

**Proceedings of the 6th International
Cryptology and Information Security
Conference 2018**

9th – 11th July 2018
Port Dickson, Negeri Sembilan,
Malaysia

Published by
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
43400 UPM Serdang
Selangor Darul Ehsan

©Institute for Mathematical Research (INSPEM), 2018

All rights reserved. No part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in a review written for inclusion in a magazine or newspaper.

First Print 2018

Perpustakaan Negara Malaysia Cataloguing – in – Publication Data

International Cryptology and Information Security Conference

(2018: Port Dickson, Negeri Sembilan)

Proceedings of the 6th International Cryptology and Information

Security Conference 2018 / Editors: Muhammad Rezal Kamel Ariffin,

Goi Bok Min, Hailiza Kamarulhaili, Heng Swee Huay, Moesfa Soeheila

Mohamad, Mohamad Rushdan Md. Said.

ISBN 978-983-44069-6-7

1. Cryptography - - Research - - Malaysia - - Congresses.

2. Computer security - - Research - - Malaysia- - Congresses.

3. Government publications - - Malaysia.

I. Muhammad Rezal Kamel Ariffin II. Goi, Bok Min.

III. Hailiza Kamarulhaili. IV. Heng, Swee-Huay.

V. Moesfa Soeheila Mohamad. VI. Mohamad Rushdan Md. Said. VII. Title.

005.8072

OPENING REMARKS

First and foremost, I would like to thank the Malaysian Society for Cryptology Research (MSCR) in collaboration with CyberSecurity Malaysia together with Universiti Putra Malaysia (UPM) for their continuous efforts and commitment to host this premier event, the International Cryptology and Information Security Conference for the sixth time. This biannual conference series which started in 2008 has been organized and



hosted at several locations in Malaysia, beginning from Kuala Lumpur to Melaka, and then on to Langkawi, Putrajaya and Kota Kinabalu. This year, Port Dickson has been chosen as the venue.

CRYPTOLOGY2018 is the sixth among a series of open forum conferences for avid researchers of theoretical foundations, applications and any related issues in cryptology, information security and other underlying technologies to contribute to this body of knowledge. For this year's CRYPTOLOGY2018, participation of researchers from various disciplines is impressive and this signifies the interdisciplinary nature of the topic of the conference. A total of 17 research results are scheduled to be delivered in this forum. Two speakers from Japan and Turkey will conduct workshops for the benefit of the participants. I hope that through this kind of events and activities, we can promote new research interests and eventually developed more expertise in this field.

Cryptology is the last line of defence in protecting information. In the absence of cryptographic measures protecting one's critical information, it cannot be ascertained that information is secured from adversaries. Hence, Malaysia must be prepared in protecting her Critical National Information Infrastructure (CNII) in the coming years as criminals and other adversaries will gain access to new technology and skills to obtain these critical information. In view of the importance of cryptography in national cyber security, National Cryptography Policy or *Dasar Kriptografi Negara* was established under the purview of National Security Council. It has seven strategic thrusts that focus on the aspect of competency and self reliant in cryptography towards ensuring the protection of national security, citizens privacy and safety; and making cryptography industry as a contributor to the nations wealth creation. CyberSecurity Malaysia together

with National Security Council are the joint secretariat to monitor the implementation of the policy.

Universiti Putra Malaysia (UPM), Universiti Sains Malaysia (USM), Multimedia University (MMU), Universiti Teknikal Malaysia Melaka (UTeM), Universiti Tunku Abdul Rahman (UTAR) and several other renowned universities have been in collaboration with CyberSecurity Malaysia and the Malaysian Society for Cryptology Research for about 10 years now. I believe that these universities somehow or rather, have to a certain extent, provided a platform for R&D in this area. Other than R&D, it is also important for the experts in cryptology and researchers working in this field, to work hand in hand and enhance their networking and communications. Therefore, this conference provides a good platform for information sharing, and to showcase new technology in internet security.

