INNOV 2014

The Third International Conference on Communications, Computation, Networks and Technologies

ISBN: 978-1-61208-373-5

October 12 - 16, 2014

Nice, France

INNOV 2014 Editors

Pascal Lorenz, University of Haute-Alsace, France

Eugen Borcoci, University Politehnica of Bucharest, Romania

INNOV 2014

Forward

The Third International Conference on Communications, Computation, Networks and Technologies (INNOV 2014), held between October 12 - 16, 2014 in Nice, France, aimed at addressing recent research results and forecasting challenges on selected topics related to communications, computation, networks and technologies.

Considering the importance of innovative topics in today's technology-driven society, there is a paradigm shift in classical-by-now approaches, such as networking, communications, resource sharing, collaboration and telecommunications. Recent achievements demand rethinking available technologies and considering the emerging ones.

The conference had the following tracks:

- Networking
- Telecommunications

We take here the opportunity to warmly thank all the members of the INNOV 2014 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to INNOV 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the INNOV 2014 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that INNOV 2014 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the areas of communication, computation, networks and technologies. We also hope that Nice, France provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

INNOV 2014 Chairs

Pascal Lorenz, University of Haute-Alsace, France
Eugen Borcoci, University Politehnica of Bucharest, Romania

INNOV 2014

Committee

INNOV 2014 Technical Program Committee

Omar Alhazmi, Taibah University, Saudi Arabia
Alargam Elrayah Elsayed Ali, University of Khartoum, Sudan
Wan D. Bae, University of Wisconsin-Stout, USA
Henri Basson, University of Lille North of France, France
Michael Bauer, The University of Western Ontario, Canada
Khalid Benali, LORIA - Université de Lorraine, France
Eugen Borcoci, University Politehnica of Bucharest, Romania
Albert M. K. Cheng, University of Houston, USA
Grzegorz Chmaj, University of Nevada - Las Vegas, USA
Li-Der Chou, National Central University, Taiwan
Morshed U. Chowdhury, Deakin University-Melbourne Campus, Australia
Matteo Dell'Amico, EURECOM, France
Jacques Demongeot, IMAG/University of Grenoble, France
Uma Maheswari Devi, IBM Research, India
Nima Dokoohaki, Royal Institute of Technology (KTH), Sweden
Tarek El-Ghazawi, George Washington University, USA
Mohamed Y. Eltabakh, Computer Science Department - Worcester Polytechnic Institute, USA
Agata Filipowska, Poznan University of Economics, Poland
David A. Gustafson, Kansas State University, USA
Fred Harris, University of Nevada - Reno, USA
Houcine Hassan, Universitat Politecnica de Valencia, Spain
Pao-Ann Hsiung, National Chung Cheng University, Taiwan
Shih-Chang Huang, National Formosa University, Taiwan
Yo-Ping Huang, National Taipei University of Technology, Taiwan
Sajid Hussain, Fisk University, Nashville, USA
Tazar Hussain, King Saud University (KSU) - Riyadh, Kingdom of Saudi Arabia
Wen-Jyi Hwang, National Taiwan Normal University, Taiwan
Sergio Ilarri, University of Zaragoza, Spain
Abdessamad Imine, LORIA-INRIA, France
Wassim Jaziri, Taibah University, Saudi Arabia
Miao Jin, University of Louisiana - Lafayette, USA
Eugene John, University of Texas at San Antonio San Antonio, USA
Khaled Khankan, Taibah University, Saudi Arabia
Igor Kotenko, St. Petersburg Institute for Informatics and Automation, Russia
Raquel Trillo Lado, University of Zaragoza, Spain
Marcela Castro León, Universitat Autònoma de Barcelona, Spain

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Security Threats in Mobile Ad Hoc Networks

Hande Bakiler, Aysel Şafak
Department of Electrical & Electronics Engineering
Baskent University
Ankara, Turkey
21020013@baskent.edu.tr, asafak@baskent.edu.tr

Ilgın Şafak
Progress R&D Center
Provus Information Technologies
Sisli, Istanbul, Turkey
ilgin.safak@provus.com.tr

***Abstract*—Mobile Ad Hoc Networks (MANET) are continuously self-organizing wireless networks with no fixed infrastructure, where network communication is established without a centralized administration. Security is an important issue for mobile ad hoc networks, due to the vulnerable nature of MANETs. This paper describes the effects of Pulse Jammer attack, Misbehavior Node attack and Byzantine attacks on the network performance under different traffic loads using Geographic Routing Protocol (GRP), Proactive Routing Protocol such as Optimized Link State Routing (OLSR) Protocol and Reactive Routing Protocols such as Ad Hoc On Demand Distance Vector (AODV) Routing Protocol and Dynamic Source Routing (DSR) Protocol. The impact of security attacks on MANET performance is evaluated by investigating which attack is more harmful to the network. IEEE 802.11b and 802.11g standards are compared with respect to the Pulse Jammer attack, Misbehavior Node attack and Byzantine attack for AODV Routing Protocol. Simulation results using OPNET simulator show that the efficient utilization of the network reduces considerably in the presence of the mentioned attacks.**

Keywords- mobile ad hoc networks (MANETs); routing attacks; network security; OPNET

I. INTRODUCTION

Next generation wireless communication systems will require a rapid deployment of independent mobile users. An emerging wireless technology, mobile ad hoc networks (MANETs), are efficient, effective, quick, and easy to deploy in networks with changing topologies. Each mobile node acts as a host, and also acts as a router. Nodes communicate with each other without the intervention of access points or base stations [1]. Ad-hoc networks are suitable for applications where it is not possible to set up a fixed infrastructure and have a dynamic topology so that nodes can easily join or leave the network at any time. Possible MANET scenarios include communications in military and rescue missions in connecting soldiers on the battlefield or establishing new networks where a network has collapsed after a disaster like an earthquake [2]. Nodes cooperate by forwarding data packets to other nodes in the network to find a path to the destination node using routing protocols. However, due to security vulnerabilities of the routing protocols, wireless ad-hoc networks are unprotected to attacks of the malicious nodes. These nodes destroy the network, thereby degrading the network performance.

The effects of Pulse Jammer attack and Misbehavior nodes using Optimized Link State Routing Protocol (OLSR), Reactive routing protocol, Ad Hoc On Demand Distance Vector (AODV) and Geographical are studied in [3], where the impact of attack on MANET performance is evaluated in finding out which protocol is more vulnerable to these attacks. No single protocol that was studied had an overall better performance under Pulse Jammer attack and Misbehavior nodes security threats.

Various protocol aware jamming attacks that can be launched in an access point based 802.11b network are studied in [4]. It is shown that misbehaving nodes that do not adhere to the underlying MAC protocol significantly degrade the network throughput. Several hybrid attacks that increase the effectiveness of the attack or the decrease the probability of detection of the attack are also presented in the paper.

In this paper, the effects of Pulse Jammer Attack, Misbehavior Node attack and Byzantine security attacks on MANET network topology are studied using different routing protocols. The purpose of this work is access security attacks on MANETs that lead to a reduced network performance, reliability and availability. Additionally, several security routing protocols are investigated for MANET. For each scenario, normal network traffic is compared to the network traffic with five disruptive nodes that are placed in the network separately.

The main contribution of this work is providing insight about network security challenges and potential harmful attacks in MANET security under different traffic loads using various routing protocols. In this work, wlan_wkstn (Wireless LAN Workstation) mobile nodes are used, so the network traffic loads, i.e., http, ftp, email, voice and video conferencing can be enabled on these mobile nodes in the network. Performance metrics are provided for different network applications in addition to the whole network performance using different routing protocols. The IEEE 802.11b and 802.11g standards are compared for the normal network with and without network attackers.

 ISBN: 978-1-61208-373-5

The paper is organized as follows: in Section II, an overview of the OLSR, GRP, DSR and AODV routing protocols are provided. In Section III, Pulse Jammer attack is described. In Section IV, Misbehavior Node attack is described and in Section V, Byzantine attack is described. Performance metrics which are used in the simulations are presented and described in Section VI. Simulation results are given in Section VII, followed by the conclusion in Section VIII.

II. Overview of Current Routing Protocols

In this section, various existing routing protocols are described.

A. The Dynamic Source Routing (DSR) Protocol

DSR [5] is a reactive unicast routing protocol that utilizes source routing algorithm. The sender knows the complete hop-by-hop route to the destination, where the routes are stored in a route cache. When a node in the ad hoc network attempts to send a data packet to a destination for which it does not know the route, it uses a route discovery process to dynamically determine one. Route discovery works by flooding the network with route request (RREQ) packets. A route reply is generated when the route request reaches either the destination itself, or an intermediate node which contains in its route cache an unexpired route to the destination. By the time the packet reaches the destination or an intermediate node, it contains a route record yielding the sequence of hops taken.

B. The Ad Hoc On-demand Distance Vector (AODV) Routing Protocol

AODV routing protocol [1] is a reactive unicast routing protocol for mobile ad hoc networks which only needs to maintain the routing information about the active paths. In AODV, routing information is maintained in routing tables at nodes. Every mobile node keeps a next-hop routing table, which contains the destinations to which it currently has a route to. A routing table entry expires if it has not been used or reactivated for a pre-specified expiration time.

C. Optimized Link State Routing (OLSR) Protocol

OLSR protocol, as defined in [6], is a proactive routing protocol based on the periodic exchange of topology information. Generally, three types of control messages are used in the OLSR protocol, namely, a HELLO message, a TC (Topology Control) message and a MID (Multiple Interface Declaration) message. The HELLO message is transmitted for sensing neighbors and for Multi-Point Distribution Relays (MPRs) calculation. Topology control is link state signaling that is performed by OLSR. MPRs are used to optimize the messaging process. MID messages contains the list of all IP addresses used by any node in the network. OLSR exchanges the topology information always with other nodes. Nodes maintain information of neighbors and MPRs by sending and receiving HELLO messages from its neighbors.

D. Geographic Routing Protocol (GRP)

GRP [7][8] is a well researched approach for ad hoc routing where nodes are aware of their own geographic locations and also of its immediate neighbors and source node are aware of the destination's position. The data packets are routed through the network using the geographic location of the destination and not the network address. GRP operates without routing tables and routing to destination depends upon the information each node has about its neighbors. Geographic routing is simple and efficient.

III. Pulse Jammer Attack

The most trivial way of disrupting a wireless network is by generating a continuous high power noise across the entire bandwidth near the transmitting and/or receiving nodes. The device that generates such a noise is called a jammer and the process is called jamming [4]. The reason to call jammer as intelligent is because its pulse off time and pulse on time are the main parameters which act on jammer to behave on and off at certain time as define to generate the transmission [3].

IV. Misbehavior Nodes Attack

The purpose of misbehaving nodes [9] is not to function properly in the network and they achieve their goal by acting maliciously. They stop forwarding packets to the other nodes by simply start dropping the packets, or consume the bandwidth of the network by broadcasting route when it is not necessary. The misbehavior nodes stop performing the basic task; as a result, the network becomes congested and the traffic on the network leads to delay of data and degrade the performances of the network.

V. Byzantine Attack

In Byzantine attacks, a compromised intermediate node or a set of compromised intermediate nodes collectively carries out attacks such as creating routing loops, routing packets on non-optimal paths and selectively dropping packets [10]. Byzantine attack drops, modifies and misroutes the forwarding packets in an attempt to disrupt the routing service [11].

VI. Performance Metrics

The performance of the whole network under different routing protocols is analyzed by four metrics: throughput, network load, delay and data dropped.

A. Throughput (bits/sec)

The average rate at which the data packet is delivered successfully from one node to another over a communication network is known as throughput.

 ISBN: 978-1-61208-373-5

B. *Network Load (bits/sec)*

Network load is the total packet sent and received across the whole network at a particular time.

C. *Delay (sec)*

The packet end to end delay is the average time of the packet passing through inside the network.

D. *Data Dropped (bits/sec)*

Data dropped shows that how many packets are successfully sent and received across the whole network.

VII. SIMULATION RESULT AND ANALYSIS

The simulation is performed in analyzing the effects of Pulse Jammer attack, Misbehavior Node attack and Byzantine attack on the network performance under different traffic loads. Simulation parameters used are depicted in Table 1.

TABLE I. SIMULATION PARAMETER

Simulation Parameter	Value
Simulator	OPNET 14.5
Area	800x800 (m)
Number of Nodes	30 Nodes
Operation Mode	802.11b, 802.11g
Data Rate of Each Node	11 Mbps, 54 Mbps
Routing Protocols	DSR, AODV, OLSR, GRP
Mobility Model	Random Waypoint
Traffic Type	HTTP, FTP, Email, Voice, Video Conferencing
Simulation Time	300 sec.
Packet Reception Power Threshold	-95 dBm

A. *Performance of DSR under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for Voice Application*

In the simulation enviroment, five jamming nodes, five misbehaving nodes and five Byzantine nodes were placed separately in the normal network with different scenarios. Then, packet end-to-end delay statistics are represented for voice application in the same graph.

Figure 1 represents the packet end-to-end delay statistics for voice application on the normal network traffic with the average value of 7.667 seconds. It shows the “packet end-to-end delay” with jamming nodes in the network as 10.864 seconds, with misbehaving nodes as 9.748 seconds and with Byzantine nodes in the network as 9.235 seconds with respect to the DSR.

The delay increases in presence of the network attacks on the network when it is compared to the normal network.

Jamming nodes deny the network transmission services to authorized users by generating noise on the wireless medium in order to block the access for authorized nodes. Misbehaving nodes consume a lot of bandwidth and do not collaborate with the other nodes in the network. Byzantine nodes drop the packets in the network which degrades the network routing services.

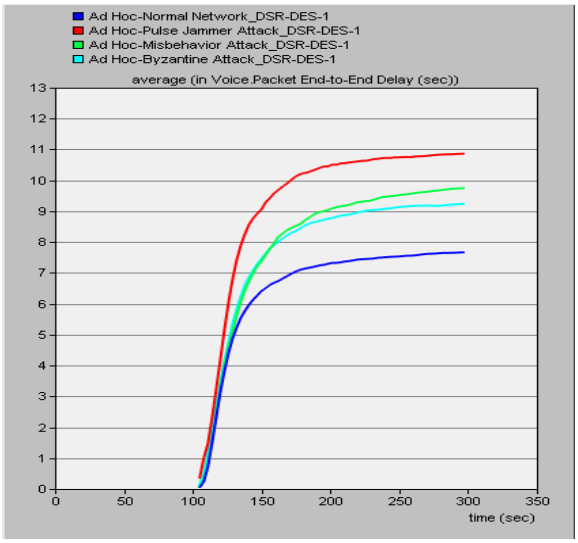


Figure 1. Packet end-to-end delay results of the normal network’s voice application with and without network attacks for DSR

B. *Performance of AODV under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for Voice Application*

In this section, the performance of AODV routing protocol under jamming nodes, misbehaving nodes and Byzantine nodes are compared. First, normal traffic is generated under AODV, and then the scenario was duplicated with a jitter parameter for different attacks. For each network attack scenario, five malicious nodes are placed in the normal network. Jitter [12] is the ratio of transmission delay of the current packet and the transmission delay of the previous packet.

In Figure 2, jitter statistics are represented for voice application in the same graph.

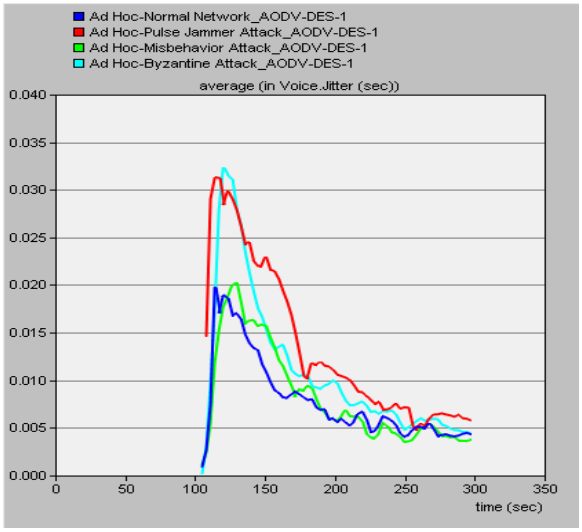


Figure 2. Jitter results of the normal network’s voice application with and without network attacks for AODV routing protocol

In the graph above, it is clearly seen that jitter increases in the beginning of the simulation up to a certain point and from that point onwards it degrades rapidly. This is due to the fact that the utilization of the network reaches a steady state after some time.

Figure 2 shows that the average value of the normal network traffic jitter in voice applications is 0.0043 seconds. On the other hand, the network with jammer nodes shows the jitter with the average value of 0.0057 seconds; with Byzantine nodes the value it is noted as 0.0044 seconds and with misbehaving nodes it is recorded as 0.004 seconds with respect to the AODV routing protocol.

The results show significant changes in jitter for voice application, especially for the network with jamming nodes and with Byzantine nodes. Due to malicious activities of the jamming nodes and Byzantine nodes, the jitter increment is more than the normal network for AODV routing protocol. Also for the network with misbehaving nodes, the jitter increment is more than the normal network in general. However, it reduces at some certain points. The reason of this reduction could be that misbehaving nodes start dropping the packets and do not forward the packets to the other nodes on the network, then the misbehaving nodes start sending the packets and forwarding packets faster than the normal nodes. As a result, normal nodes are not able to process the packets.

C. *Performance of OLSR under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for Email Application*

In this section, the performance of OLSR protocol under jamming nodes, misbehaving nodes and Byzantine nodes are compared. For each network attack scenario, five malicious nodes are placed in the normal network.

In Figure 3, the traffic received statistics for email application on the normal network traffic with and without malicious nodes are analyzed. The normal network's traffic received statistics is recorded as 153.9 bytes/sec. Then, it is noted as 140.5 bytes/sec with jammer nodes in the network. The traffic received statistics average value is 127.1 bytes/sec with misbehaving nodes and with Byzantine nodes in the network its value is noted as 100.32 bytes/sec with respect to the OLSR.

When placing the malicious nodes in the network, the MANET traffic received is recorded lower than the normal network traffic. There is significant traffic destruction of the packets transmission on the network when applying network attacks.

D. *Performance of GRP under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for Video Conferencing Application*

To implement the network attacks on MANET nodes network, five jamming nodes, five misbehaving nodes and five Byzantine nodes are deployed separately in the network for GRP with different scenarios.

The packet end-to-end delay statistics for voice application of the normal network is noted as 0.269 seconds at the duration time of simulation 300 seconds in Figure 4. After implementing the five jamming nodes, it increases to 0.928 seconds. The reason for this is because pulse jammer nodes generate a noise on radio frequency in pulse time which increases the packet end-to-end delay statistics on the network for GRP. The graph represents the packet end-to-end delay statistics of voice application as 0.40 seconds for the network with misbehaving nodes. Due to the misbehaving nodes, the network becomes congested.

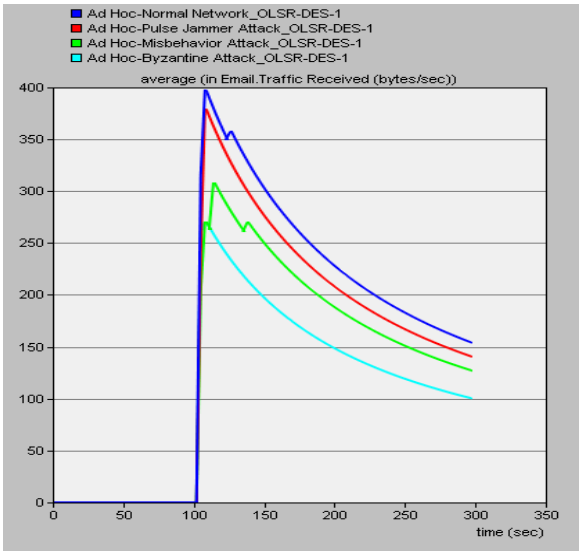


Figure 3. Traffic received results of the normal network's email application with and without network attacks for OLSR protocol

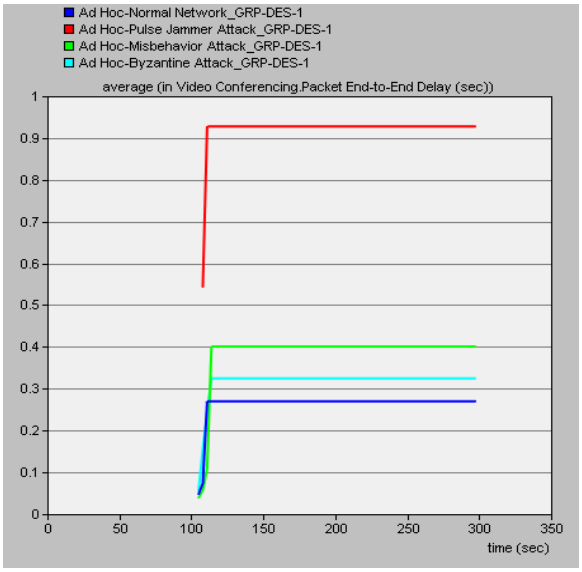


Figure 4. Packet end-to-end delay results of the normal network's video conferencing with and without network attacks for GRP

 ISBN: 978-1-61208-373-5

Figure 4 shows the packet end-to-end delay with Byzantine nodes in the network as 0.325 seconds with respect to the GRP. The Byzantine attack has a negative impact on the transmission and network traffic.

E. Performance of DSR under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for the Network with respect to "Throughput" Statistics

In this section, five jamming nodes, five misbehaving nodes and five Byzantine nodes were placed separately in the normal network with different scenarios. The throughput statistics are represented for the whole network in the same graph in Figure 5.

The throughput of the network nodes with normal traffic is noted as 741,085 bits/sec, whereas the throughput with jamming nodes is noted as 544,661 bits/sec, both for a simulation of 300 seconds duration. As seen in Figure 5, the throughput of the network with Byzantine nodes is recorded as 699,863 bits/sec and with misbehaving nodes as 715,089 bits/sec. The largest reduction of the network throughput statistic is represented for the network with jamming nodes and the least reduction is indicated for the network with misbehaving nodes with respect to the DSR protocol.

F. Performance of AODV under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for the Network with respect to "Network Load" Statistics

In this section, different network attack scenarios were designed separately to examine the AODV routing protocol under five Byzantine nodes, five misbehaving nodes and five jamming nodes.

