers. After that, different clustering techniques are applied on the data. Finally, for each clustering result, five performance measures are computed and will be explained in Section 3.2. The flowchart in Figure 1 summarize the proposed approach on this paper.

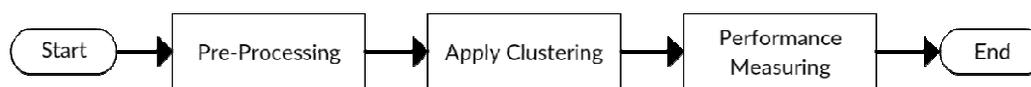


Figure 1. Proposed approach

3.1. Data Description

Four different data sets are used for experiments: Botswana, KSC, Pavia, and PaviaU. Here are brief description and information about each data set:

3.1.1. Botswana

Obtained by NASA-EO1 satellite over Okavango Delta, Botswana in 2001-2004. The data is acquired by EO1 Hyperion sensor at 30m pixel resolution from an altitude of 7.7km in 242 bands of 400-2500nm wavelength and 10nm width [17]. Feature selection is conducted in order to remove the noise of water absorption to result in 145 bands. The data is used for analysis purposes such as classification, where there are 14 classes representing various land cover types in seasonal and occasional swamps and drier woodlands in Delta.

3.1.2. Kennedy Space Centre (KSC)

Obtained by NASA AVIRIS over Kennedy Space Centre, Florida in 1996. The data is acquired at 18m pixel resolution from an altitude of 20km in 224 bands of 400-2500nm wavelength and 10nm width [17]. Feature selection is applied to remove the water absorption and SNR bands and the resulted data have 176 bands. For classification purposes, there are 13 classes representing various land cover types that defined and usually occurred on the environment site.

3.1.3. Pavia Centre and University

Obtained by the ROSIS sensor during a flight campaign over Pavia, Italy in 2002. The data is acquired at 1.3m per pixel spectral resolution with 102 bands and 103 bands for Pavia centre and Pavia university (PaviaU) respectively [18]. Pavia centre data set have 9 ground truth classes which are: water, trees, asphalt, self-blocking bricks, tiles, bitumen, shadows, meadows, and bare soil. Pavia university also have the same 9 classes but with gravel instead of tiles and painted metal sheets instead of self-blocking bricks.

3.2. Experiments Parameters

The following settings have been adopted in our experiments:

- Using cosine distance in K-means, K-medoids, and hierarchical clustering.
- Average function is used as a linkage strategy in hierarchical clustering.
- FCM is applied with an overlapping scaler equals to 1.1 to make it fuzzier.
- Due to huge memory requirements of hierarchical and DBSCAN clustering; data selection is applied. The process is achieved by selecting $p\%$ of each ground-truth class in the data set to be $p\%$ of the whole data set. The chosen percentages are **25%** and **10%** for hierarchical and DBSCAN clustering respectively.

In order to assess the performance of the proposed approach, five performance measurements are computed. The first one is clustering accuracy that calculated by assigning label for each cluster based on its elements true labels majority. Then the weighted accuracy of each cluster is computed as the clustering correct rate multiplied by the number of elements in each cluster. The final clustering accuracy is computed as the average of all clusters weighted accuracies.

To compute the other measurements, four parameters should be defined for each cluster C_i , which are: (1) number of correctly clustered elements (TP), (2) number of correctly rejected elements (TN), (3) number of incorrectly clustered elements (FP), and (4) number of incorrectly rejected data elements (FN).

The other four measurements are: Rand statistics (Q_{Rand}), Jaccard coefficient ($Q_{Jaccard}$), Fowlkes-Mallows index (Q_{FM}), and Hubert index (Q_{Hubert}) [19]. These measurements are defined in Equation (5), (6), (7) and (8) :

$$Q_{Rand} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Q_{Jaccard} = \frac{TP}{TP + FP + FN} \quad (6)$$

$$Q_{FM} = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}} \quad (7)$$

$$Q_{Hubert} = \frac{TP (TP + TN + FP + FN) - ((TP + FP)(TP + FN))}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

3.3. Results and Analysis

For all considered data sets, FCM [9, 10] is giving better clustering results with respect to accuracy measure. However, since the data sets are unbalanced, we investigate the other considered measures.