I would also like to take this opportunity to give special thanks to Augmented Technology Sdn Bhd in supporting this event as its corporate sponsor. The willingness of this one of the few companies in Malaysia to engage directly with the academia shows growing support from the industry in this field to enhance its competitiveness especially among the companies that have direct application of cryptography in their business products and services. I hope this effort can be continued in the future.

Finally to all the participants, I wish you every success in your future endeavor and a fruitful and productive conference.

Thank you.

YBRS. TUAN IR. MD. SHAH NURI BIN MD. ZAIN

Chief Executive,

National Cyber Security Agency,

National Security Council.

WELCOMING NOTES

I am very pleased to welcome speakers from countries across the world to the 6th International Cryptology and Information Security Conference 2018 (CRYPTOLOGY2018). It is our hope that participants will grab this opportunity and gain valuable experience either through formal or informal discussion during this intellectual meeting.



Cryptography is an area of research that has tremendous impact especially in the area of communication technology. In this respect, CRYPTOLOGY2018 will provide an avenue for participants to engage on current topics related to cryptology. It is also aimed at promoting and encouraging exchange of ideas and at the same time identifying areas of collaborative research between local and foreign researchers.

Information security has never become as important in our daily lives as we are experiencing today. We are now on the brink of experiencing cryptography and its deployment in every corner of our day to day experiences. Thus, research in this area has become extremely important that without continuous effort to conduct research in the area one would not be able to ascertain the degree of security being deployed. Therefore it is our responsibility to ensure this biennial gathering is held in a best possible manner such that pool of excellent ideas can be brought together to solve current and future problems.

In this conference, we have two workshop speakers namely from Japan and Turkey who are renowned researchers in their respective areas. We also have 17 papers scheduled to be presented encompassing various areas of cryptology such as theoretical foundations, applications, information security and other underlying technologies in this interesting mathematical field. I hope this conference will bring Malaysia further towards realizing and translating research into a good cryptography practices.

It goes without saying that a conference of this kind could not have been held without the committed efforts of various individuals and parties. I would like to take this opportunity to congratulate and thank everyone involved for their excellent work and in particular to Universiti Putra Malaysia (UPM) and CyberSecurity Malaysia for taking up the challenge of organizing

this conference. I would also like to thank our corporate partner Augmented Technology, who has helped to realize this event. I wish CRYPTOLOGY2018 will gives all participants great experience, enjoyable and meaningful moments. With that, I once again thank all speakers, presenters and participants in making this conference possible and a successful event.

Thank you.

PROF. DR. HAILIZA KAMARULHAILI
President,
Malaysian Society for Cryptology Research

EDITORIAL PREFACE

Since the time of Julius Caesar and possibly up until the Greek era, cryptography (a word that is derived from the Greek term cryptos) has been an integral tool for organizations (and indeed for individuals too) to ensure information that is intended only for authorized recipients remain confidential only to this set of people. Cryptography had far reaching implications for organizations in the event information leakage occurred. Often referred to as the “last bastion of defence” after all other mechanisms had been overcome by an adversary, encrypted information would still remain useless to the attacker (i.e. that is, under the usual security assumptions). Nevertheless, this simple fact has remained oblivious to the practitioners of information security omitting cryptographic mechanism for data being transferred and also during storage.

Fast forward to World War 2, the war between cryptographic and cryptanalytic techniques. While the Germans were efficiently transferring information via the Enigma encryption machine, the Allies in Bletchley Park, England were busy intercepting these ciphered information being transmitted via telegraph by the Germans. Leading mathematicians, linguists, engineers etc. were all working to cryptanalyze these ciphers in the most information way. It is here that the first electrical machine (i.e. the bomba) was born and revolutionized computing. Post World War 2 saw the emergence of the computer. Every organization that had to process data had to acquire a computer so as not to be left behind by their competitor. The banking sector advanced on a global scale due to the invention of the computer. Techniques to secure information among the headquarters of these banks had to be developed. Encryption procedures using the same key (i.e. symmetric encryption) played this role in the early days. Then came the unthinkable problem computers were being deployed almost everywhere. How is it possible to deploy cryptographic keys in secure manner so that symmetric encryption could take place? Thus, leading to the so-called key distribution problem. It was not until 1975, when Diffie and Hellman provided us with a secure key exchange method and in 1976 when Rivest, Shamir and Adleman with the asymmetric encryption scheme (i.e. to encrypt using key e and decrypt using key d , where $e \neq d$). Since then, cryptographic procedures evolved, not only playing the role of ensuring confidentiality of data, but also to ensure integrity and authenticity of data. It is also able to ensure that non-repudiating of data does not occur.