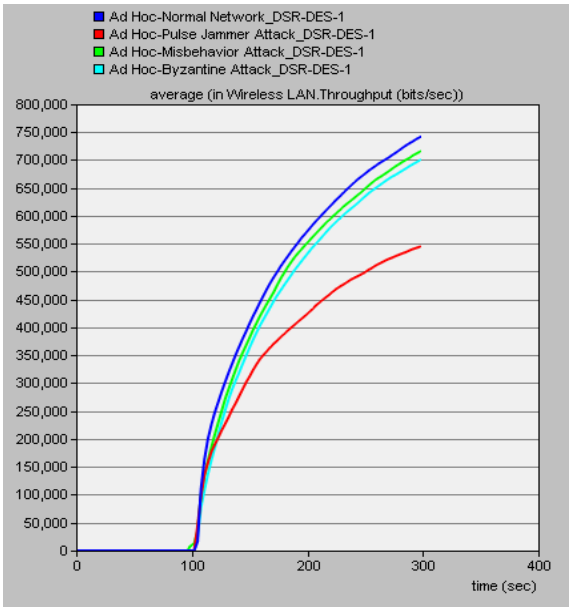


Figure 5. Throughput results of the normal network with and without network attacks for DSR protocol

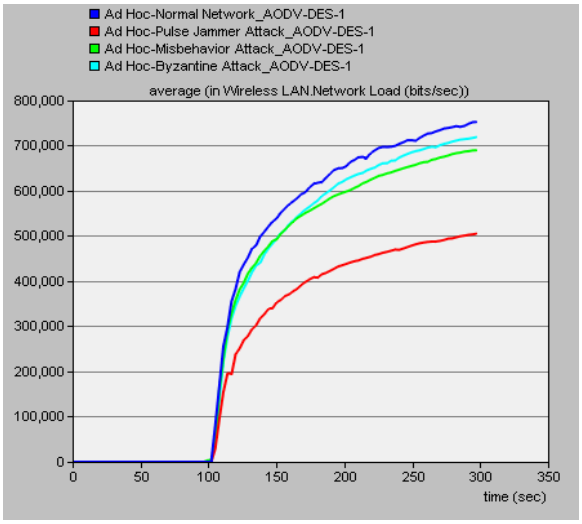


Figure 6. Network load results of the normal network with and without network attacks for AODV routing protocol

The network scenarios for different attacks are depicted in Figure 6. The network load of the normal network has the average value of 752,620 bits/sec and with the jamming nodes in the network it is noted as 505,130 bits/sec. For the network with misbehaving nodes, its average value is 690,004 bits/sec and the network load statistics according to the network with Byzantine nodes is recorded as 718,929 bits/sec. The largest reduction of the network load statistic is represented for the network with jamming nodes and the least reduction is represented for the network with Byzantine nodes with respect to AODV routing protocol.

According to Figure 6, AODV routing protocol is more vulnerable to jamming nodes. Jamming nodes deny service by generating noise and causes protocol packets lost. Jamming nodes block the access for authorized users.

As a result, the network traffic effected negatively when malicious nodes are placed in the normal network and they start dropping the forwarding packets to the other the nodes on the network.

G. Performance of GRP under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for the Network with respect to "Delay" Statistics

Five jamming nodes, five misbehaving nodes and five Byzantine nodes were placed separately in the normal network with different scenarios.

Different network scenarios for the mentioned network attacks are represented in Figure 7 according to GRP protocol.

Figure 7 represents that the normal network traffic delay average value is 3.27 seconds. On the other hand, the network with jammer nodes shows the delay with the average value of 4.42 seconds, with Byzantine nodes the value it is recorded as 3.92 seconds and with misbehaving nodes it is noted as 3.51 seconds with respect to the GRP.

 ISBN: 978-1-61208-373-5

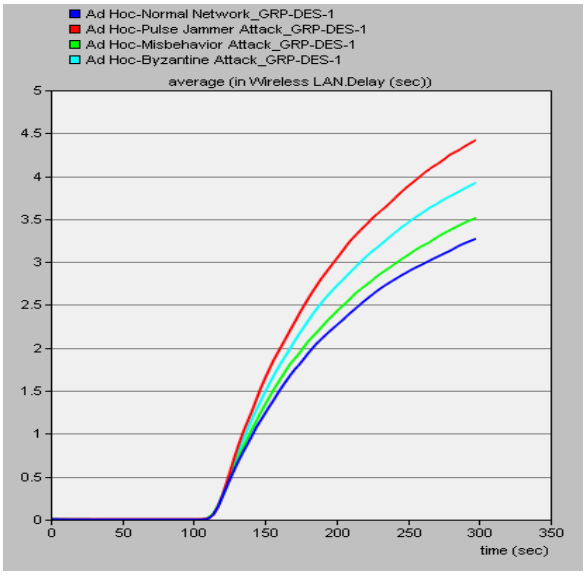


Figure 7. Delay results of the normal network with and without network attacks for GRP

The largest increment of the delay statistic is depicted for the network with jamming nodes and the least increment is represented for the network with misbehaving nodes with respect to GRP. The jamming node attack on GRP shows a significant result. The jamming nodes stop performing the basic task of the network; as a result, the network becomes congested and the traffic on the network leads to delay of the data and degrading of the performances of the network.

H. Performance of OLSR under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for the Network with respect to "Data Dropped" Statistics

In this section, five jamming nodes, five misbehaving nodes and five Byzantine nodes are placed in the network separately for OLSR protocol with different scenarios in implementing the network attacks on MANET nodes network. The data dropped statistics are shown for the whole network in the same graph.

Figure 8 shows the normal network data dropped statistics average value as 22,577 bits/sec. For the network with jamming nodes, the average data dropped value is recorded as 23.074 bits/sec; with misbehaving nodes the data dropped statistics is 24,437 bits/sec and with Byzantine nodes its value is 28,353 bits/sec.

It is seen that the largest increment of the data dropped statistic is represented for the network with misbehaving nodes and the least increment is represented for the network with jamming nodes with respect to the OLSR protocol. That means that the OLSR protocol is more vulnerable to the network with misbehaving nodes.

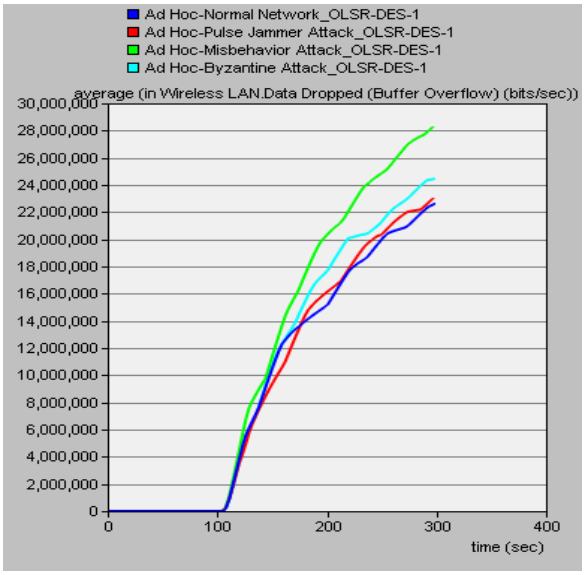


Figure 8. Data dropped results of the normal network with and without network attacks for OLSR

I. Performance of AODV under Pulse Jammer Attack, under Misbehavior Node Attack and under Byzantine Attack for IEEE 802.11g Standard with respect to "Network Load" Statistics

In this section, different network attack scenarios were designed for the AODV routing separately under Byzantine nodes, misbehaving nodes and jamming nodes in order to examine the IEEE 802.11g standard. For each network attack scenario, five malicious nodes are placed in the normal network.

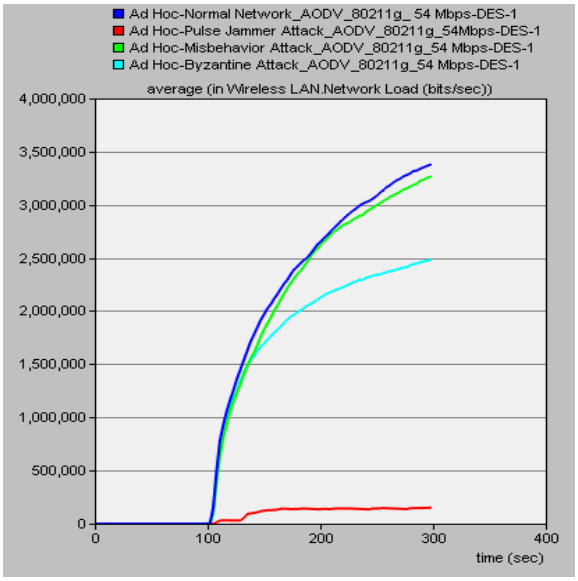


Figure 9. Network load results of the normal network with and without network attacks for AODV with respect to the IEEE 802.11g standard

As seen in Figure 9, the network load performance of the network nodes with normal traffic is 3,376,409 bits/sec and with misbehaving nodes in the network it is represented as 3,262,975 bits/sec. The network load of the network with Byzantine nodes is noted as 2,480,452 bits/sec and with jamming nodes it is recorded as 150,486 bits/sec.

The largest reduction of the network load statistic is represented for the network with jamming nodes and the least reduction is represented for the network with misbehaving nodes with respect to the IEEE 802.11g standard for AODV routing protocol. Hence, networks using 802.11b standard are more vulnerable to jamming nodes in the network.

Compared to the networks using IEEE 802.11b and 802.11g standards, networks using IEEE 802.11b standard are more vulnerable to networks with jamming nodes. On the other hand, networks using IEEE 802.11g standard are the least affected from the network with jamming nodes for AODV routing protocol.

VIII. CONCLUSION AND FUTURE WORK

In this work, the routing protocols GRP, Proactive Routing Protocol (OLSR), and Reactive Routing Protocols (AODV and DSR) are studied in IEEE 802.11b networks. The network performance under Pulse Jammer attack, Misbehavior Node attack and Byzantine attack is investigated. The network contains http (heavy browsing), ftp (high load), email (high load), voice (PCM Quality Spech) and video conferencing (low resolution video) applications. The normal network is compared with the networks which contain jamming nodes, misbehaving nodes and Byzantine nodes in terms of performance metrics, i.e., delay, network load, throughput, data dropped, jitter and traffic received by using different routing protocols. Then, the IEEE 802.11b and 802.11g standards, which share the same propagation characteristics, are compared for networks with and without security attacks using the AODV routing protocol. Results show that routing protocols are more vulnerable to networks with jamming nodes, and placing the intruder nodes in the network reduces the reliability, availability and the performance of the network. Networks using the IEEE 802.11b standard are more vulnerable in networks with jamming nodes for the AODV routing protocol. Jammer attack generates noise on the wireless radio frequency medium to stop the communication in order to trigger the network. Jamming nodes cause corruption of the packets or they cause packet lost. Misbehavior Node attack stops forwarding packets to the other nodes and drop the packets, it stop performing the basic task and the network performance degrades. Also, Byzantine attack drops the routing forwarding table or drops the forwarding packets to the other nodes. Several security breaches are represented under these three attack models using OPNET. They provide useful insight in understanding MANET in terms of the network security.

Future work encompasses extending results to other security attacks and wireless protocols, and adding detection and defense mechanisms that can protect the network from the intruders.

ACKNOWLEDGMENT

This work was supported by ITEA2 ADAX Project No. 10030 and TUBITAK TEYDEB Project No. 9130016.

REFERENCES

[1] C. Liu and J. Kaiser, “A Survey of Mobile Ad Hoc Network Routing Protocols,” The University of Magdeburg, October 2005.

[2] S. Vrutik, D. N. Modi, and P. Ashwin, "AODVGAP-An Acknowledgement Based Approach to Mitigate Selective Forwarding Attacks in MANET," International Journal of Computer Engineering and Technology (IJCET), vol. 3, no. 2, July-September 2012, pp. 458-469.

[3] S. Salim, "Mobile Ad hoc Network Security Issues," M.Sc. Thesis, University of Central Lancashire, 2010, pp. 1-81.

[4] D. J. Thuente and M. Acharya, “Intelligent Jamming in Wireless Networks with Applications to 802.11b and Other Networks,” North Carolina State University.

[5] C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina, "Performance Comparison of Two On-Demand Routing Protocols for Ad Hoc Networks," IEEE Personal Communications, February 2001, pp. 16-28.

[6] S. Ehrampoosh and A. K. Mahani, "Secure Routing Protocols: Affections on MANETs Performance," First International Conference on Communications Engineering, 22-24 December 2010, pp. 77-82.

[7] A. Tamizhselvi and Dr. R. S. D. W. Banu, "Performance Evaluation of Geographical Routing Protocol under Different Traffic Scenario," International Journal of Computer Science and Telecommunications, vol. 3, no. 3, March 2012, pp. 64-67.

[8] J. A. Sanchez, P. M. Ruiz, and R. Marin-Perez, "Beacon-Less Geographic Routing Made Practical: Challenges, Design Guidelines, and Protocols," IEEE Communications Magazine, August 2009, pp. 85-91.

[9] R. K. Jha, I. Z. Bholebawa, U. D. Dalal, and A. V. Wankhede, "Detection and Fortification Analysis of WiMAX Network: With Misbehavior Node Attack," International Journal on Communications, Network and System Sciences, vol. 5, April 2012, pp. 353-367.

[10] N. K. Pani, “A Secure Zone-Based Routing Protocol for Mobile Ad Hoc Networks,” Department of Computer Science and Engineering, National Institute of Technology, May 2009.

[11] B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens, "An On-demand Secure Routing Protocol Resilient to Byzantine Failures," Proceedings of the ACM Workshop on Wireless Security (WiSe), September 2002, pp. 21-30.

[12] H. Paul and P. Sarkar, "A Study and Comparison of OLSR, AODV and ZRP Routing Protocols in Ad Hoc Networks," International Journal of Research in Engineering and Technology (IJRET), vol. 2, no.8, August 2013, pp. 370-374.

 ISBN: 978-1-61208-373-5

A Network-disaster Recovery System using Area-based Network Management

Toshiaki Suzuki, Hideki Endo, Isao Shimokawa,
and Kenichi Sakamoto
Central Research Laboratory
Hitachi, Ltd.
Yokohama, Kanagawa, Japan
{toshiaki.suzuki.cs, hideki.endo.es, isao.shimokawa.sd,
and kenichi.sakamoto.xj}@hitachi.com

Hidenori Inouchi, Taro Ogawa, Takanori Kato,
and Akihiko Takase
Telecommunications & Network Systems Division
Hitachi, Ltd.
Kawasaki, Kanagawa, Japan
{hidenori.inouchi.dw, taro.ogawa.tg, takanori.kato.bq,
and akihiko.takase.wa}@hitachi.com

***Abstract*—A "network-disaster recovery system" using area-based network management is proposed. In this system, a whole network is separated into multiple areas. A network-management server calculates recovery paths for every possible network-area failure and distributes them with a recovery identifier (ID) for each area-failure pattern before starting network operations. Network nodes receive and store the recovery IDs and recovery configurations. The network-management server determines after detecting the network-area failures and distributes the recovery ID to related network nodes. The network nodes that received the recovery ID start data transmission according to the path configurations specified by the recovery ID. After these procedures are completed, the network-area failures are swiftly recovered. A prototype system composed of a network-management server and 96 simulated packet-transport nodes was configured and evaluated. According to the evaluation results, the network-management server could transmit the recovery ID to the related network nodes within 100 milliseconds after it detected network-area failures. That is to say, the network could immediately start to recover from the network-area failures. On the other hand, the calculation time for 500 Pseudo Wires (PWs) is about 344 milliseconds, which is longer than the time taken to distribute the recovery ID (i.e., 100 milliseconds). In other words, if there are over 500 PWs, the proposed system can recover more swiftly than a conventional system (which re-calculates recovery PWs after detecting the network-area failures) under the same evaluation conditions used for the proposed system.**

Keywords-network management; disaster recovery; packet transport; reliable network

I. INTRODUCTION

Lately, as reflected in the rising number of Internet users [1] and the popularity of cloud services [2], applications and services provided by way of networks have become indispensable in daily life. Network services must, therefore, be highly reliable and "always available" [3]. When extensive disasters occur, network services could be out of service for a long time. Consequently, networks must be robust enough so that they can continue to provide network services even if network facilities are extensively damaged.

As recovery procedures for network failures, two major techniques are applied. One is "protection," by which recovery paths are physically prepared in advance of network failures by allocating extra network resources. The other approach is "restoration," by which recovery paths are "calculated" after network failures are detected.

Protection is easily applied to recovery procedures for multi-layer networks, and recovery is immediate because recovery paths are prepared in advance (that is, before network operations are started). However, if the prepared recovery paths are not available when network failures occur, network-connection services will become out of service. On the other hand, if restoration is applied, network connections can be recovered if recovery paths are recalculated after network failures are detected. However, a little more time is needed to recalculate the recovery path if the operated networks are huge and have many network nodes. Therefore, if huge quantities of paths are used to transmit data packets, much time is needed to recalculate all recovery paths, and the network will not recover from a disaster expeditiously. In addition, even if network connections are recovered, all network flows will try to use the same recovery path. As a result, the network will easily become congested, making it difficult to guarantee network-transmission quality.

In light of the above-described issues, a robust network-management scheme is required. Specifically, it controls multi-layer network resources so as to provide and maintain network-connection services at times of a network disaster. To achieve that control, a network-management system has to monitor and control the multi-layer network resources.

The overall aim of the present study is to develop a network-management scheme to provide robust networks that can swiftly recover from a network disaster by monitoring and controlling multi-layer network resources. To recovery swiftly from a network disaster, three steps should be taken. The first step is to find network failures in a short time. The second is to promptly determine how to recover the network. The third is to immediately configure recovery paths. In the present study, the second step is focused on, and a "network-disaster recovery system" using an area-based network-management scheme that controls networks composed of IP networks and packet-transport networks, such as the Multi Protocol Label Switching - Transport Profile (MPLS-TP) network, is proposed.

The rest of this paper is organized as follows. Section II explains the requirements concerning a network-disaster recovery system. Section III proposes the network-disaster recovery system. Section IV describes a prototype system and presents some results of evaluations of the system

 ISBN: 978-1-61208-373-5

performance. Related works are described in Section V, and Section VI concludes the paper.

II. REQUIREMENTS CONCERNING NETWORK-DISASTER RECOVER SYSTEM

The target network structure is shown conceptually in Figure 1. The target network is composed of an IP network layer and a Packet-Transport-Node (PTN) network layer such as an MPLS-TP network. The core network is composed of PTNs. On the other hand, the access network is composed of IP network nodes. In this study, recovery from multiple network failures on IP and PTN networks (for example, the two network failures shown in the figure), is focused on as follows.

One of the critical issues concerning network recovery is the time taken to recover numerous established paths of a packet network in the case of a network disaster. Here, each path is configured by a Label-Switched Path (LSP) and a Pseudo Wire (PW). Specifically, the issue is the time taken to recalculate numerous recovery paths one by one after disconnected paths are detected by monitoring network conditions.

In the case of a packet-transport network, the bandwidth of a network path is guaranteed. Guaranteeing the quality of a recovery path, such as bandwidth and/or end-to-end delays before (as well as after) a network failure, is therefore also an issue.

To tackle the above-mentioned issues, the proposed system should be managed in accordance with the following four requirements.

① Manage multi-layer networks
② Recover from multiple network failures
③ Rapidly establish recovery paths
④ Guarantee quality of recovery paths after network failures are recovered

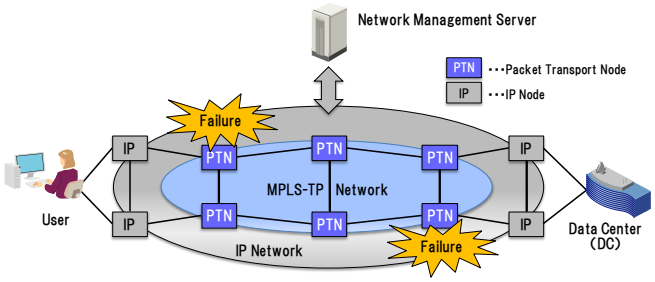


Figure 1. Target network structure

To meet these requirements, the network-disaster recovery system is designed on the basis of the following policy. If there are plenty of paths, recovery paths should not be recalculated after multiple network failures are detected (since it takes much time to recalculate them). On the other hand, recovery paths that guarantee bandwidths and delays for each possible network failure should be calculated preliminarily, and paths should be promptly recovered by using the prepared paths after the network failures are detected.

III. PROPOSED NETWORK DISASTER RECOVERY SYSTEM

In the proposed network-disaster recovery system, a network-management server centrally manages an entire network. In the target network, a core-network segment is composed of PTNs, and an access-network segment is composed of IP network nodes. In addition, the network-management server manages the entire network by dividing it into multiple network areas and controlling each of them by using an area-based network-management scheme.

A. Structure of proposed system

The structure of the proposed network-disaster recovery system is shown in Figure 2. As an example of an area-based management, the network-management server divides the whole PTN network into eight areas and manages them by using the area-based management scheme. The eight areas are shown as network areas (1) to (8) in the figure. In addition, the network-management server is connected to all PTNs, a user terminal, and servers in a datacenter (DC) by another management network (not shown in the figure). The network-management server monitors all PTNs and executes swift network-disaster recovery after detecting catastrophic network failures.

As for the proposed disaster-recovery system, the user terminal is connected to a server in a DC by way of IP networks and PTN networks, and it can get various application services from the server. To provide the user terminal with robust network access, the user terminal is connected to two "PTN network areas," at least. In addition, the DC is connected to two other "PTN network areas", at least.

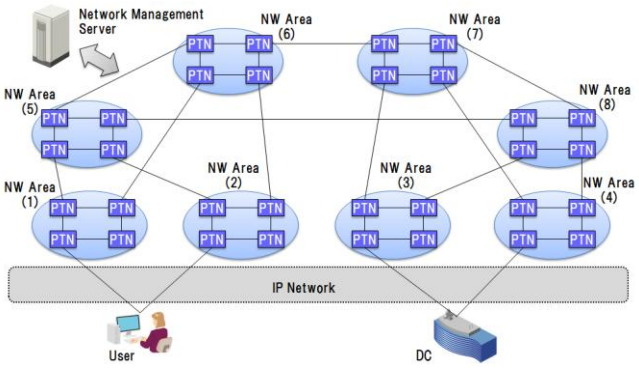


Figure 2. Proposed network-disaster recovery system

B. Overview of network-disaster recovery system

The procedures used by the proposed system for network-disaster recovery are overviewed in Figure 3. In the first procedure, the network-management server divides an entire PTN network into eight network areas, labelled (1) to (8) in the figure, and controls them by using the area-based network-management scheme. In addition, it configures the path shown as a solid line in the figure as the current path so that the user can access the server in the DC and use application services.

In the second procedure, the network-management sever calculates all recovery paths preliminarily by considering all

possible area-based failures. Specifically, the number of possible area-based failure patterns is 255 (since there are eight areas, and each area could be independently active or not active), namely, 256 (i.e., 2^8) patterns minus a "no area failure" pattern that is the current network operation. The network-management server assigns a recovery ID for each area-based network failure pattern and stores each recovery ID with information on the recovery paths. It then distributes all recovery IDs and the recovery-path information to all PTNs preliminarily. In Figure 3, it is assumed that network areas (1), (3), and (6), as stated in the figure, fail. In the case of these failures, a path depicted by a dashed line is prepared as a recovery path, and the recovery-path information is distributed to PTNs related to the recovery path before network operations are started.

During network operations, the network-management server monitors area-based network failures. When it detects area-based network failures, it determines a failure pattern and a recovery ID, and it then distributes the recovery ID to related PTNs, a user terminal, and a server in the DC. The PTNs that receive the recovery ID start to promptly recover and transmit packet data according to the recovery-path information specified by the ID. In addition, the network-management server configures IP networks to transmit packet data from the user terminal to network area (2). Alternately, it transmits a request that asks the user terminal to change an output port so as to transmit packet data to another active network area, if necessary. Besides, the network-management server configures IP networks to transmit packet data from network area (4) to the server in the DC.