For Botswana data set[17] (Figure 2), it is confirmed that FCM is outperforming the other considered clustering algorithms with respect to all performance measures. Similarly, for Pavia data set[18] (Figure 4), FCM is either the best clustering results or slightly behind hierarchical clustering algorithm [11, 12]. For KSC data set[17] (Figure 3), K-means [5, 6] and K-medoids [7, 8] are giving the best results. Concerning PaviaU data set[18] (Figure 5), FCM is giving better

results according to the accuracy and Q_{Rand} measures. However, according to Q_{Jaccard} , Q_{FM} , and Q_{Hubert} , the hierarchical clustering is giving better results.

In summary, we noticed that FCM is doing better than the other considered clustering algorithms. However, the accuracy is between 0.5 and 0.6. Moreover, FCM result depends on the value of the overlapping scaler m that needs to be tuned in order to find the optimal result.

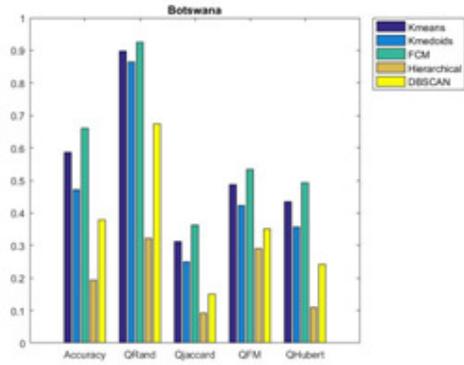


Figure 2. Botswana data set clustering results

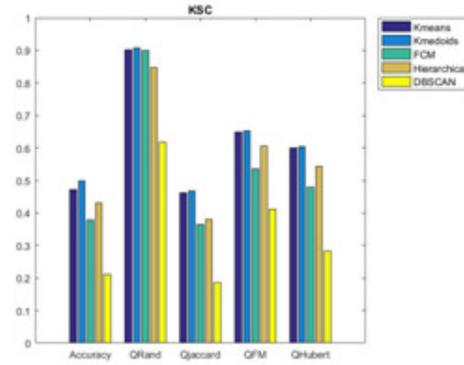


Figure 3. KSC data set clustering results

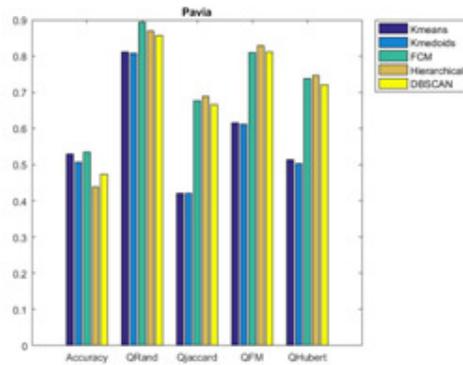


Figure 4. Pavia data set clustering results

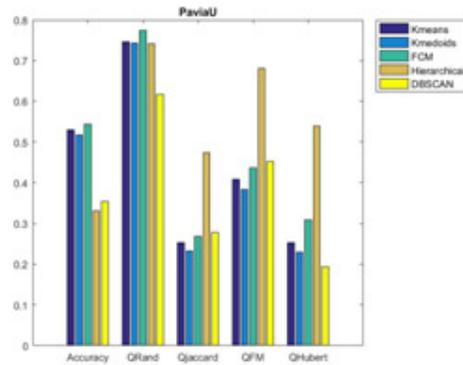


Figure 5. PaviaU data set clustering results

4. CONCLUSION

Recently, hyperspectral imaging (HSI) attracted researchers' interests. Classification and clustering algorithms are conducted to analyse and study their results on HSI. In this paper we investigated the behaviour of known clustering techniques on HSI, which are: k-means[5, 6], K-medoids [7, 8], fuzzy C-Means (FCM)[9, 10], hierarchical[11, 12], and DBSCAN[13]. Four different HSI data sets are chosen in our experiments, which are: Botswana[17], KSC[17], Pavia[18], and PaviaU[18]. We found that FCM is doing better on Botswana and Pavia data sets, K-means and K-medoids are giving better results on KSC data set, and for PaviaU the hierarchical clustering is better.