Mechanisms to transfer and store data has changed of the centuries and more so every 5 years (in this modern age). Cryptography that has long existed before mechanisms changed from manual

telegraphic electrical electronic (WAN/LAN/internet) wired until wireless procedures, has to be properly deployed in order to maintain a high level of security confidence among the stakeholders of a certain organization. The concept of securing information via encryption procedures has to be properly understood in order to avoid a null intersection to occur between cryptography and computer security practitioners. This scenario would not be to the best interest for stakeholders. As a friendly reminder, this scenario could already been seen in other discipline of knowledge where the minuting (minute-ting) of knowledge has forced the original body of knowledge to look as though it is independent and disassociated. Ever since mass usage of computers became a reality, computer security issues have never been this complicated. However, as the human race advances so will ingenious ideas emerge to overcome challenges.

It is hoped that Cryptology2018 will not only provide a platform for every participant to exchange ideas in their respective fields, but also to exchange new ideas on a broader scale for the advancement of the field of cryptology and computer security. The organizing committee hopes every participant will have an enjoyable and beneficial conference.

Thank you.

**Editorial Board,
CRYPTOLOGY2018**

Table of Contents

Opening Remarks	iii
Welcoming Notes	v
Editorial Preface	vii
Board of Editors	xii
Blockchain	xiv
Kazue Sako	
Implementation of Lattice-based Cryptographic Primitives: Efficiency and Security	xv
Erdem Alkim	
BCTRU: A New Secure NTRUCrypt Public Key System Based on a Newly Multidimensional Algebra	1
Hassan R. Yassein & Nadia M. G. Al-Saidi	
On The Unsolvability of the Diophantine Equation $x_1^{a_1} + x_2^{a_2} + \dots + x_m^{a_m} = ny^b$ and Its Cryptographic Consequences	12
Abdulrahman Balfaqih & Hailiza Kamarulhaili	
Elliptic Net Scalar Multiplication using Generalized Equivalent Elliptic Divisibility Sequence	19
Norliana Muslim & Mohamad Rushdan Md Said	
Construction of Endomorphisms with J-Invariant 1728 for ISD Method	26
Siti Noor Farwina Mohamad Anwar Antony & Hailiza Kamarulhaili	
Garbage-man-in-the-middle (type 2) Attack on the Lucas Based El-Gamal Cryptosystem in the Elliptic Curve Group Over Finite Field	35
Izzatul Nabila bin Sarbini, Wong Tze Jin, Koo Lee Feng, Mohamed Othman, Mohd. Rushdan Md. Said & Yiu Pang Hung	

On the Underlying Hard Lattice Problems of GGH Encryption Scheme	42
Arif Mandangan, Hailiza Kamarulhaili & Muhammad Asyraf Asbullah	
Detecting General Algebraic Manipulation Attacks	51
Kim Ramchen	
Hierarchical Twin-Schnorr Identity-Based Identification Scheme	64
Apurva Kiran Vangujar, Ji-Jian Chin, Syh-Yuan Tan & Tiong-Sik Ng	
Revisiting the Invisibility of Yuen et al.'s Undeniable Signature Scheme	76
Jia-Ch'ng Loh, Swee-Huay Heng & Syh-Yuan Tan	
SSE Query Security	85
M.S. Mohamad, S.Y. Tan & J.J. Chin	
Enhanced AA_β Cryptosystem: The Design	94
Muhammad Asyraf Asbullah, Muhammad Rezal Kamel Ariffin & Zahari Mahad	
Extending Pollard Class of Factorable RSA Modulus	103
Amir Hamzah Abd Ghafar, Muhammad Rezal Kamel Ariffin & Muhammad Asyraf Asbullah	
A New Simultaneous Diophantine Attack Upon RSA Moduli $N = pq$	119
Saidu Isah Abubakar, Muhammad Rezal Kamel Ariffin & Muhammad Asyraf Asbullah	
New Vulnerability on $N = p^2q$ Using Good Approximation of $\phi(N)$	139
Normahirah Nek Abd Rahman, Muhammad Rezal Kamel Ariffin, Muhammad Asyraf Asbullah & Faridah Yunos	
Accelerating DGHV's Fully Homomorphic Encryption With GPU	151
Jia-Zheng Goey, Bok-Min Goi, Wai-Kong Lee & Raphael C.-W. Phan	
Evaluation Criteria on Random Ambience for Cryptographic Keys	160
Nur Azman Abu, Shekh Faisal Abdul Latip & Shahrin Sahib	