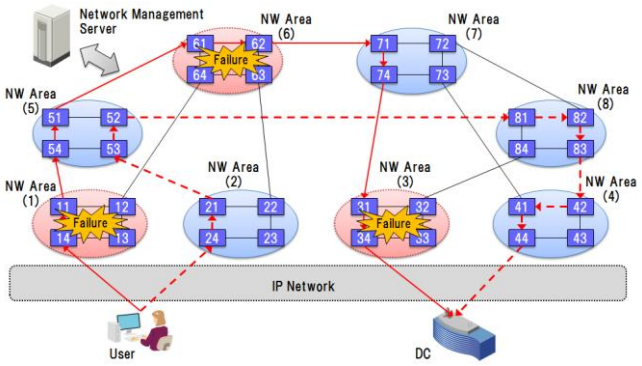


Figure 3. Proposed procedures for network-disaster recovery

C. Sequence of network-disaster recovery

The proposed network-disaster recovery follows the sequence shown in Figure 4. First, the network-management server divides the entire PTN network into multiple network areas and manages each area by using the area-based network-management scheme, labeled "area mgmt" in Figure 4. Specifically, the PTN networks are divided into eight areas and managed as shown in Figure 3. Subsequently, the network-management server calculates the current path (which is composed of the LSP and PW) for transmitting packet data from the user terminal to the server in the DC, shown as "current path" in Figure 4. It starts network operations by configuring the calculated path to related PTNs.

As for the calculation of a path, a route that can provide required bandwidths and transmit packet data within allowed delays is selected as the current path. In addition, the network-management server configures the current path to PTNs, shown as "current-path configuration" in the figure.

The network-management server then calculates all recovery paths by considering all possible area-based network failures, shown as "recovery path". Specifically, it calculates a recovery path for each possible area-based network failure, as shown in Table I. Each recovery path (labelled "P1" in the table) is identified by a recovery ID from "0" to "255." The top row of the table, containing recovery ID "0", indicates current-recovery-path configurations for no area-based network failure. The next row in the table, containing recovery ID "1", indicates recovery-path configurations for a failure of network area (1). In this case, the network failure in the area (1) is assumed. The recovery path for "P1" is calculated on the basis of available network resources. In other words, network resources in area (1) are excluded from the available resources, and the recovery path is calculated. The next row in the table, containing recovery ID "2", indicates the recovery-path configurations for a failure of network area (2). The row of the table containing recovery ID "38" indicates the recovery-path configurations in the case of failures of network areas (1), (3), and (6). As an example recovery path, the dashed line in Figure 3 is that for the current path depicted by the solid line. In Figure 3 and Table I, the recovery-path information for only path "P1" is shown as an example. However, the proposed system is able to manage multiple paths.

As the next step of the recovery sequence, the network-management server calculates recovery-path configurations for each node in case of each area-based network-failure pattern according to the recovery-path information shown in Table I. In addition, the recovery-path tables for PTN 53 and 54 are shown in Table II. The top row of the table, containing recovery ID "0" on PTN 53, shows the current configuration (i.e., "Connection 1" and "Connection 2"). With regard to PTN 53, path P1 (composed of an LSP and a PW) is not configured, since it does not transmit the related packet data. The next row of the table, containing recovery ID "1", indicates the configuration for recovery path P1 in case of a failure of network area (1). Specifically, it is shown that PTN 53 transmits packet data of P1 from PTN 21 to PTN 54 and from PTN 54 to PTN 21. In addition, the row of the table containing recovery ID "2" indicates the recovery-path configurations in the case of a failure of network area (2). However, P1 is not configured. Besides, the row of the table containing recovery ID "38" indicates the configurations of recovery path P1 in the case of failures of network areas (1), (3), and (6). Specifically, it is shown that PTN 53 transmits packet data of path P1 from PTN 21 to PTN 52 and from PTN 52 to PTN 21.

In the lower half of the table, recovery-path configurations on PTN 54 are indicated. The row of the table containing recovery ID "0" shows the current configuration. As shown in the table, PTN 54 transmits packet data of path P1 from PTN 11 to PTN 51 and from PTN 51 to PTN 11.

 ISBN: 978-1-61208-373-5

The row of the table containing recovery ID "2" indicates the recovery-path configurations in case of a failure of network area (2). PTN 54 transmits packet data of path 1 from PTN 11 to PTN 51 and from PTN 51 to PTN 11. In addition, the row of the table containing recovery ID "38" indicates the recovery-path configuration in the case of failures of network areas (1), (3), and (6). However, recovery path 1 is not configured, since PTN 54 does not transmit path-1-related packet data. After the network-management server calculates all recovery-path tables shown in Table II, it distributes them to all PTNs. When each PTN receives the configurations, it stores them with each recovery ID.

In the next step of the recovery sequence, the network-management server monitors operations of all PTNs and area-based network failures, shown as "monitoring" in Figure 4. For example, the network-management server detects failures of network areas (1), (3), and (6) shown in Figure 3. In this case, the network-management server selects recovery ID 38 to recover the configured path, shown as "recovery decision". The PTNs receive recovery ID 38 and configure a data-transmission function to transmit packet data according to the recovery-path information specified by the recovery ID 38.

In the next step, the network-management server configures IP networks to transmit packet data from the user terminal to PTN 24. In addition, it configures IP networks to transmit packet data from PTN 44 to the server in the DC. By executing the above-described recovery procedures, failures of network areas (1), (3), and (6) are recovered.

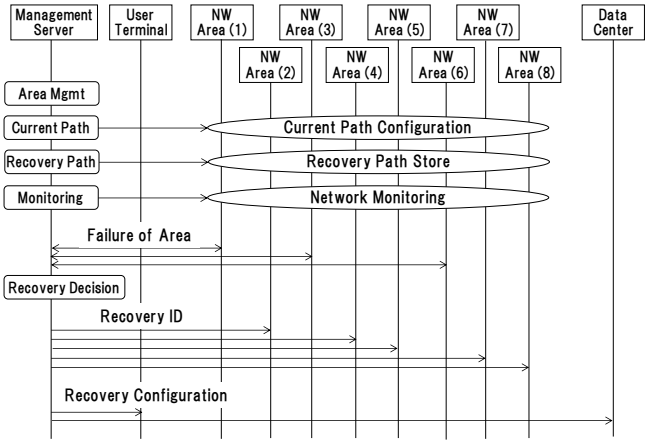


Figure 4. Sequence of network-disaster recovery

TABLE I. RECOVERY-PATH CONFIGURATION

Failure Pattern	Recovery ID	Path	Recovery Path Configuration
No failure	0	P1	14, 11, 54, 51, 61, 62, 71, 74, 31, 34
Area (1) failure	1	P1	24, 21, 53, 54, 51, 61, 62, 71, 74, 31, 34
Area (2) failure	2	P1	14, 11, 54, 51, 61, 62, 71, 74, 31, 34
- - -	- - -	- - -	- - -
Area (1), (3), (6) failures	38	P1	24, 21, 53, 52, 81, 82, 83, 42, 41, 44
- - -	- - -	- - -	- - -
All area failures	255	P1	No Recovery

TABLE II. RECOVERY-PATH TABLE FOR EACH PTN

PTN	Recovery ID	Path(LSP/PW)	Connection 1	Connection 2
53	0	- - -	- - -	- - -
	1	P1	21	54
	2	- - -	- - -	- - -
	- - -	- - -	- - -	- - -
	38	P1	21	52
	- - -	- - -	- - -	- - -
54	0	P1	11	51
	1	- - -	- - -	- - -
	2	P1	11	51
	- - -	- - -	- - -	- - -
	38	- - -	- - -	- - -
	- - -	- - -	- - -	- - -

D. Calculation of recovery paths for possible failure patterns

The flow for calculating a recovery path for an area-based network failure is shown in Figure 5. After the recovery-path calculation starts, delays and available bandwidths between PTNs are calculated from a database that includes topology information and available resources such as link bandwidths. Next, one of the possible area-based network failures, for example, failure of network area (1), is assumed. After that, the PTNs belonging to the assumed area failure are excluded from the available resources to calculate recovery paths. After available resources such as PTNs and bandwidth are fixed, one of the established PWs is selected as the recovery path. Then, the minimum delay path that has the same starting and ending points is selected as the recovery path. If the recovery path is not found because of link disconnection, etc., a message indicating "lack of resources" to find the recovery path is displayed, and the recovery-path calculation process moves on to the next step, namely, selection of another PW. If the recovery path is found, whether it meets the allowed delay time or not is checked. If the path does not meet the allowed delay time, a "lack of available resources" message is displayed, and the process moves on to the next step to find a recovery path for another PW. If the path meets the allowed delay time, it is determined as the recovery path. After the recovery path is confirmed, available bandwidth is decreased by the amount of bandwidth consumed by the recovery path itself. Subsequently, if the route of the LSP path is not the same as the previously calculated route, it is stored as a new LSP route. Then, whether all recovery paths for a selected area-based network-failure pattern have been calculated or not is checked. If all the recovery paths are not calculated, the process moves on to the next step, that is, selection of another PW. If all recovery paths for one area-based network-failure pattern are calculated, whether all recovery paths for all possible area-based network-failure patterns are calculated or not is checked. When all recovery paths for all possible area-based network failure patterns are calculated, the recovery-path calculation process stops.

 ISBN: 978-1-61208-373-5

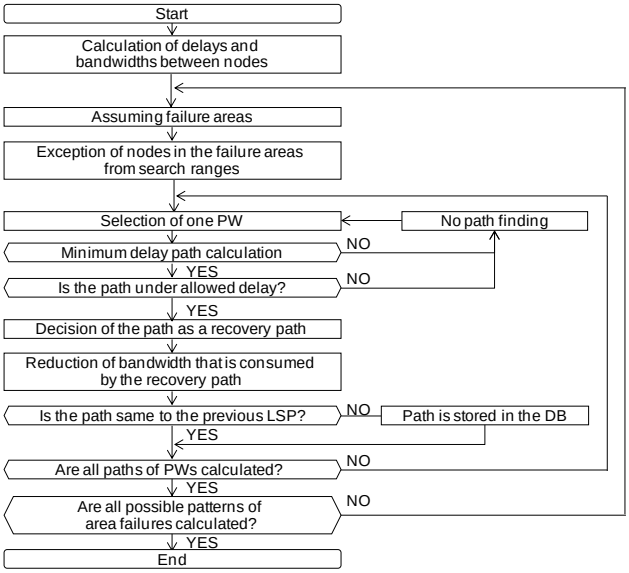


Figure 5. Calculation of recovery paths

All recovery paths are calculated, and the recovery- path information is distributed to all network nodes before network operations are started. The nodes can therefore select an appropriate recovery path swiftly when a network fails.

IV. EVALUATION AND RESULTS

The above-described recovery procedures were evaluated in the case of multiple area-based network failures. The evaluations were intended for networks composed of IP and PTN networks. First, current paths composed of LSPs and PWs were configured to allow users to access application servers in the DC and use applications provided by the server. In the evaluation, the recovery procedure to recover from multiple area-based network failures by using recovery paths was evaluated in terms of whether users can access the application servers or not. In addition, time for calculating the current recovery paths and distributing the information concerning the calculated paths to all PTNs was evaluated by changing the numbers of LSPs and PWs used to construct the current paths.

A. Evaluation system

The system used for the evaluation is depicted in Figure 6. It is composed of a network-management server, PTNs, a user terminal, and an application server in a DC. An entire PTN network is divided into eight network areas. Each network area is composed of 12 PTNs, as shown in area 7, which is an example network composing of about 100 network nodes. These PTNs are connected in a reticular pattern of 96 PTNs in total. In addition, the user terminal is directly connected to PTN-network areas (1) and (2) by IP networks. In addition, the application server is connected to PTN-network areas (3) and (4) directly by IP networks.

Note that the PTN networks (composed of 96 PTNs) are simulated by a physical server. In addition, the user terminal and application server in the DC are also simulated by the same physical server. The specification of the physical server that simulates the PTN networks, user terminal, and application server is listed in Table III. In addition, another physical server that executes the network-management function has the same specifications as the simulator server.

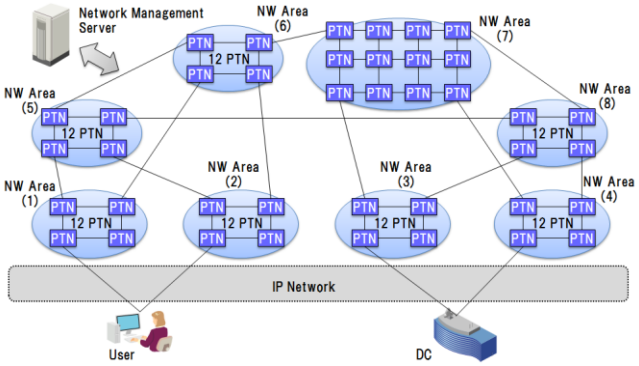


Figure 6. Evaluation system

TABLE III. SPECIFICATIONS OF SERVER

#	Item	Specifications
1	CPU	1.8 GHz, 4 cores
2	Memory	16 GB
3	Storage	600 GB

B. Evaluation conditions

The time taken to calculate PWs by using one route between the user and the application server in the DC was evaluated. As an evaluation condition, one LSP between the user and the application server was established. The LSP houses 10 PWs since it usually houses multiple PWs. The evaluations were executed according to the patterns listed in Table IV. Specifically, time to calculate current paths and recovery paths for 255 area-based network failure patterns was evaluated by changing the number of PWs (namely, 100, 500, and 1000). In addition, the time to distribute all calculated recovery-path configurations and recovery IDs was also evaluated.

TABLE IV. EVALUATION ITEMS

#	Item	Specifications
1	Current path calculation time	Time to calculate 100, 500, and 1000 PWs
2	Recovery path calculation time	Time to calculate recovery 100, 500, and 1000 PWs for 255 possible area failure patterns
3	Distribution time	Time to distribute all calculated recovery PWs and LSPs for 255 possible area failure patterns
4	Recovery ID distribution time	Time to distribute a recovery ID after detecting a first area failure

C. Evaluation result

1) Current-path calculation time

The times taken to calculate current PWs using one route are plotted in Figure 7. The evaluation condition is that 10 PWs are housed in one LSP. As shown in the figure, the times taken to calculate 100 current PWs, 500 current PWs, and 1000 current PWs were about 64, 344, and 769 milliseconds, respectively.

 ISBN: 978-1-61208-373-5

2) Recovery-path calculation time

The times taken to calculate all recovery PWs for 255 possible area-based network-failure patterns by using one route are plotted in Figure 8. The evaluation condition is that 10 PWs are housed in one LSP. As shown in the figure, the time taken to calculate all recovery PWs for 255 area-based network-failure patterns and 100 current PWs, 500 current PWs, and 1000 current PWs are about 5.2, 35.9, and 114.2 seconds, respectively.

3) Distribution time for recovery paths

The times taken to distribute all configurations of calculated recovery PWs to all PTNs are plotted in Figure 9. The evaluation condition is that 10 PWs are housed in one LSP. As shown in the figure, times taken to distribute all configurations of recovery PWs for 255 area-based network-failure patterns and the 100 current PWs, 500 current PWs, and 1000 current PWs are about 245, 282, and 455 milliseconds, respectively.

4) Recovery ID distribution time

The evaluated times taken to distribute the recovery ID to related PTNs after the first area-based network failures are detected are shown in Figure 10. The evaluation condition is that 10 PWs are housed in one LSP. The evaluations were executed for 100 current PWs, 500 current PWs, and 1000 current PWs. In the case of three area-based network-failure patterns, namely, a failure of network area (5), failures of network areas (1), (5), and (8), and failures of network areas (1) and (2) are evaluated. As shown in the figure, the time taken to distribute the recovery ID depends on the number of area-based network failures. According the figure the time taken to distribute the recovery ID ("recovery-ID distribution time" hereafter) in case of one area failure is the shortest for the three area-based network-failure patterns. This tendency is the same for each current PW number. In addition, the recovery-ID distribution time in the case of three area failures is the longest. In particular, the recovery-ID distribution time for 1000 PWs in the case of failures of network areas (1), (5), and (8) is about 3.4 seconds. However, the recovery-ID distribution time includes finding multiple network failures. Therefore, the recovery-ID distribution time depends on the number of area-based network failures. If the time taken to find network failures is excluded, the recovery-ID distribution time itself is independent of the number of the area-based network failures and is always less than 100 milliseconds, as shown in the case of the failure of network area (5). In this sense, the proposed system is useful for not only single area failure but also multiple area failures.

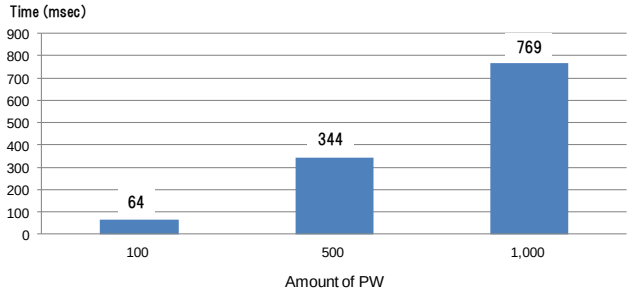


Figure 7. Time for calculating current paths

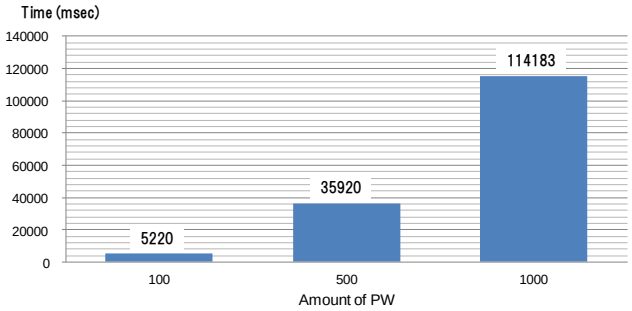


Figure 8. Time for calculating recovery paths

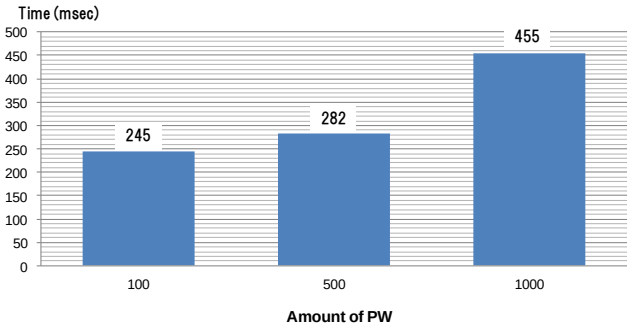


Figure 9. Time for distributing recovery paths

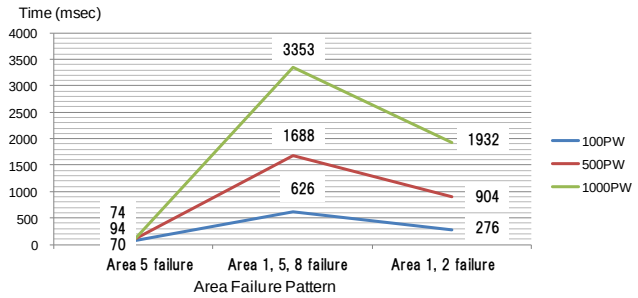


Figure 10. Time for distributing recovery ID

D. Discussion

As shown in Figure 10, the time taken to distribute the recovery ID after the first area-based network failure is detected was evaluated. If two or three area-based network failures occur, the time taken to distribute the recovery ID includes the time taken to detect the second and third area-based network failures after the first area-based network failure is detected. Consequently, the more area-based network failures occur, the longer the time taken to distribute the recovery ID. On the other hand, "pure" recovery-ID distribution time (namely, recovery-ID distribution time excluding the time taken to detect network failures) is shown as the time in Figure 10 in the case of only one area-based network failure. Namely, recovery-ID distribution time in the case that network area (5) fails is equivalent to the pure recovery-ID distribution time. The recovery-ID distribution time is under 100 milliseconds for any number of PWs (i.e., 100, 500, and 1000).

 ISBN: 978-1-61208-373-5

The time taken to calculate 100 current PWs is about 64 milliseconds, which is shorter than the time taken to distribute the recovery ID (i.e., 100 milliseconds). However, the time to calculate 500 current PWs is about 344 milliseconds, which is longer than the time taken to distribute the recovery ID. As a result, if there are over 500 PWs, the proposed network-disaster recovery system can start to recover faster by using preliminarily calculated configurations of recovery PWs and distributing the recovery ID than by recalculating the recovery PWs after an area failure is detected under the evaluation conditions described above.

E. Comparison of proposed system and conventional system

In case of a conventional network system, a restoration scheme is basically used when catastrophic network failures occur. In other words, a large number of setup paths are recalculated after finding the network failures. According to Figure 7, it takes 769 milliseconds to calculate paths for 1000 PWs. If there are 100,000 PWs, it may take over 10 minutes to calculate the paths. That is, over 10 minutes are needed to calculate recovery paths for the 100,000 PWs setup after the network failures were found. On the other hand, in the case of the proposed system, information to recover all the setup paths is distributed to all the network nodes (such as PTNs). The recovery-ID distribution time after finding the network failures is less than 100 milliseconds. Therefore, even if there are 100,000 PWs, the proposed system can start recovery within 100 milliseconds after finding network failures.

With regard to cost, compared to conventional systems (which use a restoration scheme), the proposed system needs more memory (storage) capacity to keep the recovery paths. However, memory and/or storage costs have been gradually decreasing, so the proposed system is promising for the near future.

V. RELATED WORK

Regarding highly available and reliable network management, several standardization activities have been ongoing. For example, MPLS-TP-related Operation, Administration, and Maintenance (OAM) has been standardized. In the first stage, in the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T)[4], specifications such as Transport – Multi Protocol Label Switching (T-MPLS) were discussed. In the next stage, the ITU-T jointly standardized MPLS-TP [6][7][8][9] specifications with the Internet Engineering Task Force (IETF)[5]. Using MPLS-TP OAM functions makes it easy to detect network failures in transport networks. In relation to the proposed system, it is useful to detect network failures promptly to determine areas that are out-of-service.

With regards to failure recoveries, two major techniques have been proposed: “protection” [10][11][12][13][14] and “restoration” [15][16][17][18][19]. With the protection technique, a standby path is preliminarily calculated and established by using extra physical resources. When network failures are detected, an active path is promptly changed from a current path to the standby path. With this technique, when a large number of standby paths are prepared in the case of multiple network failures, physical resources might be voluminously needed. It is therefore useful for limited network failures such as failures of a few link only. On the other hand, with the restoration technique, recovery paths are calculated one by one after network failures are detected. This scheme is useful for catastrophic network failures since all reroutes are calculated after the failures are detected. However, if there are a large number of current paths, much time might be needed to calculate all recovery paths to the current paths.

VI. CONCLUSION

A “network-disaster recovery system” using area-based network management is proposed. As for this system, a whole network is separated into multiple areas. Each area is composed of multiple network nodes, such as MPLS-TP nodes. The system is managed by a network-management server that monitors the condition of every network node. The network-management server manages the network by detecting area-based failures. It calculates recovery paths for every possible area failure and distributes them with a recovery ID for each area-failure pattern before starting network operations. The network nodes receive and store the recovery-path configuration and recovery ID. The network-management server detects the network-area failures during network operations and determines a pattern of area failures. Specifically, it determines numbers and positions of area failures. After determining the pattern of area failures, the network-management server selects a recovery ID to recover the area failures and distributes the ID to recovery-related network nodes. The network nodes receive the recovery ID and start data transmission based on the path configuration specified by the distributed ID. After these procedures are completed, the area failures are swiftly recovered.