In order to enhance further the results, we plan to use fusion techniques on the considered clustering algorithms. We will also consider semi-supervised clustering techniques [19, 20] where some of the known information is used to guide the clustering processes to a better partition. In addition, we intend to conduct advanced noise reduction process by applying some filtering techniques[21, 22]. As an example of filtering techniques that we may consider are: maximum likelihood adaptive filter for partially observed Boolean dynamical systems[23], optimal state

estimation for Boolean dynamical systems using a Boolean Kalman smoother[24], and automated lane detection by k-means clustering [25].

REFERENCES

- [1] H. Grahn and P. Geladi, Techniques and applications of hyperspectral image analysis. John Wiley & Sons, 2007.
- [2] G. Vane and A. F. Goetz, "Terrestrial imaging spectroscopy," *Remote Sensing of Environment*, vol. 24, no. 1, pp. 1-29, 1988.
- [3] C.-I. Chang, *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007.
- [4] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003.
- [5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, vol. 1, no. 14, pp. 281-297: Oakland, CA, USA.
- [6] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108, 1979.
- [7] Y. Dodge, *Statistical data analysis based on the L1-norm and related methods*. Birkhäuser, 2012.
- [8] A. Struyf, M. Hubert, and P. Rousseeuw, "Clustering in an object-oriented environment," *Journal of Statistical Software*, vol. 1, no. 4, pp. 1-30, 1997.
- [9] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," 1973.
- [10] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
- [11] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval* (no. 1). Cambridge university press Cambridge, 2008.
- [12] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241-254, 1967.
- [13] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, 1996, vol. 96, no. 34, pp. 226-231.
- [14] S. Li, B. Zhang, A. Li, X. Jia, L. Gao, and M. Peng, "Hyperspectral imagery clustering with neighborhood constraints," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 588-592, 2013.
- [15] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote sensing of environment*, vol. 113, pp. S110-S122, 2009.
- [16] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern recognition*, vol. 40, no. 3, pp. 825-838, 2007.
- [17] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to knowledge transfer for hyperspectral data analysis," in *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on, 2006*, pp. 541-544: IEEE.

- [18] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874-887, 2005.
- [19] O. Bchir, "Unsupervised and Semi-supervised Clustering with Learnable Cluster Dependent Kernels," University of Louisville, 2011.
- [20] E. Bair, "Semi-supervised clustering methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 349-361, 2013.
- [21] R. Chandel and G. Gupta, "Image Filtering Algorithms and Techniques: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 10, 2013.
- [22] S. Tania and R. Rowaida, "A comparative study of various image filtering techniques for removing various noisy pixels in aerial image," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 3, pp. 113-124, 2016.
- [23] M. Imani and U. M. Braga-Neto, "Maximum-Likelihood Adaptive Filter for Partially Observed Boolean Dynamical Systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 359-371, 2017.
- [24] M. Imani and U. Braga-Neto, "Optimal state estimation for boolean dynamical systems using a boolean kalman smoother," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 972-976: IEEE.
- [25] A. Gupta and P. S. Merchant, "Automated Lane Detection by K-means Clustering: A Machine Learning Approach," *Electronic Imaging*, vol. 2016, no. 14, pp. 1-6, 2016.

A PROCESS OF LINK MINING

Dr. Zakea Il-agure and Hicham Noureddine Itani

Higher Colleges of Technology, United Arab Emirates

ABSTRACT

Many data mining and knowledge discovery methodologies and process models have been developed, with varying degrees of success, there are three main methods used to discover patterns in data; KDD, SEMMA and CRISP-DM. They are presented in many of the publications of the area and are used in practice. To our knowledge, there is no clear methodology developed to support link mining. However, there is a well known methodology in knowledge discovery in databases, known as Cross Industry Standard Process for Data Mining (CRISPDM), developed by a consortium of several industrial companies which can be relevant to the study of link mining. In this study CRISP-DM has been adapted to the field of Link mining to detect anomalies. An important goal in link mining is the task of inferring links that are not yet known in a given network. This approach is implemented through the use of a case study of real world data (co-citation data). This case study aims to use mutual information to interpret the semantics of anomalies identified in co-citation, dataset that can provide valuable insights in determining the nature of a given link and potentially identifying important future link relationships.