A Practical SCADA Testbed in Electrical Power System Environment for Cyber-security Exercises

176

Norziana Jamil, Qais Qassim, Maslina Daud, Izham Zainal Abidin, Norhamadi Ja'ffar
& Wan Azlan Wan Kamarulzaman

Board of Editors

International Program

Abderrahmane Nitaj

Committee

Rennato Renner

Amr M. Youssef

Yanbin Pan

Keith Martin

Kaoru Kurosawa

Lamberto Rondoni

Bharathwaj Muthuswamy

Kamel Ariffin Mohd Atan

Shahrin Sahib

Mohamed Ridza Wahiddin

Ahmad Izani Md. Ismail

Mohd Salmi Md Noorani

Solahuddin Shamsuddin

Maslina Daud

Sazali Sukardi

Executive Editor

Muhammad Rezal Kamel Ariffin

Technical Editors

Goi Bok Min

Hailiza Kamarulhaili

Heng Swee Huay

Moesfa Soeheila Mohamad

Mohamad Rushdan Md. Said

Muhammad Rezal Kamel Ariffin

Committee Members

Amir Hamzah Abd Ghafar

Aniza Abdul Ghani

Muhammad Asyraf Asbullah

Nor Azlida Aminudin

Nurul Nur Hanisah Adenan

Wan Nur Aqlili Wan Mohd Ruzai

Wan Zariman Omar

Zahari Mahad

Cover Design

Zahari Mahad

Blockchain

Kazue Sako

*Security Research Laboratories, NEC Corporation,
Japan*

k-sako@ab.jp.nec.com

ABSTRACT

Blockchain is a technology used in a cryptocurrency called Bitcoin. Now the technology is being considered to be applied to other usecases, with various modifications. In the presentation, we will discuss how blockchain works in Bitcoin, what are its cryptographic background, what are the variations of the technology, and what are some suitable/not suitable use cases.

Implementation of Lattice-based Cryptographic Primitives: Efficiency and Security

Erdem Alkim

*Ondokuz Mayıs University, Samsun,
Turkey*

erdemalkim@gmail.com

ABSTRACT

In this workshop, my aim is to give a point of view on the design principles of post-quantum cryptographic schemes. This will mainly include recent schemes submitted NIST Post-Quantum Cryptography Project. Since there is a great interest for the implementation of lattice-based cryptographic schemes, the focus will be given to lattice-based cryptography. The organization of this workshop is as follows:

- Introduction to post-quantum cryptography, Overview of the NIST Post-Quantum Cryptography Project submissions with focusing on similarities and differences of the submissions.
- Lattice-based primitives, Google's post-quantum experiment, pros/cons with comparing nowadays cryptosystems as well as post-quantum alternatives. The talk will focus on three lattice based submissions, NewHope, FrodoKEM, and qTesla. I will start with explaining what are the differences from nowadays systems, and how efficient they are. Then I will compare these schemes with other NIST submissions.
- Efficiency, security and size issues of building blocks in lattice-based schemes and how they can be solved. In this talk I will first separate implementations in to building blocks and explain them separately.

Then I will give overview of the implementation issues, and attacks. Finally I explain how we solved these during the submission process.