A prototype system composed of a network-management server and 96 simulated packet-transport nodes was constructed and evaluated. The system could calculate 500 PWs as current paths that are accommodated in 50 LSPs in about 344 milliseconds. That is, it takes about 344 milliseconds to calculate recovery paths in the case of a network-area failure. Recovery paths of all the current PWs for 255 network-area failure patterns were calculated in about 36 seconds. With the proposed system, however, this calculation is done before network operations start. The network-management server could transmit the recovery ID to the related network nodes within 100 milliseconds after a network-area failure is detected, and the system could immediately start to recover from the failure. The recovery-ID distribution time is shorter than the time required for calculating recovery paths for 500 PWs. In other words, if there are over 500 PWs, the proposed system can start to recover faster than a conventional system (which recalculates the recovery paths after detecting the network failures) under the same conditions as those in the present evaluation.

 ISBN: 978-1-61208-373-5

As for the prototype system, the whole network is divided into eight areas as one of examples to divide the whole network into multiple area networks. However, scalability of this approach is an issue. For example, a recovery scheme is needed when only one link or node failure occurs. The prototype system will therefore be further developed so it can manage a larger range of failures (from small ones to large ones).

ACKNOWLEDGMENT

Part of this research was included in Research Project 0_3 (Open, Organic, Optima) and was supported by the MIC (Japanese Ministry of Internal Affairs and Communications) program, "Research and Development on Virtualized Network Integration Technology".

REFERENCES

[1] Internet World Stats, http://www.internetworldstats.com/stats.htm [retrieved: August, 2014].

[2] C. L. Belady, Microsoft Corporation, "Projecting annual new datacenter construction market size," Mar. 2011 http://cdn.globalfoundationservices.com/documents/Projecting_Annual_New_Data_Center_Construction_PDF.pdf[retrieved: August, 2014].

[3] A. Bianco, J. Finochietto, L. Giraudo, M. Modesti, and F. Neri, "Network planning for disaster recovery," 16th IEEE Workshop on Local and Metropolitan Area Networks, LAMAN 2008, PP. 43-48, Sept. 2008.

[4] International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) http://www.itu.int/en/ITU-T/Pages/default.aspx [retrieved: August, 2014].

[5] The Internet Engineering Task Force (IETF), http://www.ietf.org/ [retrieved: August, 2014].

[6] B. Niven-Jenkins, D. Brungard, M. Betts, N. Sprecher, and S. Ueno, "Requirements of an MPLS transport profile," RFC 5654, Sept. 2009.

[7] M. Bocci, S. Bryant, D. Frost, L. Levrau, and L. Berger, "A Framework for MPLS in transport networks," RFC 5921, July 2010.

[8] T. Busi and D. Allan, "Operations, administration, and maintenance framework for MPLS-based transport neworks," RFC 6371, Sept. 2011.

[9] N. Sprecher and A. Farrel, "MPLS transport profile (MPLS-TP) survivability framework," RFC 6372, Sept. 2011.

[10] M. Pickavet, P. Demeester, and D. Colle, "Recovery in multilayer optical networks," Journal of Lightwave Technology, Vol. 24, no. 1, pp. 122-134, Jan. 2006.

[11] J. Zhang, J. Zhou, J. Ren, and B. Wang, "A LDP fast protection switching scheme for concurrent multiple failures in MPLS network," 2009 MINES '09. International Conference on Multimedia Information Networking and Security, pp. 259-262, Nov. 2009.

[12] Z. Jia and G. Yunfei, "Multiple mode protection switching failure recovery mechanism under MPLS network," 2010 Second International Conference on Modeling, Simulation and Visualization Methods (WMSVM), pp. 289-292, May 2010.

[13] G. Kuperman and E. Modiano, "Network protection with guaranteed recovery times using recovery domains," INFOCOM, 2013 Proceedings IEEE, pp. 692-700, April 2013.

[14] J. Rack, "Fast service recovery under shared protection in WDM networks," Journal of Lightwave Technology, Vol. 30, no. 1, pp. 84-95, Jan. 2012.

[15] A. Valenti, P. Bolleta, S. Pompei, and F. Matera, "Experimental investigations on restoration techniques in a wide area gigabit Etherneti optical test bed based on virtual private LAN service," 11th International Conference on Transparent Optical Networks, ICTON '09, We.B3.4, June 2009.

[16] M. Lucci, A. Valenti, F. Matera, and D. Del Buono, "Investigation on fast MPLS restoration technique for a GbE wide area transport network: A disaster recovery case," 12th International Conference on Transparent Optical Networks (ICTON), Tu.C3.4, June 2010.

[17] D. Sheela, M. Smitha Krishnan, and C. Chellamuthu, "Combined link weight based restration Strategy in optical networks," 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), pp. 687-690, March 2012.

[18] T. S. Pham. J. Lattmann, J. Lutton, L. Valeyre, J. Carlier, and D. Nace, "A restoration scheme for virtual networks using switches," 2012 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 800-805, Oct. 2012.

[19] X. Wang, X. Jiang, C. Nguyen, X. Zhang, and S. Lu, "Fast connection recovery against region failures with landmark-based source routing," 2013 9th International Conference on the Design of Reliable Communication Networks (DRCN), pp. 11-19, Mar. 2013.

[i] Ethernet is a registered trademark of Xerox Corporation.

 ISBN: 978-1-61208-373-5

What's Happening: A Survey of Tweets Event Detection

Amina Madani
Blida 1 University
LRDSI Laboratory
Blida, Algeria
a_madani@esi.dz

Omar Boussaid
Lumière lyon 2 University
ERIC Laboratory
Lyon, France
Omar.Boussaid@univ-lyon2.fr

Djamel Eddine Zegour
High School of Computer Science
LCSI Laboratory
Algiers, Algeria
d_zegour@esi.dz

***Abstract*—Twitter is now one of the main means for spread of ideas and information throughout the Web. Tweets discuss different trends, ideas, events, and so on. This gave rise to an increasing interest in analyzing tweets by the data mining community. Twitter is, in nature, a good resource for detecting events in real-time. In this survey paper, we are going to present four challenges of tweets event detection: health epidemics identification, natural events detection, trending topics detection, and sentiment analysis. These challenges are based mainly on clustering and classification. We review these approaches by providing a description of each one.**

Keywords—tweets mining; tweets event detection; health epidemics identification; natural events detection; trending topics detection; sentiment analysis

I. INTRODUCTION

In recent years, hundreds of millions of users participate in online social networks and forums, subscribe to microblogging services or maintain web diaries (blogs).

Twitter, in particular, is currently the major micro-blogging service. In January 2014, the total number of active registered users on Twitter was 645,750,000 [1]. In this system, participants post short status messages that are often available publicly, or semi-publicly (e.g., restricted to the user's designated contacts). In January 2014, more than 58 millions of Twitter messages were sent every day through the world, with 9,100 the number of tweets that happen every second [1].

Recently, the exponential growth of Twitter messages has started to draw the attention of researchers from various disciplines. Numerous works in the data mining field have examined Twitter. How to automatically understand, extract and summarize useful Twitter content to detect events has therefore become an important and emergent research topic. Tweets event correspond to a lot of content generated on Twitter, which is opinions, reactions and information from users. Currently, the most promising events that exist on Twitter are: natural disasters such as earthquakes, health epidemics like influenza, trends such as world-cup and opinion about products, services or events like political election results.

The focus of this survey paper is to discuss research works done for four challenges of tweets event detection: health epidemics identification, natural events detection, trending topics detection, and sentiment analysis. Tweets event detection helps understanding what the users are really discussing about in Twitter.

This paper is organized as follows: In the next section, we explain Twitter service and its characteristics. In Section 3, we review popular tweets event detection approaches by providing a description of each one. Section 4 discusses and highlights works studied above. Finally, we conclude the paper and further work is outlined.

II. TWITTER, A POPULAR SAS

Twitter is a popular microblogging service, one of the main means for spread of ideas and information throughout the web. Twitter enables its users to send and read short text-based messages of up to 140 characters about "what's happening" (either one or two sentences), known as "tweets". Users can tweet via the Twitter website, compatible external applications (such as for smartphones), or by Short Message Service (SMS) available in certain countries.

One of the main characteristics of Twitter is that its core functions represent a *Social Awareness Stream* model. The *Social Awareness Streams* (SAS) are typified by three factors distinguishing them from other communications [2]:

- The public (or personal-public) nature of the communication and conversation;
- The brevity of posted content; and,
- A highly connected social space, where most of the information consumption is enabled and driven by articulated online contact networks.

Twitter has users of the order of hundreds of millions (Figure 1). A huge amount of content is generated every second, minute and hour of the day. Every tweet is associated with an explicit timestamp that declares the exact time it was generated. Important characteristic of Twitter is its real-time nature [3]. Tweets are data stream arriving in real-time. Data stream are data that arrive at high speed and its nature or distribution changes over time.

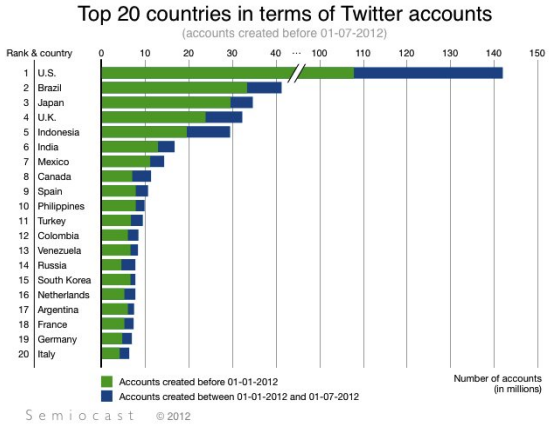


Figure 1. Top 20 countries in terms of Twitter accounts [4].

 ISBN: 978-1-61208-373-5

Also, another property of tweets is the rich set of fields associated with the content, which is usually presented in the form of semi-structured documents. Figure 2 presents an example of a tweet. While the contents are ostensibly 140 unstructured characters, the anatomy of a tweet reveals lots of metadata (e.g., location, location) and even the contents contain some structural information (e.g,. RT indicating re-tweet or #hashtags serving as topical metadata). Moreover, every user has a well-defined profile with personal information (name, location, biographical sketch).

Figure 2. An example of a tweet.

All these characteristics gave rise to an increasing interest in analyzing tweets by the data mining community. Exploiting these characteristics can be helpful for improving tweets event detection.

Tweets event correspond to a lot of content generated on Twitter, which is opinions, reactions and information from users.

III. TWEETS EVENT DETECTION: STATE OF THE ART

In the knowledge discovery context, there are two fundamental data mining tasks that can be considered in conjunction with Twitter data [5]:

- Graph mining based on analysis of the links amongst messages.
- Text mining based on analysis of the messages content.

In our study, we are interested in the tweets content using text mining. Tweets content have become an important channel for reporting real-world events. We describe an event by four main dimensions:

- *Event type*: what is happening?
- *Time*: when an event is happening?
- *Location*: where an event is happening?
- *Entities*: who is involved in an even?

Health epidemics identification (e.g. influenza), natural events detection (e.g. earthquakes), trending topics detection, and sentiment analysis (e.g. political events) are four considerable challenges for tweets event detection presented in Figure 3. These challenges are based mainly on clustering and classification.

In this part, we are going to explain more specifically, these challenges and present several research efforts that have focused on them for identifying events on Twitter.

A. Health epidemics identification

In recent years, a new research area has been developed, namely "Infodemiology". It can be defined as the "science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [6].

Twitter presents a promising new data source for Internet-based surveillance because of message volume, frequency, and public availability. Recent studies have begun to use Twitter data to understand their applicability in the context of Health epidemics identification. Twitter can be a low cost alternative source for tracking health epidemics. Health epidemics identification is based on classification in order to predict some illnesses while analyzing textual content of tweets.

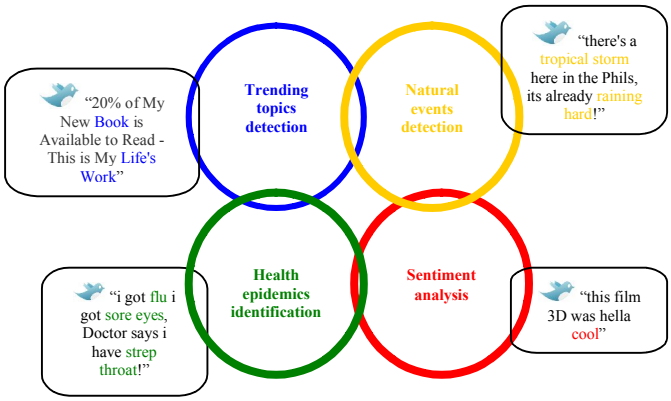


Figure 3. Challenges of tweets event detection.

Previous work in this area has focused specifically on influenza. Culotta [7] considers tweets as a valuable resource for tracking influenza for several reasons:

- The high message posting frequency enables up-to-the-minute analysis of an outbreak.
- Twitter messages are longer, more descriptive, and publicly available.
- Twitter profiles often contain semi-structured meta-data (city, state, gender, age), enabling a detailed demographic analysis.
- Despite the fact that Twitter appears targeted to a young demographic, it in fact has quite a diverse set of users.

The work of Culotta [7] explores the possibility of detecting influenza outbreaks by analyzing Twitter data. The author uses a bag-of-words classifier in order to predict *influenza-like illnesses* (ILI) rates in a population, based on the frequency of messages containing certain keywords. He compares rates with the *U.S. Centers for Disease Control and Prevention* (CDC) statistics.

A new method called *Ailment Topic Aspect Model* (ATAM) for extracting general public health information from millions of health related tweets is introduced by Paul and Dredze [8]. The approach discovers many different ailments (diseases), such as flu, allergies, or cancer and learns symptom and treatment associations. This model discovers a larger number of more coherent ailments than *Latent Dirichlet Allocation* (LDA) [9]. It produces more detailed ailment information

 ISBN: 978-1-61208-373-5

(symptoms/treatments) and tracks disease rates consistent with published government statistics (influenza surveillance) despite the lack of supervised influenza training data.

B. Natural events detection (Disasters)

An important characteristic of Twitter service is their real-time nature. Users write tweets several times in a single day. Sakaki et al. [3] consider each Twitter user as a sensor and each tweet as sensory information. A sensor detects a target event and makes a report probabilistically. Each tweet is associated with a time and location, which is a set of latitude and longitude. These virtual sensors, which they call social sensors, are of a huge variety and have various characteristics. Detecting a natural event in real time is reduced to detecting an object and its location by regarding a tweet that has both spatial and temporal regions.

Some researchers have shown how these social sensors can be useful to describe the current situation during natural disasters. The objective is to mine tweets to detect natural disaster in real-time events such as earthquakes, floods or volcanic eruptions.

The automatic detection of natural events from tweets consists in developing *Business Intelligence* (BI) tools to detect information before they appear in the agencies.

A method for automatic detection of disasters in real-time is proposed by Rosoor et al. [10]. They develop a tool for journalists / librarians to detect the information before they appear in the agencies. The automatic detection method uses heterogeneous resources from the Web (blogs, tweets, RSS, etc.). This tool is based on a "salton" representation [11] of a corpus classified into different topics (flood, earthquake, and so forth). Each category is represented as a vector of words. The new text to be classified is compared to these vectors in order to identify the relevant category.

Sakaki et al. [3] investigate the real-time interaction of events, such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, they devise a classifier of tweets using a support vector machine [12] based on features, such as the keywords in a tweet, the number of words, and their context. Subsequently, they produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. They can detect an earthquake with high probability (96% of earthquakes of Japan Meteorological Agency (JMA) seismic intensity scale 3 or more are detected) merely by monitoring tweets. The system detects earthquakes promptly and sends e-mails to registered users. Notification is delivered much faster than the announcements that are broadcast by the JMA and possibly before an earthquake actually arrives at a certain location.

The work of Cheong and Cheong [13] performs analysis on tweets during Australian floods of 2011 to identify active players and their effectiveness in disseminating critical information. They identified the most prominent users among Australian floods to be: local authorities (Queensland Police Services), political personalities (Premier, Prime Minister, Opposition Leader and Member of Parliament), social media volunteers, traditional media reporters, and people from not-for-profit, humanitarian, and community associations.

C. Trending topics detection

Twitter messages are being posted with vast amount of new information and changes continuously. They are a live stream that contains a great wealth of information where topics of discussion shift dynamically with time. However, it is not practical for us to browse tweets manually all the time for searching about the latest most discussed issues and thus revealing the emerging topics of our interest.

Analyze tweets to detect trending topics (trends) in real time is a big challenge. Trending topics detection has primarily involved analyzing the content of tweets. We define a trending topic as an emerging keyword which links to a very recent event. This keyword is experiencing an increase in usage in Twitter messages. Trending topics are typically driven by emerging events, breaking news and general topics that attract the attention of a large fraction of Twitter users [14]. Unspecified trending topics of interest are typically driven by topics that attract the attention of a large number of users.

Twitter allows users to observe the top ten popular terms or topics of discussion at any given moment. But, we must browse related tweets manually all the time for viewing more detail about these trending topics. It is important to automatically analyse, understand, extract and summarize useful tweets content.

Twitter is a new form of newspapers. Trending topics detection in real-time is thus of high value to news reporters, analysts, and e-marketing specialists.

We present several approaches effo*rts* that have focused on trending topics detection for identifying events on Twitter.

Cheong and Lee [15] analyze tweets to research the anatomy of trending topics. They split them into 3 categories: long-term, medium-term and short-term topics. Long-term topics occur infrequently, but over a long amount of time in the public time-line, while medium-term topics occur more frequently. However, the medium-term topics are limited to a time range of a few days. Short-term topics are heavily discussed topics, and often refer to current events. Also, they categorize the users into 3 major groups: “Personal”, “Aggregator” and “Marketing”. The results show that mostly users who talk about their personal life contribute to emerging trending topics.

TwitterMonitor, a system that performs trend detection is proposed by Mathioudakis and Koudas [14]. The system identifies emerging topics (trends) on Twitter in real-time. A trend is identified as a set of bursty keywords that occur frequently together in tweets. TwitterMonitor provides meaningful analytics that synthesize an accurate description of each topic. It extracts additional information from the tweets that belong to the trend, aiming to discover interesting aspects of it. Users interact with the system by ordering the identified trends using different criteria and submitting their own description for each trend.

The main objective in the work of Naaman et al. [2] is to identify different types of user activity, specifically focusing on message content and its relationship to patterns of use. To characterize the type of messages posted on Twitter they use a grounded approach to thematize and code a sample of 200 tweets. First, the three authors independently assigned categories to the downloaded messages. They then proceeded

 ISBN: 978-1-61208-373-5

to analyze the affinity of the emerging themes to create an initial set of coding categories. Next, they downloaded a second set of 200 posts, categorized them, then reflected on and adapted the initial categories based on the additional input.

Benhardus [16] outlines methodologies of detecting and identifying trending topics from streaming data. *Term frequency-inverse document frequency* (TF-IDF) analysis and relative normalized term frequency analysis are performed on the tweets to identify the trending topics. Relative normalized term frequency analysis identifies unigrams, bigrams, and trigrams as trending topics, while term frequency-inverse document frequency analysis identifies unigrams as trending topics.

Cataldi et al.'s approach [17] proposes a novel approach to detect in real-time emerging topics on Twitter. They extract the contents (set of terms) of the tweets and model the term life cycle according to a novel aging theory intended to mine terms that frequently occur in the specified time interval and they are relatively rare in the past. Moreover, considering that the importance of content also depends on its source, they studied the social relationships in the user network in order to determine the authority of the users. Finally, they formalized a keyword-based topic graph which connects the emerging terms with their co-occurrent ones, allowing the detection of emerging topics under user-specified time constraints.

Budak et al. [18] introduce new methods for identification of important topics that utilize the network topology. They propose two novel trend definitions called coordinated and uncoordinated trends that detect topics that are popular among highly clustered and distributed users, respectively. A novel information diffusion model called *independent trend formation model* (ITFM) has also been introduced to distinguish viral diffusion of information from diffusion through external entities, such as news media, and to capture the diffusion of an arbitrary number of topics in a social network.

D. Sentiment analysis

With the rapid growth of Twitter messages, users are enabled to express their opinions in terms of views, sentiments, evaluations, attitudes, appraisals and emotions towards entities, events and their properties on almost anything. Opinions can be expressed by persons or by organizations.

In recent years, sentiment analysis (also known as opinion mining, sentiment detection or sentiment classification) has emerged as a new method to study user's opinions (or feelings) in regard to some topic. Twitter sentiment analysis focuses on analyzing tweets but it is difficult to extract opinions, read them, summarize them, and organize them into usable forms. Thus, automated Twitter sentiment analysis is needed. Sentiment analysis can be cast as a classification problem where the task is to classify messages into two categories depending on whether they convey positive or negative feelings [5].

A large number of tweets include opinions about products and services. It is interesting for companies interested in knowing how users feel about their products. Another stream of research focuses on the analysis of tweets as *electronic Word Of Mouth* (eWOM) in the area of product marketing. Word of mouth is the passing of information from person to person by oral communication.

Given its distinct communication characteristics, Twitter deserves serious attention as a form of online WOM (oWOM) or electronic WOM (eWOM). Tweets are underutilized as a source for evaluating customer sentiment. Sentiment analysis consists generally to classify an opinionated text as either positive or negative, according to the overall sentiment expressed by the author within it. A Tweet can contain polarity sentiments. For example, the word "kill" has a negative polarity, and the word "love" has a positive one.

Jansen et al. [19] consider a tweet as eWOM. They have found that 19% of a random sample of tweets contained mentions of a brand or product and that an automated classification was able to extract statistically significant differences of customer sentiment (i.e., the attitude of a writer towards a brand). Using sentiment detection, market researchers have a valuable tool to monitor how a product is accepted.

Twitter messages can also be used for political communication. It is important for political institutions to get a feel of prevalent sentiment and determine whether the public opinions are positive or negative.

The results of Tumasjan et al. [20] demonstrate that Twitter can also be considered as a valid real-time indicator of political sentiment. First, they examine Twitter messages. Their results indicate that people are finding interesting political information on Twitter which they share with their network of followers. They found that Twitter is indeed used as a platform for political deliberation. Second, they analyze the political sentiment of tweets and found that tweets reflect the current offline political sentiment in a meaningful way. To extract the sentiment of these tweets automatically, they use *Linguistic Inquiry and Word Count* (LIWC2007) [21], a text analysis software developed to assess emotional, cognitive, and structural components of text samples using a psychometrically validated internal dictionary. Third, after analyzing whether the activity on Twitter, they find that Twitter is as a predictor of the election result and even comes close to traditional election polls.

Twitter was also used to monitor the U.S. presidential debate in 2008 [22]. Tweets tended to favour Obama over McCain, and Obama really won the election afterwards. This shows that Twitter can also be used to predict political election results.

IV. COMPARISON AND DISCUSSION

In this section, we compare (Table 1) and discuss the approaches studied above. We begin by presenting a detailed description of the comparison criteria used:

a) Event challenge: is one of the fourth challenges of tweets event detection presented in section 3: Health epidemics identification, natural events detection, trending topics detection, or sentiment analysis.

b) Event type: describe the event type for an event challenge, specifically "what happens?". For example earthquakes, floods and volcanic eruptions are different event type for the event challenge natural events detection.