KEYWORDS

Link mining, anomalies, mutual information

1. INTRODUCTION

In the last decade we have seen an increasing interest in the study of anomalies detection in data mining applied to law enforcement, financial fraud, and terrorism. In recent years, this study has been applied to social networks and online communities to identify influential networks participants and predict fraudulent or malicious activities. To our knowledge, the study of anomaly detection in link mining relied mostly on statistical or machine learning methods in order to gain insight to the structure of their networks. We believe that we can achieve a better understanding of these anomalies if we apply mutual information to the data entities and objects and links to reveal their semantic relationship. The aim of this research is to show how mutual information can help in providing a semantic interpretation of anomalies in data, to characterise the anomalies, and how mutual information can help measuring the information that object item X shares with another object item Y. This paper attempted to demonstrate the contribution of mutual information to interpret anomalies using a case study. This paper presents a novel approach to anomaly detection in link mining methodology based on mutual information.

2. LINK MINING METHODOLOGY

As CRISP-DM methodology is well developed and applied in knowledge discovery; this research has adapted it to the emerging field of link mining. While data mining addresses the discovery of patterns in data entities, link mining is interested in finding patterns in objects by exploiting and modeling the link among the objects. The approach to link mining is still an adhoc approach. The proposed adopted CRISP-DM methodology can help provide a structured approach to link mining in Figure 1. This consists of six stages:

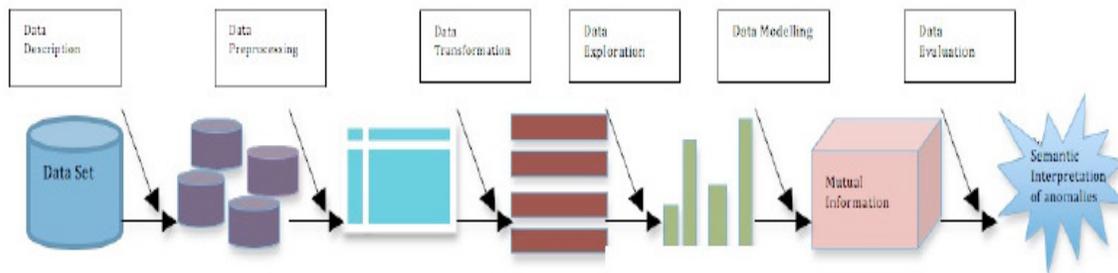


Figure 1. Link mining methodology

The aim of this methodology is to define the link mining task and determining the objectives of link mining.

1. **Data description:** The data description phase starts with initial data collection and proceeds with activities that enable the researcher to become familiar with the data. The aim is to check data quality and any associated problems in order to discover first insights into the data, and identify interesting subsets to form hypotheses regarding hidden information.
2. **Data pre-processing:** The data pre-processing phase covers activities related to data cleansing and data integrity needed to construct the final dataset from the initial raw data. While outliers can be considered noise, or anomalies and thus discarded in data mining, they become the focus of this study as they can reveal important knowledge in link mining.
3. **Data transformation:** This involves syntactic modifications applied to the data; this maybe required by the modelling tool. Selecting an appropriate representation is an important challenge in link mining. The objects in link mining (e.g. people, events, organisation, and countries) have to be transformed into feature factors to represent and capture the connectivity and the strength of the links among those objects.
4. **Data exploration:** This stage is concerned with the distribution of the data and using relevant graphical tools to visualise the structure of the objects and their links. This stage helps identify the existence of anomalous objects or links.

5. **Data modeling:** This stage aims to identify all entities and the relationship between them. Data modeling puts algorithm in general in a historical perspective rooted in mathematics, statistics, and numerical analysis. For more complex data sets, different techniques are used such as nearest neighbour, statistical, classification, and information/context based approaches.
6. **Evaluation:** Data cleaning solutions will clean data by cross checking with a validated data set in phase 2. The clustering model in phase 5, explains natural groupings within a dataset based on a set of input variables. The resulting clustering model is sufficient statistics for calculating the cluster group norms and anomaly indices. Mutual information is useful in validating the model as it provides a semantic.