BCTRU: A New Secure NTRUCrypt Public Key System Based on a Newly Multidimensional Algebra

Hassan R. Yassein ^{*1} and **Nadia M. G. Al-Saidi** ²

¹*Department of Mathematics, College of Education, University of Al-Qadisiyah, Iraq*

²*Department of Applied Sciences, University of Technology, Iraq*

E-mail: hassan.yaseen@qu.edu.iq

**Corresponding author*

ABSTRACT

BCTRU is a newly generated multi-dimensional NTRU like public key cryptosystem. It is based on a newborn algebraic structure utilized instead of the classical NTRU-polynomial ring called bi-cartesian algebra, which is commutative and associative. The probability of successful of BCTRU decryption is discussed, it has good resistant against attacks, such as alternate key and brute force attacks. The performance analysis in terms of key and message security is demonstrated.

Keywords: NTRU, BCTRU, bi-cartesian algebra, security analysis.

1 INTRODUCTION

With the rapid advancement of our reset technological world, most of the important information which transferred through public channels required high-level protection mechanism. Cryptography is one of such mechanism used to provides security of information in public networks.

Most of the modern cryptographic techniques are best on arithmetic operations that defined on commutative algebraic structure which is considered nowadays as weak due to the fast increasing in computing process of newly computer devices. An alternate fast public key system is a challenge and of great demand. In 1996 at Crypto96 conference, three mathematicians researchers (Hoffstein et al., 1998) introduced as to a new filed of research called non commutative algebraic cryptography through introducing of NTRU (Number Theory Research Unit) cryptosystem.

They aimed to develop the cryptographic technique based on a non-commutative algebraic structures after demonstrating of NTRU as a secure protocol. It was generalized by many researchers through developing of its algebraic structure. Some of those developed protocols based

on different Euclidean ring free modules and algebras beyond \mathbb{Z} are as follows: Basic collection of objects used by the NTRU public key cryptosystem occurs in a truncated polynomial ring of degree $N - 1$ with integer coefficients that belong to $\mathbb{Z}[x]/(x^N - 1)$. NTRU is the first public key cryptosystem that does not depend on factorization and discrete log problems. Compared with the RSA (Rivest et al., 1978) and ECC (Schoof, 1985) cryptosystems, NTRU is faster and exhibits significantly smaller keys (Rivest et al., 1978, Schoof, 1985). Based on polynomial ring on $F_2[x]$ is proposed by Gaborit et al. (2002). They constructed a CTRU which is a NTRU variant cryptosystem. Matrices of polynomials of size $k \times k$ in $\mathbb{Z}[x]/(x^N - 1)$ is proposed by Coglianesi and Goi (2005). This NTRU analog is called MaTRU.

Suri and Puri (2007) presented the concept of crossbred technology using symmetric key as a stream cipher for encryption and decryption. NTRU public key cryptosystem was used in sending the secret key. Accordingly, the studies doubled the security, thereby avoiding brute force attacks because the attacker needs first to find the secret key that was encrypted through public key cryptography. In the same year, Malekian and Zakerolhosseini (2010) used the actual performance results of NTRU with respect to current asymmetrical cryptosystems.

Atici et al. (2008) presented a low-power and compact NTRU design that was suitable for security applications such as RFID and sensor nodes. Their design involved two architectures, namely, one that can perform both encryption and decryption and another for encryption only. The researchers compared the design with the original NTRU and found that the new design saves a factor of more than two. This design improved the speed of NTRU.

Other NTRU variant cryptosystem was proposed by Malekian et al., in 2009 and 2010 respectively (Malekian and Zakerolhosseini, 2010, Malekian et al., 2009). They rely on designing their cryptosystems on Quaternion algebra and Octonion algebra respectively. In the same period, Vats (2009) introduced NTRU variant which is operated in the non-commutative ring $M = M_k(\mathbb{Z})[x]/(x^N - I_{k \times k})$, where M is a matrix ring of the $k \times k$ matrices of polynomials in $\mathbb{Z}[x]/(x^N - 1)$.