TABLE I. COMPARISON OF TWEETS EVENT DETECTION APPROACHES

Approach	Event challenge	Event type	Representation	Technique	Algorithm	Structural content	Textual content	Semantic
Culotta 2010	Health epidemics identification	Influenza	Bag-of-words	Linear regression	Supervised classifier	-	+	-
Paul and Dredze 2011		Diseases	SVM (*Support Vector Machine)*	Machine learning approach	Supervised	-	+	-
Rosoor et al. 2010	Natural events detection	Disasters	Salton representation Bag-of-words	Statistics approach	Supervised	Time	+	-
Sakaki et al. 2010		Earthquakes	SVM (*Support Vector Machine)*	Hybrid approach	Supervised	Time Retweets Location	+	+
Cheong and Cheong 2011		Floods	Undefined	Hybrid approach	Supervised	Username	+	-
Cataldi et al. 2010	Trending topics detection	Unspecified	Vectors of terms	Hybrid approach	Supervised Unsupervised	Time Username Retweets	+	+
Cheong and Lee 2009		Unspecified	Undefined	Statistics approach	Supervised	Username Time	+	-
Mathioudakis and Koudas 2010		Unspecified	Vectors of keyword	Statistics approach	Unsupervised	Time	+	-
Naaman et al. 2010		Unspecified	Undefined	Statistics approach	Supervised	Username	+	-
Benhardus 2010		Unspecified	Bag of words	Statistics approach	Unsupervised	Time	+	-
Budak et al. 2011		Unspecified	Undefined	Hybrid approach	Unsupervised	Username	+	+
Jansen et al. 2009	Sentiment analysis	Customer sentiment towards a brand	Undefined	Statistics approach	Unsupervised	-	+	+
Tumasjan et al. 2010		Political sentiment	Undefined	Linguistics approach	Unsupervised	Time Username Retweets	+	+

c) Representation: it is a transformation of a tweet in a format which is easier to understand. We present here the usual tweets representations.

d) Technique: four tweets mining techniques has been widely used for tweets:

- *Statistics Approach*: in these methods, the statistical information of the words can be used for tweets mining. Statistical methods include word frequency, TF*IDF, word co-occurrence, etc.
- *Linguistics Approach*: it uses the linguistic features of the words including the lexical analysis, syntactic analysis, discourse analysis and so on.
- *Machine Learning Approach*: it employs the extracted keywords from training tweets to learn a model and applies the model to find keywords from new tweets. This approach includes Naïve Bayes, Support Vector Machine, etc.
- *Hybrid approach*: that combines the techniques mentioned above.

e) Algorithm: the major algorithms used for tweets event detection, are subdivided into supervised, unsupervised and hybrid algorithms. For each approach, we mention the direction of the algorithm used.

f) Structural and Textual content: textual content of a tweet is 140 unstructured characters, the structural content reveals lots of metadata (e.g., time, re-tweet). Moreover, every user has a well-defined profile with personal information that represents the structural content (name, location, biographical sketch). When dealing with tweets, according to the prior information available on the collection, it may be relevant to consider textual content alone or to consider both structure and textual content.

g) Semantic: Semantics is the study of meaning in language [23]. In tweets, semantic treatment (lexical not grammatical) has for goal to study the semantic relationships

 ISBN: 978-1-61208-373-5

between words. Hence, the problem is how to distinguish between many different senses that a word may have (polysemy) or between different words that can have the same significance (synonymy)… The objective is to exploit the semantic similarity of terms composing the textual content of tweets. The semantic treatment can use external semantic resources like ontologies, thesauruses and taxonomies. In our comparison, we study if the existing approaches take into account this aspect or not.

Since the size of tweets is small, most traditional data mining algorithms are not adapted for tweets. Twitter messages contain little informational value but the aggregation of millions of tweets can generate important knowledge.

Several supervised classification algorithms have been proposed for specified events, including for instance support vector machines [8], [3]. Most techniques are unsupervised and rely on clustering [14], [16].

The majority of approaches of tweets event detection work on tweets content. Structural information is less used. Twitter profiles often contain semi-structured metadata (city, state, gender, age), enabling a more detailed statistical analysis. For example, we can associate geographic information with each tweet in order to perform a more fine automatic detection of natural events. For health epidemics identification, we can include temporal and geospatial dynamics to track diseases across a population. Public health information can also be correlated with user location.

The majority of approaches ignore semantic of information inside tweets. Synonymy and polysemy can cause difficulties (different label that describe the same concept or a label denoting different concepts).

The major problem consists in determining, the information and the knowledge to extract from tweets to serve in different fields. Societies wish to detect some information from tweets before even their apparition in the press agencies of news. For example, automatic detection of natural events from tweets consists in developing *Business Intelligence* tools to detect information before it appears in the agencies.

Various studies have been focused for trending topics detection which is another considerable challenge for tweets event detection based mainly on clustering or classification. Analysing tweets content in real-time can help specialists (news reporters, analysts…) to understand what is happening, what emergent trends are exchanged between people.

Trend detection is also important for online marketing professionals and opinion tracking companies, as trends point to topics that capture the public's attention. The requirement for real-time trend detection is only natural for a live stream where topics of discussion shift dynamically with time. Twitter content could become key applications in the attention economy. Given the ease of monitoring any brand's sentiment, one can view tweets as a competitive intelligence source.

We think that looking at Twitter data in real-time can help people to understand what is happening, what people are thinking about brands, organisations and products, and more importantly, how they feel about them.

Other works demonstrate that Twitter can also be considered as a valid real-time indicator of political sentiment. We note that mining public opinion from freely tweets could be a faster and less expensive alternative to traditional polls.

V. CONCLUSION AND FUTURE WORK

These last years have been marked by the emergence of microblogs. Their rates of activity reached some levels without precedent. Hundreds of millions of users are registered in these microblogs as Twitter. They exchange and tell their last thoughts, moods or activities by tweets in some words.

Tweets reveal useful event information for a variety of events. Approaches studied in this paper are interested in automatic extraction and detection of events from tweets. Although tweets are very exchanged on the web, we note that there are few works that are interested in tweets event detection, due in part to the fact that Twitter has only been in existence since 2006.

The major problem in this domain consists in determining, the event to extract from tweets to serve in different fields. For example societies wish to detect some information from tweets before even their apparition in the press agencies of news.

In terms of perspectives, we will try to take advantage of the methods studied in this paper, to propose a new approach for detecting tweets event in real-world. As a tweet is often associated with spatial and temporal information, we want to detect when and where an event happens.

REFERENCES

[1] Statistic brain, Twitter Statistics : Statistic Verification, http://www.statisticbrain.com/twitter-statistics/, 2014.

[2] M. Naaman, J. Boase, and C. H. Lai, "Is it really about me?: message content in social awareness streams", In CSCW '10: Proceedings of the 2010 ACM conference on Computer supported cooperative work, February 2010, pp. 189-192, Savannah, Georgia, USA.

[3] T. Sakaki, M. Okazaki, and Y. Matsuon, "Earthquake shakes Twitter users: real-time event detection by social sensors", Proceedings of the 19th international conference on World wide web (WWW), April 2010, pp. 851-860, New York, NY, USA.

[4] Semiocast study, http://semiocast.com/fr/publications/2012_07_30 _Twitter_ reaches_half_a_billion_accounts_140m_in_the_US, 2012.

[5] A. Bifet and E. Frank, "Sentiment knowledge discovery in Twitter streaming data", In Proc 13th International Conference on Discovery Science, October 2010, pp. 1-15, Springer, Canberra, Australia.

[6] G. Eysenbach, "Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search", Communication and Publication Behavior on the Internet . J Med Internet Res, Vol. 11(1):e11, 2009.

[7] A. Culotta, "Towards detecting influenza epidemics by analyzing twitter messages", In KDD Workshop on Social Media Analytics, July 2010, pp. 115–122.

[8] J.M. Paul and M. Dredze, "A Model for Mining Public Health Topics from Twitter", Technical Report. Johns Hopkins University. 2011.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan", Latent dirichlet allocation", Journal of Machine Learning Research, Vol. 3, January 2003, pp. 993-1022.

[10] B. Rosoor, L. Sebag, S. Bringay, P. Poncelet, and M. Roche, "When a tweet detect a natural event... ", Actes du colloque Veille Stratégique Scientifique et Technologique VSST'2010, Septembre 2010, Toulouse (France).

[11] G. Salton and C. Buckley,"Term-weighting approaches in automatic text retrieval", Inf. Process. Manage., Vol. 24(5), January 1988, pp. 513-523.

[12] T. Joachims, "Text categorization with support vector machines", In Proc. ECML'98, April 1998, pp. 137-142.

[13] F. Cheong and C. Cheong, "Social media data mining: A social network analysis of tweets during the 2010-2011 australian floods", In PACIS, July 2011, pp. 1-16.

 ISBN: 978-1-61208-373-5

[14] M. Mathioudakis and N. Koudas, "TWITTERMONITOR: trend detection over the twitter Stream", Proceedings of SIGMOD Conference, June 2010, pp.1155-1158.

[15] M. Cheong and V. Lee, "Integrating web-based intelligence retrievaland decision-making from the twitter trends knowledge base", InSWSM '09: Proceeding of the 2nd ACM workshop on Social web search and mining, November 2009, pp. 1-8, New York, NY, USA.

[16] J. Benhardus, "Streaming Trend Detection in Twitter", 2010 UCCS Reu for Artificial Intelligence, Natural Language Processing and Information Retrieval, Final report 1, 2010.

[17] M. Cataldi, L. D. Caro, and C. Schifanella, "Emerging Topic Detection on Twitter based on Temporal and Social Terms Evaluation", MDMKDD'10, July 2010, pp. 4-13.

[18] C. Budak, D. Agrawal, and A. El Abbadi, "Structural trend analysis for online social networks". Technical Report UCSB/CS-2011-04, UCSB, 2011.

[19] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth", Journal of the American Society for Information Science and Technology, Vol. 60, November 2009, pp. 1-20.

[20] A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment", In Proceedings of the 4th International Conference on Weblogs and Social Media, May 2010, pp. 178-471.

[21] J. Pennebaker, C. Chung, and M. Ireland, "The development and psychometric properties of LIWC2007", Austin, TX, 2007.

[22] N.A. Diakopoulos and D.A. Shamma, "Characterizing debate performance via aggregated twitter sentiment". CHI 2010 Proceedings of the SIGCHI Conference on Human Factors in Computing System, ACM, Atlanta Georgia, April 2010, pp. 1195-1198.

[23] J. R. Hurford, "Semantics : a coursebook", Cambridge University Press, 1983.

 ISBN: 978-1-61208-373-5

Application of the Composite Field in the Design of an Improved AES S-box Based on Inversion

Zhao Wang, Xiao Zhang, Sitao Wang, Zhisong Hao and Zhiming Zheng
School of Mathematics and Systems Science & LMIB
Beijing University of Aeronautics and Astronautics
Beijing, China
Email: wangzhao@smss.buaa.edu.cn, xiao.zh@buaa.edu.cn, wang_sitao@smss.buaa.edu.cn, haozhisong2004@sina.com, zzheng@pku.edu.cn.

***Abstract*—The hardware implementation of the Substitution-Box (S-box) of the Advanced Encryption Standard (AES) always employs composite field $GF((2^n)^2)$ to obtain better efficiency. In this paper, an improved class of S-boxes by direct inversion in composite field is presented, and the choice of the subfield leading to the most efficient implementation is discussed. Eliminating the field isomorphic transformations, such a composite field is easier to fix and the resulting hardware implementation is more efficient than that of AES S-box. Some common cryptographic characteristics for the composite field based S-boxes are examined, and it turns out that direct inversion in composite field does not weaken the cryptographic characteristics. In addition, a demonstration for the immunity against the potential algebraic attack on AES with the replacement of our S-box is given, and it is proven that the revised AES is even more secure than the original AES against the algebraic attack. As a result of this work, it could be predicted that the isomorphism implies equal immunity from certain cryptanalysis. Our S-box is suitable for the area-limited hardware production.**

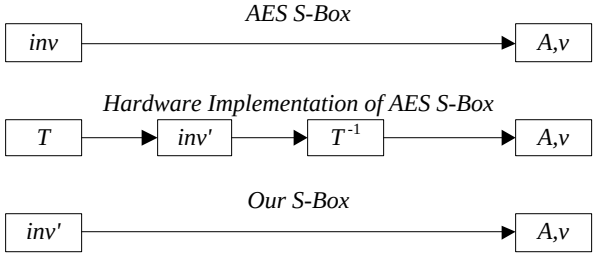
Keywords–AES; Composite field; S-box; Hardware implementation.

I. INTRODUCTION

The Substitution-Box (S-box) is a basic component of symmetric key algorithms, and should always be carefully chosen to create strong confusion and to resist certain kinds of cryptanalysis. The multiplicative inversion mapping over Galois Field are frequently employed due to their ideal cryptographic characteristics [1]. Most of the recent S-boxes in block ciphers, such as the Advanced Encryption Standard (AES) [2], Camellia (NESSIE and CRYPTREC winner) [3], CLEFIA (developed by SONY) [4] and SMS4 (used in the Chinese National Standard for Wireless LAN WAPI) [5] are created based on the inversion on $GF(2^8)$. Currently, the GF-inversion has become one of the most popular algebraic tools in block ciphers, and its hardware implementation, especially targeted for AES is still a worldwide challenge.

Among so many state-of-art designs to implement GF-inversion, one general idea is to employ the composite field representation [6]. The fields $GF(2^8)$, $GF((2^4)^2)$ and $GF(((2^2)^2)^2)$ are linear isomorphic to each other, so that the isomorphisms can be achieved by simple matrix multiplications, which means that finding the inverse in $GF(2^8)$ can be changed into calculating the low-cost addition, multiplication, square and inversion in $GF(2^4)$ [7] or $GF((2^2)^2)$ [8].

Mentens et al. [9] used $GF(((2^2)^2)^2)$, examined all possible choices for irreducible polynomials generating the extension field and all the transformation matrices mapping to the corresponding $GF(((2^2)^2)^2)$ representation, and pointed out the area of [8] was still 5% from the minimum. Later, Canright [10] considered not only polynomial bases but also normal bases, with 432 cases including all in [9], and get the most compact S-box up to date. However, the critical path delay, substructure sharing and use of NOR and NOT gates were all ignored, and then Zhang and Parhi [11] improved it and got a seemingly better result. Nikova and Rijmen [12] decomposed $GF(2^8)$ to $GF((2^4)^2)$ using normal bases and the result could compete with that in [10]. For the other recent architectures, Nogamni et al. [13] suggest mixing normal and polynomial bases for the reduction of the critical path delay of S-box.



inv The inversion on $GF(2^8)$
inv' The inversion on $GF((2^4)^2)$ or $GF(((2^2)^2)^2)$
A,v The affine transformAtion of AES
T The field conversion from $GF(2^8)$ to $GF((2^4)^2)$ or $GF(((2^2)^2)^2)$
T^{-1} The inverse transformation of *T*

Figure 1. Structures of the different S-boxes in this paper.

In this paper, we try to do the inversion directly in composite field; see Figure 1. Then, the effect caused by transformation matrixes can be ignored, and the choice of irreducible polynomial for field extension would be easier.

The outline of this paper is as follows. In Section II, the applications of composite field in S-box implementation is introduced. In Section III, our new S-box based on inversion in composite field is described. The complete cryptographical analysis of this new S-box and some potential algebraic attacks are given in Section IV and Section V, followed by the conclusion.

II. PRELIMINARIES

In this section, we show the applications of the composite field for the hardware implementation of AES S-box.

 ISBN: 978-1-61208-373-5

The AES S-box involves an inversion mapping in $GF(2^8)$ followed by a $GF(2)$-affine transformation, here $GF(2^8) = GF(2)/(O(z))$, $O(z) = z^8+z^4+z^3+z+1$. Denote the AES S-box by $S_1(x)$,

$$S_1(x) = \boldsymbol{A} \cdot x^{-1} + \boldsymbol{v} = \boldsymbol{A} \cdot inv(x) + \boldsymbol{v}, \quad (1)$$

where $inv(x)$ denotes $GF(2^8)$ inversion of x. Denote the binary representation of any $x \in GF(2^8)$ by $(x_7, x_6, \cdots, x_0)$ with x_7 the most significant bit. And

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}. \quad (2)$$

The basic idea of employing composite field comes from the field isomorphism $GF(2^n) \cong GF((2^{n/2})^2) \cong GF(((2^{n/4})^2)^2)$. However, all the basic arithmetics in $GF((2^{n/2})^2)$ are actually the ones in $GF(2^{n/2})$. Each element $\Delta \in GF((2^m)^2) = GF(2^m)/(x^2 + \alpha x + \beta)$ where $\alpha, \beta \in GF(2^m)$, can be expressed as $\Delta = \delta_1 x + \delta_0$, with $\delta_1, \delta_0 \in GF(2^m)$. The multiplicative inversion of Δ can be obtained via

$$\Delta^{-1} = (\delta_1\Gamma)x + (\delta_0 + \delta_1\alpha)\Gamma, \Gamma = (\delta_1^2\beta + \delta_1\delta_0\alpha + \delta_0^2)^{-1}. \quad (3)$$

$GF(2^8)$ can be mapped to either $GF((2^4)^2)$ or $GF(((2^2)^2)^2)$. The specific operation comprises of the following 3 steps:

(I) Field linear isomorphic transformation T maps each $GF(2^8)$ element into the composite field $GF((2^4)^2)$ (or $GF(((2^2)^2)^2)$).
(II) Inversion in composite field.
(III) Linear transformation T^{-1}.

The operations using $GF((2^4)^2)$ representation can be expressed as

$$S_1(x) = \boldsymbol{A} \cdot \boldsymbol{T}_1^{-1} \cdot inv'(\boldsymbol{T}_1 \cdot x) + \boldsymbol{v}, \quad (4)$$

where $inv'(x)$ denotes $GF((2^4)^2)$ inversion of x. In [8], there is another architecture by mapping $GF(2^8)$ to $GF(((2^2)^2)^2)$, and the operations can be expressed as

$$S_1(x) = \boldsymbol{A} \cdot \boldsymbol{T}_2^{-1} \cdot inv''(\boldsymbol{T}_2 \cdot x) + \boldsymbol{v}, \quad (5)$$

where $inv''(x)$ denotes $GF(((2^2)^2)^2)$ inversion of x.

III. Improved S-boxes by inversion in Composite Field

A. The Specification of the Improved S-boxes

According to the designers of AES [2], the choice of $GF(2^8)$ inversion only comprises cryptographical reasons but barely covers implementation efficiency. It provides very ideal input-output correlation amplitude and difference propagation probability, and the following $GF(2)$-affine transformation (A, v) complicates the algebraic expression of $S_1(x)$. However, one may find it difficult to build the optimum hardware architecture for $S_1(x)$ on composite field. We believe that three reasons might lead to such a situation:

(I) *Computational Complexity*. As revealed by the Extended Euclid's Algorithm [14], GF-inversion is essentiality more complex than any other basic GF arithmetic such as multiplication.

(II) *Overfull Factors*. Introducing the composite field does reduce the area cost, however, additional factors arise, such as the coefficients of the irreducible polynomial generating the composite field. The following factors must be considered:
   a) Subfield multiplication of two multiplicands;
   b) Subfield squaring;
   c) Subfield constant multiplication;
   d) Matrixes of $\boldsymbol{T}$ and $\boldsymbol{A} \cdot \boldsymbol{T}^{-1}$.

(III) *How to Define the "optimum"?* The criteria to build the optimum architecture of the previous contributions [7]–[13] are unambiguous, which can not be achieved simultaneously. For example, the throughput usually contradicts the area since compact construction always causes more critical path delays, and certain criterion is hard to be judged or quantified. As shown by Mentens et al. [9], further area reduction can be achieved by substructure sharing or introducing NOR gates or NOT gates, so the matrix for $\boldsymbol{T}$ with the least number of "1" might not be the best choice.

Recognizing the comparability among (1), (4) and (5), we could try to overlook the field isomorphism T_i and T_i^{-1} and directly carry out the multiplicative inversion in the isomorphic composite field. So, two new S-boxes are obtained in (6) and (7), and only the irreducible polynomials are to be fixed.

$$S_2(x) = \boldsymbol{A} \cdot inv'(x) + \boldsymbol{v}, x \in GF((2^4)^2)$$
$$\begin{cases} GF((2^4)^2) &= GF(2^4)/(M(x)) \\ M(x) &= x^2 + m_0 x + m \\ GF(2^4) &= GF(2)/(N(y)) \\ N(y) &= y^4 + n_2 y^3 + n_1 y^2 + n_0 y + n \end{cases} \quad (6)$$

$$S_3(x) = \boldsymbol{A} \cdot inv''(x) + \boldsymbol{v}, x \in GF(((2^2)^2)^2)$$
$$\begin{cases} GF(((2^2)^2)^2) &= GF((2^2)^2)/(P(x)) \\ P(x) &= x^2 + p_0 x + p \\ GF((2^2)^2) &= GF(2^2)/(Q(y)) \\ Q(y) &= y^2 + q_0 y + q \\ GF(2^2) &= GF(2)/(R(z)) \\ R(z) &= z^2 + r_0 z + r \end{cases} \quad (7)$$

First of all, fix $R(z) = z^2 + z + 1$ since it is the only irreducible polynomial over $GF(2)$ with degree 2. Then, set $m_0 = 1 \in GF(2^4)$ in (8) and $p_0 = 1 \in GF((2^2)^2)$, $q_0 = 1 \in GF(2^2)$ in (9) since they would reduce more multiplications than m and p, q, similar to the AES settings in Section II.

For $S_2(x)$, there are only three choices for $N(y)$, and for $M(x)$, only m is to be decided. When $N(y)$ is fixed, there are only limited candidate values for m to make $M(x)$ irreducible. The best value for m can be found through comparing the gate counts of multiplying m. For $S_3(x)$, since $R(z)$, p_0 and q_0 are fixed, we find there are only eight choices of (p,q) to make $P(x)$ and $Q(y)$ irreducible, so the best (p,q) can also be fixed through simple comparison. Consequently, all the settings can

 ISBN: 978-1-61208-373-5

TABLE I. NUMBER OF XOR GATE FOR T AND $A \cdot T_1^{-1}$

Design	Field	Original	After Optimizations
[7]	$GF((2^4)^2)$	43	Not available
[8]	$GF(((2^2)^2)^2)$	45	Obtained by greedy algorithm
[9]	$GF(((2^2)^2)^2)$	38	No optimization
[11]	$GF(((2^2)^2)^2)$	36	28 (Gate) + 6 (Critical Path)
[11]	$GF(((2^2)^2)^2)$	38	27 (Gate) + 6 (Critical Path)
S_2	$GF((2^4)^2)$	**32**	**18** (Gate) + **4** (Critical Path)
S_3	$GF(((2^2)^2)^2)$	**32**	**18** (Gate) + **4** (Critical Path)

be fixed:

$$S_2(x) = \boldsymbol{A} \cdot inv'(x) + \boldsymbol{v}, x \in GF((2^4)^2)$$
$$\begin{cases} GF((2^4)^2) & = & GF(2^4)/(M(x)) \\ M(x) & = & x^2 + x + \{1001\}_2 \\ GF(2^4) & = & GF(2)/(N(y)) \\ N(y) & = & y^4 + y + 1 \end{cases} \quad (8)$$

$$S_3(x) = \boldsymbol{A} \cdot inv''(x) + \boldsymbol{v}, x \in GF(((2^2)^2)^2)$$
$$\begin{cases} GF(((2^2)^2)^2) & = & GF((2^2)^2)/(P(x)) \\ P(x) & = & x^2 + x + \{1100\}_2 \\ GF((2^2)^2) & = & GF(2^2)/(Q(y)) \\ Q(y) & = & y^2 + y + \{10\}_2 \\ GF(2^2) & = & GF(2)/(R(z)) \\ R(z) & = & z^2 + z + 1 \end{cases} \quad (9)$$

B. Performance of Hardware Implementation

Defined in composite filed, the inversion function in both S_2 and S_3 could be implemented directly by (3) without field isomorphism.