3. CASE STUDY

The application of the novel approach is implemented to a case study to demonstrate how mutual information can help explore and interpret anomalies detection with a real-world data set and application area. The key challenge for this technique is to apply data representation, for example graphs to visualise the dataset and a clustering approach (hierarchical cluster method). In Figure 2 shows how this study focuses on a case study using a set of co-citation data. The link mining methodology described above is applied to this case study and includes the following stages: data description, data preprocessing, data transformation, data exploration, data modeling based on graph mapping, hierarchical cluster and visualisation, and data evaluation.

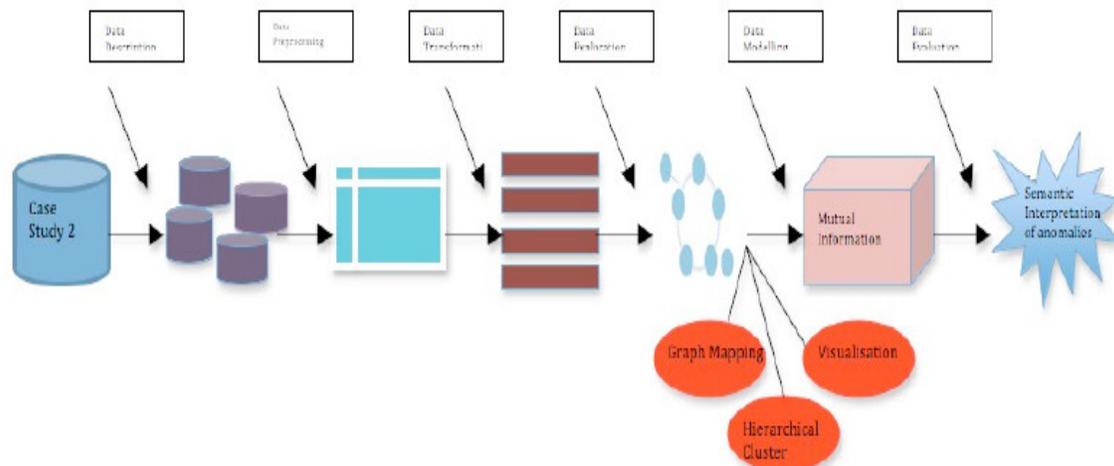


Figure 2. Link mining methodology in case study

This case study covers the three link mining tasks. It is an attempt at identifying and clustering objects, representing them into a graph structure and studying the links between these objects.

4. DISCUSSION

If the approach were to be valid when used with a data set where the anomalies and relationships are unknown, it is necessary to demonstrate that the approach could be scaled to real world data volumes and used with inconsistent and/or noisy data and with other clustering algorithms. This case study addresses these issues. The clustering approach used in this case study was hierarchical clustering. Using bibliographic data, this approach created 5 clusters. Cluster 1 was found to contain data with the strongest links and cluster 5 to contain data with the weakest links. Applying mutual information, we were able to demonstrate that the clusters created by applying the algorithm reflected the semantics of the data. Cluster 5 contained the data with the lowest mutual information calculation value. This demonstrated that mutual information could be used to validate the results of the clustering algorithm.

As the result in Table 1, cluster 1 shows high mutual information indicating higher co-citation strength; cluster 5 has a low mutual information indicating lower co-citation strength.

Table 1. Result of mutual information

	Clusters	Items	Colour	Mutual information
1	Cluster1	58		0.93
2	Cluster2	49		0.82
3	Cluster3	38		0.63
4	Cluster4	29		0.43
5	Cluster5	19		0.00

It was necessary to establish whether the proposed approach would be valid if used with a data set where the anomalies and relationships were unknown. Having clustered and then visualized the data and examined the resulting visualisation graph and the underlying cluster through mutual information, we were able to determine that the results produced were valid, demonstrating that the approach can be used with the real world data set. Analyzing each of the clusters, and the relationships between elements in the clusters was time consuming but enabled us to establish that the approach could be scaled to real world data and that it could be used with anomalies which were previously unknown. We found with the case study that the semantic preprocessing stage was an essential first step. The data from the bibliographic sources normally contains errors, such as misspelling the author's name, the journal title, or in the references list. Occasionally, additional information has to be added to the original data, for example, if the author's address is incomplete or wrong. For this reason, the analysis cannot be applied directly to the data retrieved from the bibliographic sources -a pre-processing stage over the retrieved data is necessary to overcome these issues. In this case study, the clustering approach was used to cluster the data into

groups sharing common characteristics, graph based visualization and mutual information was used to validate the approach.