Another generalized framework is proposed by Pan and Deng (2011). They used hiding the trapdoor technique, which is led to designing of a new lattice-based cryptosystem, which helps to solve the closest vector problem. Kumar et al. introduced complex problems into the existing implementation; efficiency could be achieved through reduced implementation of polynomial multiplication of inverse computation.

A new ring of cubic root of unity called Eisensteinian ring $\mathbb{Z}[\omega]$ is used to construct a new framework to NTRU called ETRU which is proposed by Jarvis and Nevins (2015). CQTRU is another NTRU variant cryptosystem proposed by Alsaïdi et al. (2015). In (Thakur and Tripathi, 2016) Thakur and Tripathi utilized the rational field to construct a ring with polynomials of one variable over this field to be used in introducing of new NTRU alternative cryptosystem called BTRU. After that Yassein and Al-Saidi constructed several high dimensional algebra as and utilized them in proposing of different NTRU analog cryptosystems presented in Al-Saidi and Yassein (2017), Alsaïdi and Yassein (2016), Yassein and Al-Saidi (2016, 2017).

In this paper, a new algebra is constructed, we called Bi-Cartesian algebra. It is used to construct BCTRUE which is a new NTRU like cryptosystem. It is also multidimensional public

key system, because it is produced two public key, which resulted in increasing the security of proposed cryptosystem that based on comparing to its variant with the same structures BCTRU has the ability to encrypt four messages sent by asingle origin or four independent messages sent by four different sources. With this important property, the proposed system will be considered as an ultimate fast new public key cryptosystem to be as a best fit in many applications with limited resources for examples smart cards, cellular phones and many others. This study is organized as follows. Section 2 we introduce new algebra which called bi-cartesian algebra. BCTRU cryptosystem described in the section 3. The probability of successful decryption of BCTRU is discussed in section 4. The security analysis of BCTRU is dedicated in section 5. Discussions are provided in section 6. Finally, section 7 is for the most important conclusions.

2 BICARTESIAN ALGEBRA

The Bi-Cartesian algebra is defined by utilizing the same parameters N, p and q used in NTRU, taking in our consideration that the integer constants d_f, d_g, d_m and d_ϕ should be less than N . Also, the truncated polynomial ring is defined as $K = Z[x]/(x^N - 1)$ with degree $N - 1$. We define a new algebra as follows:

A bi-cartesian algebra is introduced in this section as a vector space of dimension two over field F . Let $BC = \{(a, b)(1, 1) + (c, d)(k, 1) | a, b, c, d \in F\}$, $k^2 = 1$ where $\{(1,1), (k,1)\}$ forms the basis of this algebra. The operation on this space is defined as follows: Let $x, y \in BC$, such that $x = (a, b)(1, 1) + (c, d)(k, 1)$ and $y = (a_1, b_1)(1, 1) + (c_1, d_1)(k, 1)$, the addition is then defined by

$$x + y = (a + a_1, b + b_1)(1, 1) + (c + c_1, d + d_1)(k, 1).$$

The multiplication $x * y$ can be determined using Table 1.

*	(1,1)	(1,k)
(1,1)	(1,1)	(1,k)
(1,k)	(1,k)	(1,1)

Table 1: Multiplication operation.

For any scalar α , the scalar multiplication is defined by $\alpha x = (\alpha a, \alpha b)$. It is clear that the multiplication is commutative and associative. We now consider the truncated polynomial rings $K(x) = (Z/Z)[x]/(x^N - 1)$, $K_p(x) = (Z/pZ)[x]/(x^N - 1)$ and $K_q(x) = (Z/qZ)[x]/(x^N - 1)$ and define three bi-cartesian algebra as ψ, ψ_p and ψ_q as follows:

$$\begin{aligned} \psi &= \{(f_0, f_1)(1, 1) + (f_2, f_3)(k, 1) | f_0, f_1, f_2, f_3 \in K\} \\ \psi_p &= \{(f_0, f_1)(1, 1) + (f_2, f_3)(k, 1) | f_0, f_1, f_2, f_3 \in K_p\} \\ \psi_q &= \{(f_0, f_1)(1, 1) + (f_2, f_3)(k, 1) | f_0, f_1, f_2, f_3 \in K_q\}. \end{aligned}$$

The parameters N, p , and q are fixed similar to the NTRU parameters. The constants d_f, d_g, d_ϕ and d_m are defined in a similar role as in NTRU.