Observing the differences between (4) and (8), or (5) and (9), our improvement on efficiency is manifest. Most of the previous architectures for S_1 based on composite field need to do at least two matrix multiplications, $\boldsymbol{T}$ and $\boldsymbol{A} \cdot \boldsymbol{T}_1^{-1}$, while our S-boxes only need to multiply the matrix $\boldsymbol{A}$. The number of XOR gate for the two matrix multiplications of the previous architectures are listed in Table I. Note that [10] [12] [13] used different bases, while [11] provided us two optimal settings.)

The total number of XOR gate within the multiplication of $\boldsymbol{A}$ is only 32. Our S-box, S_2 or S_3 has only one regular-structured matrix $\boldsymbol{A}$, while the hardware implementation of S_1 has to deal with two irregular structured matrixes, $\boldsymbol{T}$ and $\boldsymbol{A} \cdot \boldsymbol{T}_1^{-1}$. Further optimization would reach less gate counts and critical path. If $\boldsymbol{y} = \boldsymbol{A} \cdot \boldsymbol{x}$ where $\boldsymbol{A}$ is defined in (2), then

$$\begin{cases} y_7 = x_5 + X_2 + X_1 \\ y_6 = x_6 + X_6 \\ y_5 = x_1 + X_6 \\ y_4 = X_4 + X_3 \\ y_3 = x_7 + x_3 + X_5 \\ y_2 = X_5 + X_2 \\ y_1 = X_7 + X_3 \\ y_0 = x_4 + x_0 + X_7 \end{cases} with \begin{cases} X_7 = x_5 + X_2 \\ X_6 = x_5 + X_4 \\ X_5 = X_3 + x_2 \\ X_4 = X_1 + x_2 \\ X_3 = x_1 + x_0 \\ X_2 = x_7 + x_6 \\ X_1 = x_4 + x_3 \end{cases} . \quad (10)$$

Apparently, there are totally 18 XORs in (10), and 4 XORs in the critical path. From Table I, we can see that our S-boxes have a great advantage over the known results.

The reduction in our S-box also optimises the counter-measure against side-channel attack [15], such as differential power analysis [16], which uses statistical analysis of physical quantities to deduce certain information about the secret key. S_1 can be effectively masked under composite field, so do S_2 and S_3. The only difference is that we only need to mask A for S_2 and S_3, which is clearly more effective. Also for S_3, the inversion on $GF(((2^2)^2)^2)$ can be split into that on $GF((2^2)^2)$, and can be split into $GF(2^2)$, where $x^{-1} = x^2$, and inversion becomes linear. Hence, S_3 might be easier to be masked.

Our S-boxes are suitable for the encryption(decryption) within the area-limited hardware productions, such as flash memory cards, smart cards and mobile phones. Furthermore, the idea to employ the composite filed to construct the S-box is highly recommended in the design of the lightweight block cipher [17].

TABLE II. TEST RESULTS ON ANF OF S_i

Terms	f_7	f_6	f_5	f_4	f_3	f_2	f_1	f_0	Sum
S_1	110	112	114	131	136	145	133	132	1013
S_2	130	119	131	117	126	132	132	129	1016
S_3	119	114	132	126	126	126	128	134	1005

IV. Cryptographic Characteristics of S_2 and S_3

In this Section, a security evaluation of S_2 and S_3 will be given by comparing some common cryptographic characteristics with those of S_1. Denote $S(x) = (f_{m-1}(x_{n-1}, \cdots, x_0), ..., f_0(x_{n-1}, \cdots, x_0)) : GF(2^n) \mapsto GF(2^m)$ as the S-box transformation, with $f_i(x), m-1 \geq i \geq 0$ the n-tuple Boolean function of the i_{th} output bit.

A. Non-Linearity, Differential Distribution, Algebraic Degree, and Algebraic Complexity

By simple calculations, the Non-Linearity (NL) [18], the differential distributions [18], and the algebraic degree [18] of both S_2 and S_3 stay the same as S_1, and they show almost the same number of terms in their algebraic normal form (ANF) [18]; see Table II. In terms of algebraic complexity, since the structure of both S_2 and S_3 are entirely the same as S_1, restricting the polynomial in each own field makes more sense. It has been proven that every S-box with the form $\boldsymbol{A} \cdot x^{-1} + \boldsymbol{v}$ has only 9 terms in its polynomial expression, so does the $GF((2^4)^2)$ polynomial of S_2 and $GF(((2^2)^2)^2)$ polynomial of S_3, which equally show the ability against the interpolation attack [19].

B. Algebraic Immunity

Algebraic Immunity comes from the algebraic attack [20]. For an $n \times n$ S-box, it is defined as $\Gamma = ((t-r)/n)^{\lceil t/r \rceil}$, where r denotes the total number of linear independent equations, and t denotes the number of monomials appearing in the equations, including the constant terms.

For S_1 in (1), where $b = inv(a),\ a, b \in GF(2^8)$, one can find $r = 24$ bi-affine equations between a_i and b_j. The first set of eight equations comes from simplifying the following polynomial in the bases $\{1, z, z^2, \cdots, z^7\}$:

$$\left(\sum_{i=0}^{7} a_i z^i\right) \cdot \left(\sum_{i=0}^{7} b_i z^i\right) \ mod\ O(z) = 1. \quad (11)$$

The second set of eight equations is derived from simplifying any one equation from the group of the following $GF(2^8)$ equations:

$$a = a^2 \cdot b, a^2 = a^4 \cdot b^2, \cdots, a^{128} = a \cdot b^{128}. \quad (12)$$

Since these eight $GF(2^8)$-equations in (12) are linearly equivalent with each other, every two different $GF(2^8)$-equations from (12) will generate two different but linearly dependent sets of 8 $GF(2)$-equations between $\{a_i\}$ and $\{b_j\}$. The remaining 8 equations comes from the symmetry with respect to the exchange of a and b [20]. Adding the affine relationship $\boldsymbol{c} = \boldsymbol{A} \cdot \boldsymbol{b} + \boldsymbol{v}$, all the $\{b_j\}$ can be replaced by $\{c_k\}$, and then totally 24 bi-affine equations between $\{a_i\}$ and $\{c_k\}$ are obtained. The monomials of the system are: $\{1, a_0, a_1, \cdots, a_7, c_0, c_1, \cdots, c_7, a_0c_0, a_0c_1, \cdots, a_7c_7\}$, therefore $t = 1 + 8 + 8 + 8 \times 8 = 81$.

For S_2, $b = inv'(a)$, we have

$$\Big[\Big(\sum_{i=4}^{7} a_i y^{i-4}\Big)x + \sum_{i=0}^{3} a_i y^i\Big] \cdot \Big[\Big(\sum_{i=4}^{7} b_i y^{i-4}\Big)x + \sum_{i=0}^{3} b_i y^i\Big] \equiv 1 \ mod \ M(x) \ mod \ N(y) \tag{13}$$

Unlike (11), the bases for (13) are $\{1, y, y^2, y^3, x, yx, y^2x, y^3x\}$, still $n = 8$ equations are get. In the same way, one can get another eight equations from any one of the group (12) defined in $GF((2^4)^2)$, and eight more by exchanging a and b. Our simulation proved that these $r = 24$ equations are linearly independent, and if adding eight more from expanding any one equation from the $GF((2^4)^2)$ group (12) (or exchanging a and b), does not change the rank of the system. Our simulation shows that for S_2, $r = 24$. Similarly, t stays unchanged. Therefore, the Algebraic Immunity of S_2 is the same as S_1. For S_3, the result is also the same.

C. Influence on AES

Considering the coherence for the calculational field for the sake of the analysis of the algebraic attack, we suggest all the computation in AES being defined in the same field according to the chosen S-box, which means that the matrix multiplication for the MixColumn operation will be done in $GF((2^4)^2)$ if S_2 is used or in $GF(((2^2)^2)^2)$ if S_3 is used. Denote the AES with S_1 replaced by S_2 and $GF((2^4)^2)$ matrix multiplication for MixColumn by AES$_{cf}$, and similarly with S_3 and $GF(((2^2)^2)^2)$ matrix multiplication by AES$_{tf}$.

By then, we can conclude that both AES$_{cf}$ and AES$_{tf}$ are immune against linear attack, differential attack and interpolation attack. While for the other attacks not related to S-boxes, such as the square attack [21], the collisions attack [22] and related-key attack [23], both AES$_{cf}$ and AES$_{tf}$ have the same immunity as AES.

V. Algebraic Attack on AES$_{cf}$ and AES$_{tf}$

Our improvement on S_2 or S_3 is simply the change of field. In order to deeper demonstrate the advantage of the composite field, a concrete algebraic attack on AES$_{cf}$ and AES$_{tf}$ will be given. There are two ways to develop the algebraic attack, one is put forward by N. T. Courtois and J. Pieprzyk in Asia Crypt 2002 based on a $GF(2)$-system [20], and the other is found by S. Murphy and M. J. Robshaw in CRYPTO 2002 with only simple algebraic operations in $GF(2^8)$ [24]. The $GF(2^8)$-system created in [24] is less complicated than the $GF(2)$-system derived in [20], which indicates that any change in the field evolved during the encryption should be considered in algebraic attack, that is why three case are discussed below.

A. The $GF(2)$-Algebraic Attack

Firstly, put AES$_{cf}$ in the $GF(2)$-system. From Section IV-B, the $GF(2)$-system of S_2 is very similar to that of S_1, for they have the same algebraic lmmunity. For AddRoundkey and ShiftRows, their $GF(2)$-system is apparently equal in scale. It is easily seen that in $GF(2^8)$ or $GF((2^4)^2)$, constant multiplication can be represented by a 8-order $GF(2)$-matrix-vector multiplication. The operation in MixColumn is equivalent to a 32-order $GF(2)$-matrix-vector multiplication. In [20], it was proved that the $GF(2)$-system of AES is too complicated to be solved, therefore, AES$_{cf}$ is safe from the $GF(2)$ algebraic attack.

B. The $GF((2^4)^2)$-Algebraic Attack

In [24], AES is embedded within the Big Encryption System (BES) which uses algebraic operations in $GF(2^8)$ and can be described by a system of multivariate quadratic equations in $GF(2^8)$ simpler than the $GF(2)$-system in [20]. Analogously, we could embed AES$_{cf}$ within a BES-like cipher, and get a $GF((2^4)^2)$-system of multivariate quadratic equations. Even though this is better than the $GF(2)$-system of AES$_{cf}$, the solvability remains the same as that of the $GF(2^8)$ system of BES because these two systems are equal in scale.

C. The $GF(2^4)$-Algebraic Attack

To be more accurate, the arithmetic within AES$_{cf}$ is in $GF(2^4)$ rather than $GF((2^4)^2)$. So, we may try to split the round function of AES$_{cf}$ in $GF(2^4)$. (Most of the notations from [24] will be used below with the same indication.)

First of all, embed AES$_{cf}$ within another BES-like cipher called BES$_{cf}$. Define a mapping ϕ from $GF(2^4)$ to $(GF(2^4))^4$, $\phi(a) = (a^{2^0}, a^{2^1}, a^{2^2}, a^{2^3})$, and regard $\boldsymbol{a} \in (GF((2^4)^2))^{16}$ (the state variable of AES$_{cf}$) as a column vector, where

$$\boldsymbol{a} = (a_{00}, \cdots, a_{30}, a_{01}, \cdots, a_{31}, \cdots, a_{33})^T. \tag{14}$$

Each $a_{ij} \in GF((2^4)^2)$ can be split as $a_{ij} = a_{ij1}x + a_{ij0}$, so that the state space of AES$_{cf}$ is actually $(GF(2^4))^{32}$. The function ϕ can be extended to the state space of AES$_{cf}$: $\phi(\boldsymbol{a}) = (\phi(a_{001}), \phi(a_{000}), \cdots, \phi(a_{331}), \phi(a_{330}))$, and the state space of BES$_{cf}$ is $(GF(2^4))^{128}$. Using $\phi(a_{ijm}) = (b_{ijm0}, \cdots, b_{ijm3}), i, j = 0, \cdots, 3, m = 1, 0$, every BES$_{cf}$ state vector $\boldsymbol{b}$ can be denoted as

$$\boldsymbol{b} = (b_{0010}, \cdots, b_{0013}, b_{0000}, \cdots, b_{0003}, \cdots, b_{3310}, \cdots, b_{3313}, b_{3300}, \cdots, b_{3303})^T. \tag{15}$$

Our aim is to give every operation of AES$_{cf}$ a $GF(2^4)$ expression, and extend it to BES$_{cf}$ using the way of [24].

As noted in [24], the additive constant 0X63 in S_1 (noted by $\boldsymbol{v}$ in (1)) can be removed by incorporating it within a modified key schedule, and so does S_2 for AES$_{cf}$. Our reductions are as follows.

1) Subkey addition: This is the same as BES, just a bitwise XOR operation on $GF(2^4)^{128}$.

2) S-box inversion: Since the inversion of S_2 is defined on $GF((2^4)^2)$, it will act on each pair of $(a_{ij1}, a_{ij0}), i, j = 0, \cdots, 3$. From (3), the $GF(2^4)$ expression is get

$$\begin{cases} inv'(a_{ij1}, a_{ij0}) = (a_{ij1} \cdot t^{-1}, (a_{ij1} + a_{ij0}) \cdot t^{-1}), \\ t = a_{ij1}^2 \cdot \lambda + a_{ij1} \cdot a_{ij0} + a_{ij0}^2. \end{cases}$$

For every eight consecutive elements $(b_{ij10}, \cdots, b_{ij03})$ from $\boldsymbol{b}$, the S-box inversion of BES$_{cf}$ can be expressed as:

$$\begin{cases} t_m = b_{ij1(m+1)} \cdot \lambda^{2^m} + b_{ij1m} \cdot b_{ij0m} + b_{ij0(m+1)}, \\ b_{ij1m} \mapsto t_m^{-1} \cdot b_{ij1m}, \\ b_{ij0m} \mapsto t_m^{-1} \cdot (b_{ij1m} + b_{ij0m}). \end{cases}$$

Here, $m = 0, \cdots, 3$, and $m+1$ is interpreted modulo 4.

3) S-box linear operation: Use Lagrange interpolation in $GF((2^4)^2)$, one can get

$$\begin{aligned} l(a) =' & \ 06'a +' 4B'a^2 +' F6'a^4 +' 89'a^8 \\ & +' 46'a^{16} +' C0'a^{32} +' 8F'a^{64} +' 24'a^{128}, \end{aligned} \tag{16}$$

where $l(a)$ denotes the $GF(2)$ matrix multiplication of S_2. Furthermore, set $a = a_1 x + a_0$ and by only simple calculations on $GF(2^4)$, (16) can be converted from $GF((2^4)^2)$ to $GF(2^4)$:

$$l(a_1 x + a_0) = \Big[\sum_{i=0}^{3} (l_i a_1^{2^i} + l_{i+4} a_0^{2^i}) \Big] x + \sum_{i=0}^{3} (l_{i+4} a_1^{2^i} + l_i a_0^{2^i}),$$

where $(l_0, \cdots, l_7) = ('0,' B,' 9,' D,' 4,' 8,' 7,' A')$. Also the following vector form is used,

$$\begin{aligned} a_1 &\mapsto (l_0, \cdots, l_7) \cdot \tilde{\boldsymbol{a}}, \\ a_0 &\mapsto (l_4, \cdots, l_7, l_0, \cdots, l_3) \cdot \tilde{\boldsymbol{a}}, \\ \tilde{\boldsymbol{a}} &= (a_1, a_1^2, a_1^4, a_1^8, a_0, a_0^2, a_0^4, a_0^8)^T. \end{aligned}$$

The extension to BES$_{cf}$ requires a 128×128 $GF(2^4)$ matrix $\boldsymbol{Lin_B}$, a block diagonal matrix with 16 identical blocks, that is, $\boldsymbol{Lin_B} = Diag_{16}(\boldsymbol{L_B})$, where $\boldsymbol{L_B} = \begin{pmatrix} \boldsymbol{L_{B1}} & \boldsymbol{L_{B2}} \\ \boldsymbol{L_{B2}} & \boldsymbol{L_{B1}} \end{pmatrix}$ and

$$\boldsymbol{L_{B1}} = \begin{pmatrix} l_0^{2^0} & l_1^{2^0} & l_2^{2^0} & l_3^{2^0} \\ l_1^{2^1} & l_2^{2^1} & l_3^{2^1} & l_0^{2^1} \\ l_2^{2^2} & l_3^{2^2} & l_0^{2^2} & l_1^{2^2} \\ l_3^{2^3} & l_0^{2^3} & l_1^{2^3} & l_2^{2^3} \end{pmatrix}, \boldsymbol{L_{B2}} = \begin{pmatrix} l_4^{2^0} & l_5^{2^0} & l_6^{2^0} & l_7^{2^0} \\ l_5^{2^1} & l_6^{2^1} & l_7^{2^1} & l_4^{2^1} \\ l_6^{2^2} & l_7^{2^2} & l_4^{2^2} & l_5^{2^2} \\ l_7^{2^3} & l_4^{2^3} & l_5^{2^3} & l_6^{2^3} \end{pmatrix}.$$

4) ShiftRows: This can be represented as a 128×128 $GF(2^4)$ matrix $\boldsymbol{R_B}$ when we only need to ensure every two vector conjugates (8 elements) are moved as a single entity.

5) MixColumn: We have assumed this operation on $GF((2^4)^2)$. For AES$_{cf}$ it can be represented as a 8×8 $GF(2^4)$ matrix $\boldsymbol{C_A}$, that is

$$\begin{pmatrix} a_{0i1} \\ a_{0i0} \\ a_{1i0} \\ a_{1i1} \\ a_{2i0} \\ a_{2i1} \\ a_{3i0} \\ a_{3i1} \end{pmatrix} \mapsto \begin{pmatrix} y & 0 & y+1 & 0 & 1 & 0 & 1 & 0 \\ 0 & y & 0 & y+1 & 0 & 1 & 0 & 1 \\ 1 & 0 & y & 0 & y+1 & 0 & 1 & 0 \\ 0 & 1 & 0 & y & 0 & y+1 & 0 & 1 \\ 1 & 0 & 1 & 0 & y & 0 & y+1 & 0 \\ 0 & 1 & 0 & 1 & 0 & y & 0 & y+1 \\ y+1 & 0 & 1 & 0 & 1 & 0 & y & 0 \\ 0 & y+1 & 0 & 1 & 0 & 1 & 0 & y \end{pmatrix} \cdot \begin{pmatrix} a_{0i1} \\ a_{0i0} \\ a_{1i0} \\ a_{1i1} \\ a_{2i0} \\ a_{2i1} \\ a_{3i0} \\ a_{3i1} \end{pmatrix}.$$

Here, y is a root of $N(y)$ that defines $GF(2^4)$ in (8). To maintain the conjugacy for extension to BES$_{cf}$, four matrixes are needed: $p = 0, \cdots, 3, \boldsymbol{C_B^{(p)}} = \begin{pmatrix} y^{2^p} & 0 & (y+1)^{2^p} & 0 & 1 & 0 & 1 & 0 \\ 0 & y^{2^p} & 0 & (y+1)^{2^p} & 0 & 1 & 0 & 1 \\ 1 & 0 & y^{2^p} & 0 & (y+1)^{2^p} & 0 & 1 & 0 \\ 0 & 1 & 0 & y^{2^p} & 0 & (y+1)^{2^p} & 0 & 1 \\ 1 & 0 & 1 & 0 & y^{2^p} & 0 & (y+1)^{2^p} & 0 \\ 0 & 1 & 0 & 1 & 0 & y^{2^p} & 0 & (y+1)^{2^p} \\ (y+1)^{2^p} & 0 & 1 & 0 & 1 & 0 & y^{2^p} & 0 \\ 0 & (y+1)^{2^p} & 0 & 1 & 0 & 1 & 0 & y^{2^p} \end{pmatrix}$. and $\boldsymbol{C_B^{(0)}} = \boldsymbol{C_A}$ and if $(b_0, \cdots, b_7)^T = \boldsymbol{C_A} \cdot (a_0, \cdots, a_7)^T$, then $(b_0^{2^p}, \cdots, b_7^{2^p})^T = \boldsymbol{C_B^{(p)}} \cdot (a_0^{2^p}, \cdots, a_7^{2^p})^T$. The whole operation can be represented as a 128×128 $GF(2^4)$ matrix $\boldsymbol{Mix_B}$ and by a simple basis re-ordering $\boldsymbol{Mix_B}$ is a block diagonal of sixteen 8×8 matrices.

6) Key Schedule: The key length for AES$_{cf}$ is sixteen bytes, while for BES$_{cf}$ it is sixty-four bytes. The 64-byte BES$_{cf}$ key will generate eleven subkeys with the same length. Additionally, the embedded image of the AES$_{cf}$ key k_A is the BES$_{cf}$ key $k_B = \phi(k_A)$, then for every round subkey $(k_B)_i = \phi((k_A)_i)$, the same as BES.