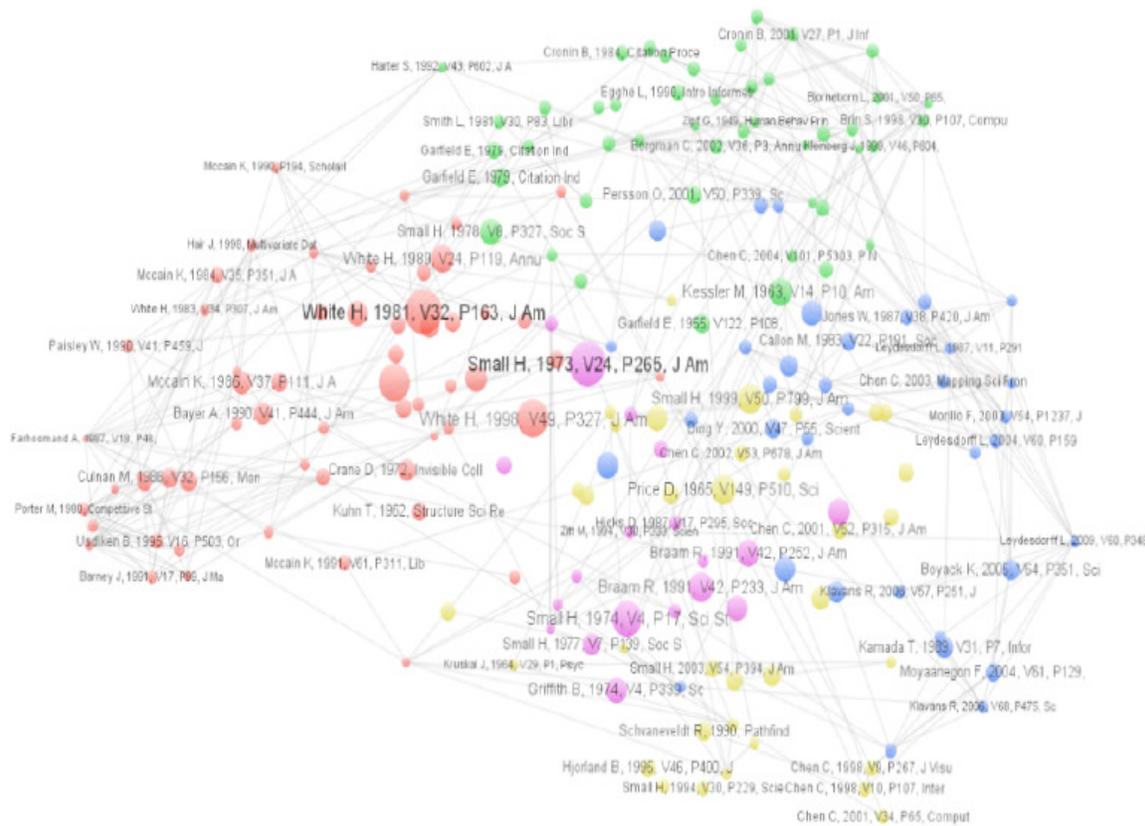


Figure 3. Mapping nodes.

Clusters are designed to classify observations, as anomalies should fall in regions of the data space where there is a small density of normal observations. The anomalies occur in this case study as a cluster among the data, such observations are called collective anomalies, defined by Chandola et al. (2009) as follows: “The individual data instances in a collective anomaly may not be anomalies by themselves, but their occurrence together, as a collection is anomalous.” Existing work on collective anomaly detection requires supporting relationships to connect the observations, such as sequential data, spatial data and graph data. Mutual information can be used to interpret collective anomalies. Mutual information can contribute to our understanding of anomalous features and help to identify links with anomalous behaviour. In this case study, mutual information was applied to interpret the semantics of the clusters. In cluster 5, for example, mutual information found no links amongst this group of nodes. This indicates collective anomalies, as zero mutual information between two random variables means that the variables are independent. Link mining considers data sets as a linked collection of interrelated objects and therefore it focuses on discovering explicit links between objects. Using mutual information allows us to work with objects without these explicit links. Cluster 5 contained documents, which had been selected as part of the co-citation data, but these documents were not themselves cited. Mutual information allowed us to examine the relationships between documents and to determine that some objects made use of self-citation meaning that they were regarded co-

cited but did not connect to other objects. We also identified a community anomaly, where the edge is considered a relationship anomaly, because it connects two communities, which are usually not connected to one another. Mutual information provided information about the relationships between objects, which could not be inferred from a clustering approach alone. This additional information supports a semantic explanation of anomalies.