7) A multivariate quadratic $GF(2^4)$-system for BES$_{cf}$: As usual, the S-box linear operation, ShiftRows and MixColumn can be combined into one Matrix denoted by $\boldsymbol{M_B} = \boldsymbol{Mix_B} \cdot \boldsymbol{R_B} \cdot \boldsymbol{Lin_B}$. Denote $\boldsymbol{p}, \boldsymbol{c} \in (GF(2^4))^{128}$ as the plaintext and ciphertext, $\boldsymbol{k}_i \in (GF(2^4))^{128}, i = 0, \cdots, 10$ as the eleven BES$_{cf}$ subkeys, and the state vectors before and after the i^{th} invocation of the $GF((2^4)^2)$ inversion layer by $\boldsymbol{w}_i \in (GF(2^4))^{128}$ and $\boldsymbol{x}_i \in (GF(2^4))^{128}$, $i = 0, \cdots, 9$ respectively, with $\boldsymbol{t}_i \in (GF(2^4))^{64}$ a temporary variable during the $GF((2^4)^2)$ inversion. For each vector above except $\boldsymbol{t}_i$, four subscripts are given, $(j, m, p, q), j, m, q = 0, \cdots, 3, p = 0, 1$, where j, m indicate the $(4*j+m)^{th}$ component corresponding $a_{jm} \in GF((2^4)^2)$ in (14), p indicates one of the two $GF(2^4)$ segments of a_{jm} and q represents the coordinate of conjugate. For $\boldsymbol{t}_i$, the subscript p is discarded since $\boldsymbol{t}_i$ is used only for $GF((2^4)^2)$ inversion where both the $GF(2^4)$ segments of a_{jm} mingle. The BES$_{cf}$ encryption can then be described by the following $GF(2^4)$ systems:

$$0 = w_{0,(j,m,p,q)} + p_{(j,m,p,q)} + k_{0,(j,m,p,q)}, \tag{17}$$

for $i = 0, \cdots, 9$,

$$\begin{aligned} 0 = & \ t_{i,(j,m,q)} + w_{i,(j,m,1,q+1)} \cdot \lambda^{2^q} \\ & + w_{i,(j,m,1,q)} \cdot w_{i,(j,m,0,q)} + w_{i,(j,m,0,q+1)}, \end{aligned} \tag{18}$$

$$0 = t_{i,(j,m,q)} \cdot x_{i,(j,m,1,q)} + w_{i,(j,m,1,q)}, \tag{19}$$

$$\begin{aligned} 0 = & \ t_{i,(j,m,q)} \cdot x_{i,(j,m,0,q)} + w_{i,(j,m,1,q)} \\ & + w_{i,(j,m,0,q)}, \end{aligned} \tag{20}$$

$$0 = w_{i,(j,m,p,q)}^2 + w_{i,(j,m,p,q+1)}, \tag{21}$$

$$0 = x_{i,(j,m,p,q)}^2 + x_{i,(j,m,p,q+1)}, \tag{22}$$

$$0 = t_{i,(j,m,q)}^2 + t_{i,(j,m,q+1)}, \tag{23}$$

for $i = 1, \cdots, 9$,

$$\begin{aligned} 0 = & \ w_{i,(j,m,p,q)} + k_{i,(j,m,p,q)} + \\ & \sum_{(j',m',p',q')} \alpha_{(j',m',p',q')} \cdot x_{i-1,(j',m',p',q')}, \end{aligned} \tag{24}$$

$$\begin{aligned} 0 = & \ c_{(j,m,p,q)} + k_{10,(j,m,p,q)} + \\ & \sum_{(j',m',p',q')} \beta_{(j',m',p',q')} \cdot x_{9,(j',m',p',q')}. \end{aligned} \tag{25}$$

TABLE III. ITEMS FOR THE $GF(2^4)$ SYSTEM OF BES$_{cf}$

Eq.	Num.	Property	Increased Variables	Increased Quadratic Terms
(17)	128	linear	256	0
(18)	640	quadratic	1792	640
(19)	640	quadratic	640	640
(20)	640	quadratic	640	640
(21)	1280	quadratic	0	1280
(22)	1280	quadratic	0	1280
(23)	640	quadratic	0	640
(24)	1152	linear	1152	0
(25)	128	linear	128	0

TABLE IV. ITEMS FOR THE SYSTEM OF BES AND BES$_{cf}$

Items	BES [24]	BES$_{cf}$
Equations in total	5248	6528
Linear Equations	1408	1408
Quadratic Equations	3840	5120
Terms in total	7808	9728
Quadratic Terms	3840	5120
State Variables	2560	3200
Key Variables	1408	1408

Note that the equations in (21), (22) and (23) indicate conjugacy. In (24) and (25), $\alpha_{(j',m',p',q')}$ and $\beta_{(j',m',p',q')}$ denote the elements in $\boldsymbol{M_B}$ and $\boldsymbol{M_B^*} = \boldsymbol{R_B} \cdot \boldsymbol{Lin_B}$ respectively, and $q+1$ is interpreted modulo 4. The numbers of items of the $GF(2^4)$ systems are listed in Table III.

The system contains 6528 equations, of which 1408 are linear and 5120 are (extremely sparse) quadratic equations. The system comprises 9728 terms made from 3200 state variables and 1408 key variables, of which 4608 are linear terms (state variables and key variables), 3200 are square terms and 1920 are quadratic terms. The details are listed in Table IV.

The effectiveness for the algebraic attack lies in the solvability of the system. Courtois and Pieprzyk [20] present a method called XSL to solve the $GF(2)$-system for AES, and it is also available for the $GF(2^8)$-system for BES. However, up till now, there exists no authentic estimation for XSL attack, so it is of no worth to give a complete comparison of the $GF(2^8)$ attack for BES and the $GF(2^4)$ attack for BES$_{cf}$. However, we find three evidences leading to the insolvability of the $GF(2^4)$-system for BES$_{cf}$:

(I) The $GF(2^4)$-system has more terms and more equations. The complexity of the XSL algorithm is on average $O(T^\omega)$ [20]; here T denotes the number of terms.

(II) The $GF(2^4)$-system for BES$_{cf}$ has more quadratic equations than the $GF(2^8)$-system for BES. The quadratic terms of BES$_{cf}$ occupy $5120/9728 = 52.6\%$ in total, more than $3840/7808 = 49.2\%$ for BES. According to the analysis in [20] , the $GF(2^4)$-system for BES$_{cf}$ would be more complex to carry out XSL attack than the $GF(2^8)$-system for BES. In fact, most of the extra terms of BES$_{cf}$ are the new variables $\boldsymbol{t}_i$ used for the special treatment with composite field inversion.

(III) In [25], there is another definition of Algebraic Immunity for the XSL S-box. For each XSL S-box of BES$_{cf}$, the related equations are (18)-(23). First, the size $n = 8$ is fixed because each XSL S-box works on eight state variables of BES$_{cf}$. The terms appearing in (18)-(23) are: $t_{i,(j,m,q)}$, $w_{i,(j,m,p,q)}$, $x_{i,(j,m,p,q)}$, $w_{i,(j,m,1,q)} \cdot w_{i,(j,m,0,q)}$, $t_{i,(j,m,q)} \cdot x_{i,(j,m,1,q)}$, $t_{i,(j,m,q)} \cdot x_{i,(j,m,0,q)}$, $w^2_{i,(j,m,p,q)}$, $x^2_{i,(j,m,p,q)}$, $t^2_{i,(j,m,q)}$. Summing up according to the subscripts p and q, the number of terms is $t = 4+8+8+4+4+4+8+8+4 = 52$ and the number of equations $r = 4+4+4+8+8+4 = 32$, then the Algebraic Immunity for XSL S-box of BES$_{cf}$ is $\Gamma = (\frac{t-r}{n})^{\lceil \frac{t-r}{n} \rceil} = 2.5^3 = 15.625$ which is larger than 9.6 of BES. The S-box based on composite field seems more immune from the potential algebraic attack.

Furthermore, adding the key schedule, the system may not be any simpler since it has the same number of S-box substitutions as BES. By then, we can conclude that AES$_{cf}$ is immune from the $GF(2^4)$-system, and the potential algebraic attack for AES$_{cf}$ may not work.

Similarly, one can consider AES$_{tf}$ in a $GF((2^2)^2)$ system and get a $GF((2^2)^2)$-system for AES$_{tf}$, with the same scale of system of equations as the $GF(2^4)$-system for AES$_{cf}$. And even more, one can think of splitting the $GF((2^2)^2)$ system into the field $GF(2^2)$, where the basic $GF((2^2)^2)$ operations, especially the inversion, have to be replaced by the basic operations on $GF(2^2)$. However, based on what we have done before, splitting $GF((2^4)^2)$ inversion into basic operations on $GF(2^4)$, which complicates the system, one can see that the expected $GF(2^2)$-system for AES$_{cf}$ may not be any simpler.

VI. CONCLUSION

In this paper, we tried to change the computational field used in AES S-box, and created a new class of S-box with better efficiency while preserving the cryptographical security. Two 8×8 S-boxes S_2 and S_3 are constructed, by direct inversion in composite fields $GF((2^4)^2)$ and $GF(((2^2)^2)^2)$ respectively, combined with a $GF(2)$ affine transformation, the same used in AES S-box to give a rational comparison of the composite field. The choice of the subfield leading to the most efficient implementation is mainly discussed. By simple comparison, our new S-boxes have better hardware implementation than AES S-box. The masking strategy against differential power attack is also more convenient.

We also studied the cryptographic characteristics with such a S-box based on composite field inversion. The results show that both S_2 and S_3 have comparatively the same cryptographic characteristics with AES S-box. Thus, the replacement to composite field does not weaken the cryptographic characteristics. Moreover, we investigated whether or not those effective cryptanalysis of AES might work if our S-box took the place, especially the algebraic attack. Due to the different fields involved, algebraic attacks applied on $GF(2)$, $GF((2^4)^2)$ and $GF(2^4)$ are discussed, respectively. And we proved that with the replacement of S_2 or S_3 and the corresponding field for MixColumn operation, the revised AES, denoted by AES$_{cf}$ or AES$_{tf}$, had no effective algebraic attack and was even more solid than the original AES with S_1.

In fact, the essence of our design is just to try to overlook the underlying computational fashion and to choose the most efficient one while preserving the cryptographic characteristics. The most compact AES S-box to date was created by normal bases [11]. The advantage for normal bases is that they have very sparse matrixes in the implementation compared with polynomial bases [11], but finding inversion will be as hard as under polynomial bases. The S-box constructed on normal bases would also survive those attacks.

 ISBN: 978-1-61208-373-5

TABLE V. COMPOSITE FIELD IN BLOCK CIPHER

Cipher	Field	Structure
AES	$GF(2^8)$	$\boldsymbol{A} \cdot x^{-1} + \boldsymbol{v}$
SMS4	$GF(2^8)$	$\boldsymbol{A} \cdot (\boldsymbol{A} \cdot x + \mathbf{v})^{-1} + \boldsymbol{v}$
CLEFIA	$GF(2^8)$	$\boldsymbol{A}_2 \cdot (\boldsymbol{A}_1 \cdot x + \boldsymbol{v}_1)^{-1} + \boldsymbol{v}_2$
Camellia	$GF((2^4)^2)$	$\boldsymbol{A}_2 \cdot (\boldsymbol{A}_1 \cdot (x + \boldsymbol{v}_1))^{-1} + \boldsymbol{v}_2$

Based on our argument, we suggest composite field $GF((2^n)^2)$ in the design of block cipher. And we think that our settings for S_2 or S_3 is indeed a balance between the implementation complexity and the theoretical security. It seems that the designers of block ciphers did not truly realize the advantage of $GF((2^n)^2)$, see Table V. Even though Camellia uses composite field, the structure is the most complex. As a result, we suggest SMS4 and CLEFIA use composite field for a more efficient hardware implementation.

ACKNOWLEDGMENT

This work is supported by Major Program of National Natural Science Foundation of China (NO.11290141), the International Cooperation Project (NO.2010DFR00700), and the Fundamental Research of Civil Aircraft (NO.MJ-F-2012-04). We are grateful to the Key Laboratory of Mathematics, Informatics and Behavioral Semantics, Ministry of Education for funding us to fulfil this work.

REFERENCES

[1] K. Nyberg, "Differentially uniform mappings for cryptography," in Advances in Cryptology – EUROCRYPT 93, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1994, vol. 765, pp. 55–64.

[2] J. Daemen and V. Rijmen, The design of Rijndael: AES–the Advanced Encryption Standard, ser. Information security and cryptography. Springer, 2002.

[3] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita, "Camellia : A 128-bit block cipher suitable for multiple platforms - design and analysis," in Selected Areas in Cryptography, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, vol. 2012, pp. 39–56.

[4] S. Corporation, "The 128-bit blockcipher CLEFIA – algorithm specification (revision 1.0)," 2007, URL: http://www.sony.net/Products/cryptography/clefia/download/data/clefia-spec-1.0.pdf [retrieved: July, 2014].

[5] W. Diffie and G. Ledin, "SMS4 encryption algorithm for wireless networks," Cryptology ePrint Archive, Report 2008/329, 2008, URL: http://eprint.iacr.org/2008/329.pdf [retrieved: July, 2014].

[6] B. Sunar, E. Savas, and C. Koc, "Constructing composite field representations for efficient conversion," IEEE Transactions on Computers, vol. 52, no. 11, nov. 2003, pp. 1391 – 1398.

[7] A. Rudra, P. K. Dubey, C. S. Jutla, V. Kumar, J. R. Rao, and P. Rohatgi, "Efficient Rijndael encryption implementation with composite field arithmetic," in Cryptographic Hardware and Embedded Systems – CHES 2001, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, vol. 2162, pp. 171–184.

[8] A. Satoh, S. Morioka, K. Takano, and S. Munetoh, "A compact Rijndael hardware architecture with S-Box optimization," in Advances in Cryptology – ASIACRYPT 2001, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, vol. 2248, pp. 239–254.

[9] N. Mentens, L. Batina, B. Preneel, and I. Verbauwhede, "A systematic evaluation of compact hardware implementations for the Rijndael S-Box," in Topics in Cryptology - CT-RSA 2005, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, vol. 3376, pp. 323–333.

[10] D. Canright, "A very compact S-Box for AES," in Cryptographic Hardware and Embedded Systems - CHES 2005, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, vol. 3659, pp. 441–455.

[11] X. Zhang and K. K. Parhi, "On the optimum constructions of composite field for the AES algorithm," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 53, no. 10, oct. 2006, pp. 1153–1157.

[12] S. Nikova, V. Rijmen, and M. Schläffer, "Using normal bases for compact hardware implementations of the AES S-Box," in Security and Cryptography for Networks, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008, vol. 5229, pp. 236–245.

[13] Y. Nogami, K. Nekado, T. Toyota, N. Hongo, and Y. Morikawa, "Mixed bases for efficient inversion in F(((2^2)^2)^2) and conversion matrices of subbytes of AES," in Cryptographic Hardware and Embedded Systems, CHES 2010, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, vol. 6225, pp. 234–247.

[14] B. Hannes, C. Andreas, and H. Max, "On computing multiplicative inverses in GF(2$^{\mathrm{m}}$)," IEEE Transactions on Computers, vol. 42, no. 8, aug 1993, pp. 1010 –1015.

[15] E. Oswald, S. Mangard, N. Pramstaller, and V. Rijmen, "A side-channel analysis resistant description of the AES S-Box," in Fast Software Encryption, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, vol. 3557, pp. 199–228.

[16] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in Advances in Cryptology – CRYPTO 99, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1999, vol. 1666, pp. 789–789.

[17] O. Özen, K. Varıcı, C. Tezcan, and Ç. Kocair, "Lightweight block ciphers revisited: Cryptanalysis of reduced round present and hight," in Information Security and Privacy, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, vol. 5594, pp. 90–107.

[18] C. Carlet, "Boolean functions for cryptography and error correcting codes," in Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Y. Crama and P. L. Hammer, Eds. Cambridge University Press, 2010, pp. 257–397, URL: http://www.math.univ-paris13.fr/~carlet/chap-fcts-Bool-corr.pdf [(a preliminary version) retrieved: July, 2014].

[19] T. Jakobsen and L. R. Knudsen, "The interpolation attack on block ciphers," in Fast Software Encryption, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1997, vol. 1267, pp. 28–40.

[20] N. T. Courtois and J. Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations," Cryptology ePrint Archive, Report 2002/044, 2002, URL: http://eprint.iacr.org/2002/044.pdf [retrieved: July, 2014].

[21] N. Ferguson, J. Kelsey, S. Lucks, B. Schneier, M. Stay, D. Wagner, and D. Whiting, "Improved cryptanalysis of Rijndael," in Fast Software Encryption, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, vol. 1978, pp. 136–141.

[22] H. Gilbert and M. Minier, "A collision attack on 7 rounds of Rijndael," in AES Candidate Conference'00, 2000, pp. 230–241.

[23] A. Biryukov and D. Khovratovich, "Related-key cryptanalysis of the full aes-192 and aes-256," in Advances in Cryptology - ASIACRYPT 2009, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, vol. 5912, pp. 1–18.

[24] S. Murphy and M. J. Robshaw, "Essential algebraic structure within the AES," in Advances in Cryptology – CRYPTO 2002, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2002, vol. 2442, pp. 1–16.

[25] S. Oh, C. H. Kim, J. Lim, and D. H. Cheon, "Remarks on security of the AES and the XSL technique," Electronics Letters, vol. 39, no. 1, Jan 2003, pp. 36–38.

 ISBN: 978-1-61208-373-5

Timing Synchronization Method for MIMO-OFDM Systems with CAZAC Sequences

Ali Rachini
IETR - INSA de Rennes, France
GET/UL - Lebanese University
Email: Ali.Rachini@ul.edu.lb

Fabienne Nouvel
IETR - INSA de Rennes
Rennes, France
fabienne.nouvel@insa-rennes.fr

Ali Beydoun and Bilal Beydoun
GET/UL - Lebanese University
Hadath, Lebanon
bilbey@ul.edu.lb
beydoual@yahoo.fr

***Abstract*—Multiple-Input Multiple-Output (MIMO) Orthogonal Frequency Division Multiplexing (OFDM) systems are very sensitive to carrier frequency offset (CFO) and timing synchronization. In this paper, a new timing synchronization preamble designed for MIMO-OFDM systems is presented. Constant Amplitude Zero Auto Correlation (CAZAC) sequences are used in order to construct this preamble. CAZAC sequence has a sharp correlation peak and zero side lobes. Simulation results show that the proposed method presents a good performance at a low Signal to Noise Ratio (SNR) in AWGN and multipath fading Rayleigh channels.**

***Keywords*-MIMO-OFDM systems; Timing Synchronization; CAZAC sequences; orthogonal preamble.**

I. INTRODUCTION

Wireless communications can be regarded as the most important development that has an extremely wide range of applications. In this new information age, high data rate and reliability features are required for any wireless communication system. MIMO-OFDM (Multiple Input Multiple Output - Orthogonal Frequency Division Multiplexing) is the most recent wireless broadband technology. This technology has gained great popularity for its capability of high rate transmission and its robustness against multi-path fading and other channel impairments. Therefore, the combinaison between MIMO and OFDM systems is proposed in 802.11n [1].

The OFDM [2] became a very popular multi-carrier modulation technique for transmission of signals over wireless channels. It converts a frequency-selective fading channel into different parallel flat fading sub-channels, thanks to the FFT's algorithm (Fast Fourier Transform) [3]. The inverse FFT algorithm (IFFT) [3] is also used to demodulate the message at the receiver. Hence, the bandwidth is utilized efficiently in OFDM systems without causing the Inter-Carrier Interference (ICI). OFDM combines multiple low-data-rate subcarriers into high-data-rate with a long symbol duration in order to eliminate the Inter-Symbol Interference (ISI).

MIMO exploits the space dimension to improve wireless systems capacity, range and reliability. It offers significant increases in data throughput and link range without additional bandwidth or increased transmit power. MIMO achieves this goal by spreading the same total transmit power over the antennas to achieve an array gain that improves the spectral efficiency (Spatial Multiplexing (SM)) or to achieve a diversity gain that improves the link reliability (Space Time Coding (STC)).

The SM transmits independent data rates over different N_t transmit antennas in order to increase the throughput between the transmitter and the receiver. Foshini [4] has shown that the theoretical capacity of the MIMO channel, with N_t transmit antennas and N_r receive antennas, increases linearly with min(N_t,N_r). The STC is increasing the performance by sending redundant data over different transmit antennas [5]. In this paper, the MIMO-OFDM system is based on the STC technique. In order to improve the link reliability, we will focus on Space-Time Block Code (STBC) Alamouti code [6] [7].

A major challenge for MIMO-OFDM system is the synchronization between transmitter and receiver. Two types of synchronization are necessary, namely, the timing and the frequency synchronization. The coarse timing synchronization is to detect the beginning of the OFDM frame and the fine timing synchronization is used for coherent detection of OFDM symbols. The frequency synchronization is to correct the phase error caused by the mismatch of the local oscillator (LO) between transmitter and receiver [8]. In this paper, we will focus on timing synchronization.

In the literature, several synchronization approaches have been proposed for MIMO-OFDM systems [9] [10]. Most of the timing synchronization methods are preamble based. Therefore, the synchronization preamble should have a good correlation function in order to detect the packet arrival at the receiver.

In this work, based on [11], we propose a new timing synchronization method for MIMO-OFDM systems with CAZAC sequences. Furthermore, Constant Amplitude Zero Auto-Correlation (CAZAC) sequence [12] has constant amplitude and zero autocorrelation for all non-zero shifts. The main characteristics of CAZAC sequences are their correlation functions. They have a good autocorrelation function and their

 ISBN: 978-1-61208-373-5

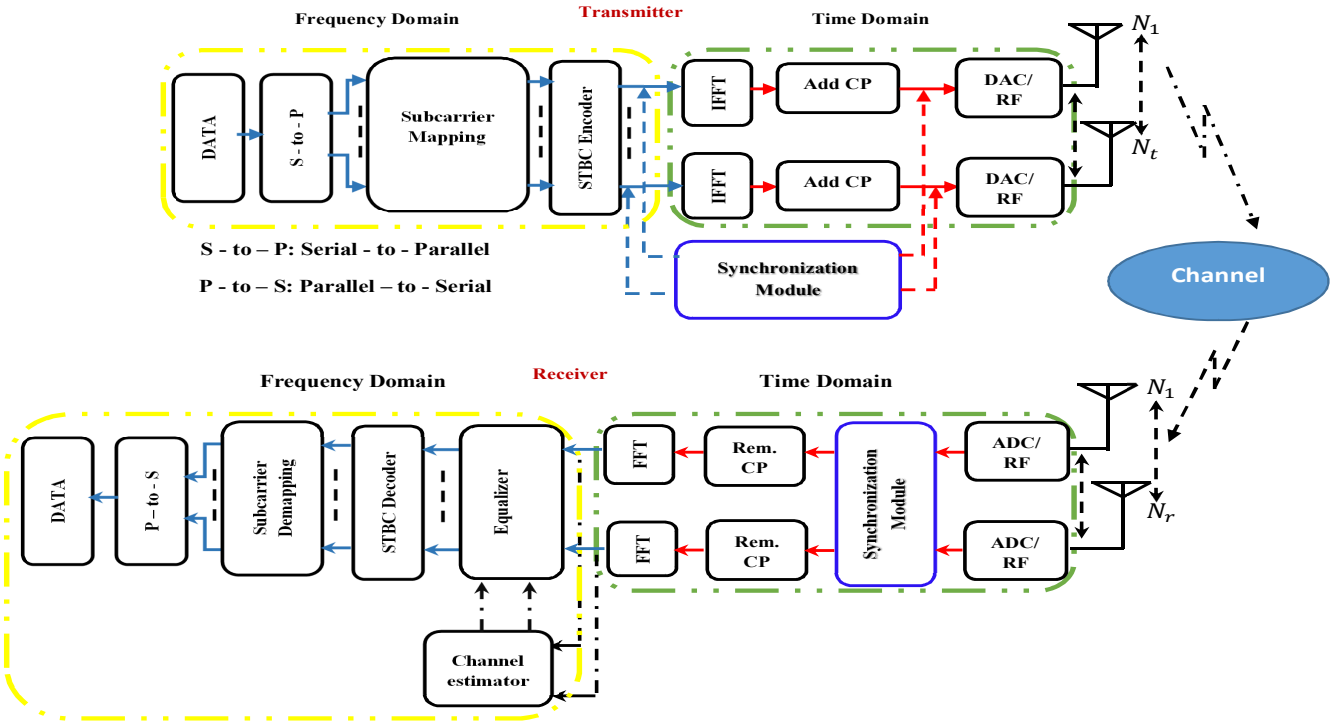


Figure 1. Transmission system MIMO-OFDM-STBC

crosscorrelation function is near zero. Due to their orthogonality, CAZAC sequences reduce inter-code interference between multiple antennas and have a lower Peak-to-Average Power Ratio (PAPR). As a result, CAZAC sequences are regarded as optimum preamble for timing synchronization in MIMO-OFDM systems.

Based on CAZAC sequences, a compact preamble design for synchronization in distributed MIMO-OFDM systems has been proposed in [10]. Training symbols, based on exclusive subband, have designed. The main drawback of this approach is the number of transmit antennas. When this number is increased, the length of the synchronization sequence decreases, and hence, the timing synchronization performance decreases. As result, [10] shows a good timing synchronization with an SNR of 15dB. Furthermore, using the same simulation parameters and the same propagation channel delay, our proposed method shows a perfect timing synchronization against the proposed method in [10] at low SNR.

The aim of this paper is to present a new timing synchronization method for MIMO-OFDM systems with CAZAC sequences. Section II briefly describes the MIMO-OFDM system structure based on STBC. Our proposed method and preamble structure are presented in Section III. Simulation results and conclusion are done in Section IV and V respectively.

II. System Model

The combinaison of MIMO and OFDM systems is one of the most effective techniques to improve spectral efficiency of radio communications. In Figure 1, we consider a MIMO system using OFDM modulation and N_{sc} subcarriers per transmit antenna, where the transmitter and receiver are respectively provided with N_t and N_r antennas ($N_t, N_r \in \{2, 4, 8\}$).

In the following, we describe the MIMO-OFDM system presented in Figure 1.

A parallel data stream is passed through a digital modulator. This modulator encodes the data stream with 16-QAM constellation. The complex symbols are then fed into an STBC encoder in order to encode the data stream with Alamouti encoder. Then, symbols pass through the OFDM modulator. This modulator can be done by using a simple Inverse Fast Fourrier Transform (IFFT) algorithm. In which, the output signal of OFDM modulator, is in time domain. After the IFFT, a Guard Interval (GI) block is presented in order to append the Cyclic Prefix (CP) at the beginning of each OFDM symbol. It refers to a copy of the last portion of the OFDM symbol appended to the front of the symbol, in order to reduce the ISI. Letting T_s is the total symbol period ($T_s = T_g + T_u$), T_g is the period of cyclic prefix, T_u is the period of useful data.

The synchronization block is used in order to insert a known synchronization preamble at the beginning of each OFDM frame. This block could be presented in time domain [13] or in frequency domain [11]. In this paper, we focus in the second approach. At the receiver, the synchronization is performed in the time domain.

The second part of MIMO-OFDM system is the receiver.

The first block of the receiver is the synchronization block. This block is presented in time domain. After a good synchronization, a CP removal block is applied in order to remove the CP from the beginning of each OFDM symbol. The OFDM symbols are demodulated using FFT algorithm. Equalizer and channel estimator are the first two blocks in frequency domain in order to estimate and detect the channel coefficients. After equalization, the STBC encoder is implemented using Alamouti decoder. After the QAM demodulator, we get the data stream.

A. Transmitted signal

At transmitter, x_k is the symbol on the frequency f_k. The OFDM transmit signal $s_i(t)$ transmitted over i^{th} transmit antenna can be expressed as follows [14]:

$$s_i(t) = \frac{1}{\sqrt{N_{sc}}} \sum_{k=0}^{N_{sc}-1} \mathfrak{Re}\left\{x_k e^{j.2\pi.f_k.t}\right\} \quad (1)$$

where N_{sc} is the number of sub-carriers.

B. Channel

The transmitted signal reaches the receiver, by undergoing many effects, over several different paths. The multipaths fading channel between transmit antenna T_i, $i \in \{1, N_t\}$ and receive antenna R_j, $j \in \{1, N_r\}$, is written as:

$$h^{i,j}(\tau, t) = \sum_{p=1}^{P_{ij}} \alpha_p(t) e^{-j2\pi f_k \mathcal{T}_p(t)} .\delta\left[\tau - \tau_p(t)\right] \quad (2)$$

where $\alpha_p(t)$ is the attenuation factor for the signal received on the p^{th} path with the propagation delay $\tau_p(t)$.

C. Received signal

The transmitted signal $s_i(t)$ from i^{th} transmit antenna undergoes fading by the channel before reaching the j^{th} receive antenna. The received signal $r_j(t)$ is written as:

$$\begin{aligned} r_j(t) &= \sum_{i=1}^{N_t} \left[h^{i,j}(\tau, t) \star x_i(t)\right] + n_{ij}(t) \\ &= \frac{1}{\sqrt{N_{sc}}} \sum_{i=1}^{N_t} \sum_{p=1}^{P_{ij}} \left[\alpha_p(t) e^{-j2\pi f_k \mathcal{T}_p(t)} .s_i\left[\tau - \tau_p(t)\right]\right] e^{j2\pi f_k(t)} \\ &\quad + n_{ij}(t) \end{aligned} \quad (3)$$

where h_{ij} is the channel between the transmit antenna T_i and the receive antenna R_j, τ is the propagation delay for the different channels paths, α_p is the attenuation for the p^{th} path, $s_i(t)$ is the OFDM transmitted signal, P_{ij} is the number of path between T_i and R_j and n_{ij} is the AWGN noise between T_i and R_j.

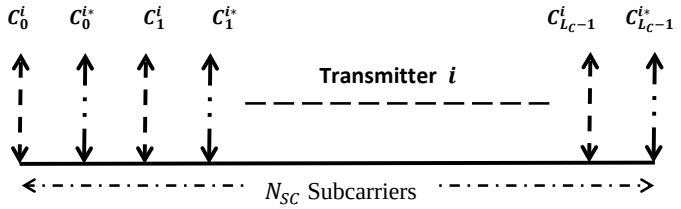


Figure 2. Preamble Structure in Frequency Domain

III. Proposed Method

Based on Rachini et al. [11], a new timing synchronization preamble is presented, in this section. This structure is used in order to estimate the beginning of the OFDM received frame and to detect the beginning of useful OFDM symbols in each frames. This preamble structure is generated in the frequency domain, as shown in Figure 2.

Let i is the transmit antenna in MIMO-OFDM system and C is a CAZAC sequence. C is given by the following equation:

$$C(k) = \begin{cases} e^{j\left(\frac{\pi M k(k+1)}{L_C}\right)} & \text{if } L_C \text{ is odd} \\ e^{j\left(\frac{\pi M k^2}{L_C}\right)} & \text{if } L_C \text{ is even} \end{cases} \quad (4)$$

where $L_C = L_{FFT}/2$ is the length of the CAZAC sequence, $n \in \mathbb{N}$, $M \in \mathbb{N}$ is a prime number with L_C and $k \in \{0, L_C - 1\}$ is the index of the sample.

In this structure, we combined a CAZAC sequence with its $conjugate$. This combinaison gives a time-domain complex envelope form that have a good cross-correlation and autocorrelation functions. This combination retains the orthogonality between different preambles over different transmit antennas. Figure 3 shows the preamble structure in time domain.

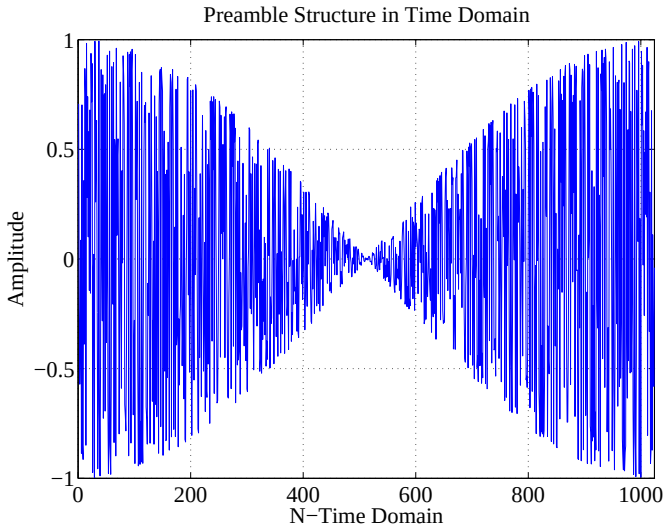


Figure 3. Preamble structure in time domain

Figure 4 represents the different orthogonal preamble structure over different N_t transmit antennas. This preamble structure can be

applied regardless of the number of transmit or receive antennas.

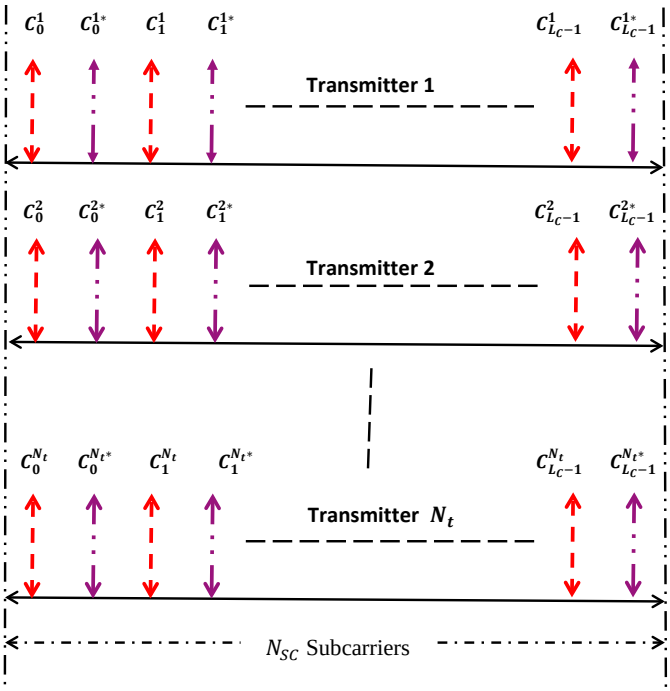


Figure 4. Frame structure in frequency domain

In Figure 4, each preamble contains a CAZAC sequence C mapped on the odd subcarrier and the conjugate of C is mapped on the even subcarrier.

At receiver, a correlation function is applied between the received signal r_j and the local sequence seq_j, in order to detect the timing synchronization peak. The correlation function $\mathcal{R}_{r_j,seq_j}$ is calculated as:

$$\mathcal{R}_{r_j,seq_j}(n) = \sum_{n=1}^{L} [r_j(n) * seq_j(n-\tau)] \quad (5)$$

where n is the index of the sample, equivalent to the subcarrier index.

The timing synchronization estimator, $\widehat{ind}_n$, is given by:

$$\widehat{ind}_n = \underset{n}{\operatorname{argmax}}\{\|\mathcal{R}_{r_j,seq_j}(n)\|\} \quad (6)$$

The $\widehat{ind}_n$ is the timing estimate where n is considered as the timing synchronization point.

IV. Simulation Result

Simulation results have been conducted to validate the performance of the proposed preamble structure, in both AWGN channel and multipaths fading channel. In order to evaluate the performance of our proposed preamble against [10], a SISO-OFDM and MIMO-OFDM systems up to 8×8 transmit and receive antennas were simulated.

A. Simulation parameters

An OFDM system was developed, with 512 and 1024 subcarriers ($L_{FFT} = 512, L_{FFT} = 1024$ resp.) was considered in Rayleigh multipaths fading channel with 6 paths sample-spaced with T_s (Sampling Time) suggested by the IEEE 802.11 Working Group [15]. The parameters used for the simulations are summarized in Tables I and II.

TABLE I: SIMULATION PARAMETERS.

Simulation Parameters	Value
MIMO system	up to 8×8
FFT/IFFT Length	1024 & 512
Cyclic Prefix Length	$L_{FFT}/4$
Channel Type	Multipath Rayleigh and AWGN channel
Sequences	CAZAC
Length of orthogonal code L_C	$L_{FFT}/2$
Number of channel taps between different antennas [15]	6 Taps
SNR over all the OFDM Frame	from 0 dB to 25 dB

TABLE II: THE AVERAGE POWER PROFILE OF THE MULTIPATH RAYLEIGH CHANNEL MODEL.

Multipath propagation delays [15]	$[0.T_s, 1.T_s, 2.T_s, 3.T_s, 4.T_s, 5.T_s]$
The power of each multipath Tap [15]	[0.8111, 0.1532, 0.0289, 0.0055, 0.0010, 0.0002]

The correlation function $\mathcal{R}_{r_j,seq_j}$, at the j^{th} receive antenna, is calculated in time domain. Due to the frequency distribution of CAZAC sequence, C and C^*, in each preamble, the $\mathcal{R}_{r_j,seq_j}$ may have a high peak's value. The timing synchronization estimator, $\widehat{ind}_n$, detects the beginning of each OFDM received frame once the value of the correlation peak reaches a defined threshold value.

B. Results

In this section, the acquisition probability (P_{SYNC}) is evaluated in term of different value of SNR (Signal to Noise Ratio). P_{SYNC} represents the probability of successful timing synchronization. Simulation parameters are shown in Tables I and II.

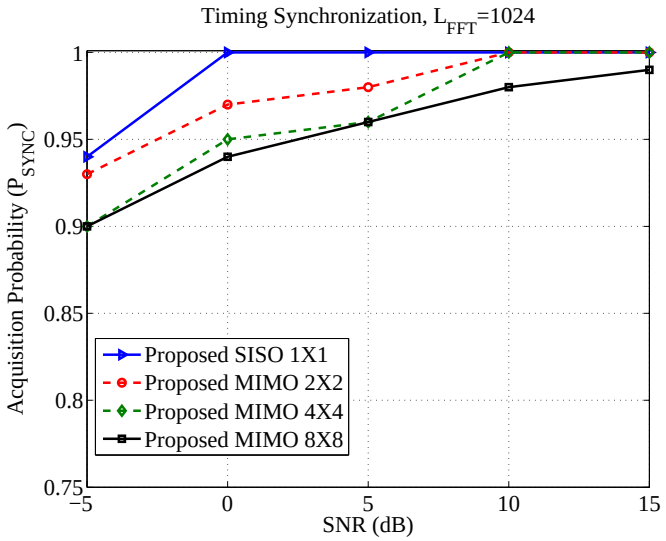


Figure 5. Timing synchronization performance of the proposed method ($L_{FFT} = 1024$)

 ISBN: 978-1-61208-373-5

Figure 5 represents the acquisition probability, in term of SNR, for different OFDM systems (SISO-OFDM 1×1, MIMO-OFDM up to 8×8) using CAZAC sequences where the length of preamble is $L_{FFT} = 1024$.

Figure 5 presents a good timing synchronization for a low SNR. For an $SNR = -5dB$, the $P_{SYNC} \geq 90\%$ for all MIMO-OFDM system up to 8×8. Therefore, for an $SNR = 0dB$, the proposed timing synchronization preamble shows a perfect timing synchronization for SISO-OFDM system. The $P_{SYNC} \geq 97\%$ for MIMO-OFDM system 2×2 for the same SNR. For a MIMO-OFDM system 4×4 the $P_{SYNC} \geq 96\%$ at an $SNR = 5dB$. On the other hand, for MIMO-OFDM system 8×8, the acquisition probability P_{SYNC} reaches 98% at an $SNR = 10dB$.

Figure 6 presents the performance of our synchronization preamble of length $L_{FFT} = 512$. In this Figure, the acquisition probability P_{SYNC} is greater than 97% for both SISO-OFDM and MIMO-OFDM 2×2 systems at an $SNR = 0\,dB$. Therefore, $P_{SYNC} \geq 90\%$ for MIMO-OFDM 4×4 system at an $SNR = 0dB$. In the other hand, the P_{SYNC} reaches 80% at an $SNR = 5\,dB$ for MIMO-OFDM system8×8.

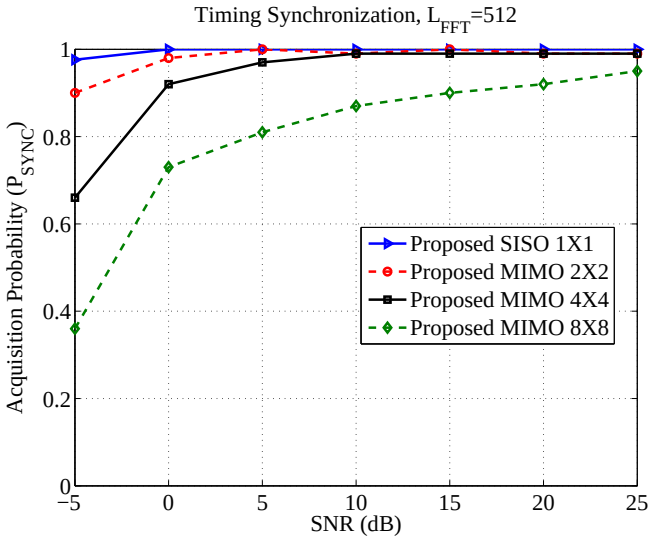


Figure 6. Timing synchronization performance of the proposed method ($L_{FFT} = 512$)

In Table III, the simulation results of Figures 5 and 6, are summarized. It can be shown that the performance of our timing synchronization method increases with the length of L_{FFT}. Moreover, the results of Figure 5 ($L_{FFT} = 1024$) show a good performance against those presented in Figure 6 ($L_{FFT} = 512$).

In order to evaluate the performance of our proposed method, we conducted an extensive comparison of our approach with the synchronization scheme of [10]. Hung and Chin [10] Wang used a subband-based preamble based on CAZAC sequences. The main drawback of this method is the number of transmit antennas. As the number of transmit antennas increases, the length of synchronization sequence, on each transmit antenna, decreases. Therefore, the value of the synchronization peak at the receiver decreases.

Figure 7 presents the performance between our proposed approach and the synchronization scheme of [10]. Simulation results in Figure 7

TABLE III: COMPARISON BETWEEN THE ACQUISITION PROBABILITY OF DIFFERENT MIMO-OFDM SYSTEMS, IN TERM OF SNR AND LENGTH OF FFT

Acquisition probability			
MIMO-OFDM system	P_{SYNC}	**SNR (dB)**	L_{FFT}
MIMO-OFDM 2x2	≥97%	>0 dB	1024
	≥96%	>0 dB	512
MIMO-OFDM 4x4	≥95%	>0 dB	1024
	≥93%	>0 dB	512
MIMO-OFDM 8x8	≥94%	>0 dB	1024
	≥78%	>0 dB	512

are done with the simulation parameters of Tables I and II with a synchronization preamble of length $L_{FFT} = 256$, and MIMO-OFDM system 2×2 and 3×3.

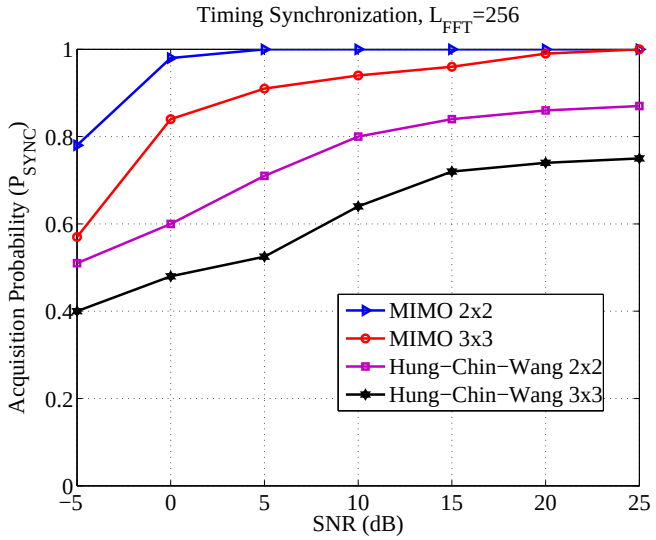


Figure 7. Comparisons between the proposed approach and subband-based preamble [10]

Simulation results of our proposed approach, have a good performance against [10] at a low SNR. The acquisition probability P_{SYNC} for our method is greater than 90% at an $SNR \geq 5\,dB$ for both MIMO-OFDM 2×2 and 3×3 system. Therefore, the proposed method in [10] shows that the acquisition probability is between 0.5 and 0.75 at the same value of SNR.

V. CONCLUSION AND FUTURE WORK

The major challenges in MIMO-OFDM communication systems are the synchronization and the channel estimation. In this paper, we proposed a new timing synchronization preamble structure, based on [11], in order to detect the timing frame synchronization. At the receiver, a correlation function, between received signal and local sequence, is applied in order to detect the beginning of OFDM received frames. Hence, due to the combination of CAZAC sequence C and C^* in the synchronization preamble, the correlation function shows a good frame detection as the number of transmit antenna increases. Simulation results of our proposed method presents good timing frame synchronization against the subband-based preamble timing synchronization method in [10]. Therefore, this preamble structure shows a good timing acquisition probability at a low SNR.

 ISBN: 978-1-61208-373-5

Furthermore, this approach can be implemented with a large number of transmit antennas of MIMO-OFDM system.

ACKNOWLEDGMENT

This work was supported by GET of the Lebanese University and IETR of INSA-Rennes, France.

REFERENCES

[1] "Ieee draft standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements - part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications," IEEE Draft P802.11-REVmb/D3.0, March 2010 (Revision of IEEE Std 802.11-2007, as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11w-2009 and IEEE Std 802.11n-2009), 11 2010, pp. 1–2228.

[2] L. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," Communications, IEEE Transactions on, vol. 33, no. 7, Jul 1985, pp. 665–675.

[3] S. G. Johnson and M. Frigo, "Implementing ffts in practice," 2009.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Communications, vol. 6, 1998, pp. 311–335.

[5] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-blast: an architecture for realizing very high data rates over the rich-scattering wireless channel," in Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on, 1998, pp. 295–300.

[6] S. Alamouti, "A simple transmit diversity technique for wireless communications," vol. 16, no. 8, Oct. 1998, pp. 1451–1458.

[7] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," Information Theory, IEEE Transactions on, vol. 45, no. 5, 1999, pp. 1456–1467.

[8] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," Communications, IEEE Transactions on, vol. 47, no. 2, Feb 1999, pp. 199–207.

[9] L. Li and P. Zhou, "Synchronization for b3g mimo ofdm in dl initial acquisition by cazac sequence," in Communications, Circuits and Systems Proceedings, 2006 International Conference on, vol. 2, June 2006, pp. 1035–1039.

[10] H.-C. Wang and C.-L. Wang, "A compact preamble design for synchronization in distributed mimo ofdm systems," in Vehicular Technology Conference (VTC Fall), 2011 IEEE, Sept. 2011, pp. 1–4.

[11] A. Rachini, A. Beydoun, F. Nouvel, and B. Beydoun, "A novel compact preamble structure for timing synchronization in mimo-ofdm systems using cazac sequences," International Conference on Communications, Computation, Networks and Technologies (INNOV), 2013, pp. 1–6.

[12] R. Frank, S. Zadoff, and R. Heimiller, "Phase shift pulse codes with good periodic correlation properties (corresp.)," Information Theory, IRE Transactions on, vol. 8, no. 6, Oct. 1962, pp. 381–382.

[13] A. Rachini, A. Beydoun, F. Nouvel, and B. Beydoun, "Timing synchronisation method for mimo-ofdm system using orthogonal preamble," in Telecommunications (ICT), 2012 19th International Conference on, 2012, pp. 1–5.

[14] A. Rachini, A. Beydoun, F. Nouvel, and B. Baydoun, "International journal on advances in telecommunications," in Information Theory, 1993. Proceedings. 1993 IEEE International Symposium on, vol. 1, no. 1 & 2, July 2014, pp. 22–33.

[15] B. O'Hara and A. Petrick, The IEEE 802.11 Handbook: A Designer's Companion. Standards Information Network IEEE Press, 1999.

 ISBN: 978-1-61208-373-5