5. SUMMARY

In this study, hierarchical clustering is applied to identify clusters and the data is visualised using graph representation. Anomalies occur as a cluster among the data, such observations are collective anomalies. Cluster validity with respect to anomalies can be difficult to evaluate because of data volumes. This research has demonstrated that mutual information can be applied to evaluate cluster content and the validity of the clustering approach. This also supports validation of the visualisation element. This case study was developed to use mutual information to validate the visualization graph. We used a real world data set where the anomalies were not known in advance and the data required pre-processing. We were able to show that the approach developed when scaled to large data volumes and combined with semantic pre-processing, allowed us to work with noisy and inconsistent data. The co-citation data applied hierarchical clustering and visualised the data as a graph where nodes represented authors and edges represented cited-by. The aim was to cluster the nodes into groups sharing common characteristics; mutual information was applied to all clusters and demonstrated strong links among the element of each cluster, except in cluster 5. Mutual information conforms that cluster 5 elements share no links with the clusters and among themselves no link was found between authors. Zero mutual information between two random variables means that the variables are independent. Mutual information supported a semantic interpretation of the clusters, as shown by the discussion of cluster 5. The experimental work confirmed the effectiveness and efficiency of the proposed methods in practice. In particular, this revealed that our method is able to deal with data sets with a large number of objects and attributes. Having clustered and then visualised the data and examined the resulting visualisation graph and the underlying cluster through mutual information, we were able to determine that the results produced were valid, demonstrating that the approach can be used with the real world data set. Anomalies detection finds applications in many domains, where it is desirable to determine interesting and unusual events in the activity, which generates such data. The core of all anomalies detection methods is the creation of a probabilistic, statistical or algorithmic model, which characterises the normal behaviour of the data. The deviations from this model are used to determine the anomalies. A good domain-specific knowledge of the underlying data is often crucial in order to design simple and accurate models, which do not over fit the underlying data. Using mutual information contributes to our understanding of the anomalous features and helps with semantic interpretation and to identify links with anomalous behavior. The problem of anomalies detection becomes especially challenging, when significant relationships exist among the different data points. This is the case for bibliographic data in which the patterns in the relationships among the data points play a key role in defining the anomalies. In the data used in this case study, there is significantly more complexity in terms of how anomalies may be defined or modelled which can be used to interpret semantic meaning. Therefore, anomalies may be defined in terms of significant changes in the underlying network community or distance structure. Such models combine network analysis and change detection in order to detect structural and temporal anomalies from the underlying data. This research has demonstrated that mutual information can be applied to evaluate cluster content

and the validity of the clustering approach. This also supports validation of the visualization element.

REFERENCES

- [1] G. Chandola V., Banerjee A., and Kumar V.(2009) Anomaly Detection. A Survey, ACM. Computing Survey. 41(3). p.15.
- [2] Shearer C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22..
- [3] IL-agure, Z. I. (2016). Anomalies in link mining based on mutual information). Staffordshire University. UK.

AUTHOR INDEX

- Arwa Alturki* 73
- Benjamin Aziz* 01
- Greeshma R* 23
- Hicham Nouredine Itani* 81
- Jean Tajer* 01
- Li Nan* 53
- Liu Kai-ming* 53
- Mo Adda* 01
- Mohamed Nacer Bouatit* 13
- Moneef Almutairi* 33
- Mostafa Hefnawi* 43
- Ouiem Bchir* 73
- Sangwon Hyun* 63
- Selma Boumerdassi* 13
- Shahanaz N* 23
- Stephen Riddle* 33
- Tai-Myoung Chung* 63
- Wang Yan-ru* 53
- Wang Yi-rong* 53
- YunSuk Yeo* 63
- Zakea Il-agure* 81
- Zhang Hao* 53