Notation	Definition
L_f	$\{(f_0, f_1)(1, 1) + (f_2, f_3)(k, 1) \in K \mid f_i \text{ has } d_f \text{ coefficients equal to } +1, (d_f - 1) \text{ coefficients equal to } -1, \text{ and the rest are } 0\}$
L_g	$\{(g_0, g_1)(1, 1) + (g_2, g_3)(k, 1) \in K \mid g_i \text{ has } d_g \text{ coefficients equal to } +1, d_g \text{ coefficients equal to } -1, \text{ and the rest are } 0\}$
L_ϕ	$\{(\phi_0, \phi_1)(1, 1) + (\phi_2, \phi_3)(k, 1) \in K \mid \phi_i \text{ has } d_\phi \text{ coefficients equal to } +1, d_\phi \text{ coefficients equal to } -1, \text{ and the rest are } 0\}$
L_m	$\{(m_0, m_1)(1, 1) + (m_2, m_3)(k, 1) \in K \mid \text{coefficients of } m_i \text{ are chosen modulo } p \text{ between } -p/2 \text{ and } p/2\}$

Table 2: The subsets of BCTRU.

Let F and $G \in \psi_p$ or ψ_q , such that:

$$F = (f_0, f_1)(1, 1) + (f_2, f_3)(k, 1)$$

$$G = (g_0, g_1)(1, 1) + (g_2, g_3)(k, 1)$$

where f_0, f_1, f_2, f_3 and $g_0, g_1, g_2, g_3 \in \psi_p$ or ψ_q .

The addition of F_1 and F_2 is performed by adding the corresponding coefficients mod p or mod q , such that

$$F + G = (f_0 + g_0, f_1 + g_1)(1, 1) + (f_2 + g_2, f_3 + g_3)(k, 1)$$

the multiplication of F and G is defined as follows:

$$F * G = (f_0g_0 + f_2g_2, f_1g_1 + f_3g_3)(1, 1) + (f_0g_2 + f_2g_0, f_1g_3 + f_3g_1)(k, 1)$$

The multiplicative inverse of any non zero element F in BC is given by:

$$F^{-1} = ((f_0^2 - f_2^2)^{-1} f_0, (f_2^2 - f_0^2)^{-1} f_2)(1, 1) \\ + ((f_1^2 - f_3^2)^{-1} f_1, (f_3^2 - f_1^2)^{-1} f_3)(k, 1)$$

3 BCTRU CRYPTOSYSTEM

Similar to NTRU, The BCTRU cryptosystem is constructed based on the same parameters, a prime number N , and two relatively prime numbers p , and q , in which q is much larger than p . The four main subsets that NTRU and any NTRU variant cryptosystem depends on are defined as:

Definition 1: The subsets L_f, L_g, L_ϕ and $L_m \subset \Psi$ are called the subsets of BCTRU, and these subsets are defined as follows: d_f, d_g and d_ϕ are also constant parameters similar to those defined in NTRU. The main cryptosystem parts of BCTRU are:

A. KEY GENERATION

In this phase, the sender is able to generate the public key by choosing F and U randomly from the set L_f and G randomly from the set L_g such that,

$$F = (f_0, f_1)(1, 1) + (f_2, f_3)(k, 1), G = (g_0, g_1)(1, 1) + (g_2, g_3)(k, 1)$$

and $U = (u_0, u_1)(1, 1) + (u_2, u_3)(k, 1)$

By considering that F must have multiplicative inverse modulo p and q referred to as F_p^{-1}, F_q^{-1} respectively, and U must have multiplicative inverse modulo p referred to as U_p^{-1} , the public keys are given by:

$$H = F_q^{-1}G \text{ mod } (q), K = UF_q^{-1} \text{ mod } (q),$$

where F, G and U are the private keys. BCTRU key generation needs sixteen convolution multiplications and eight polynomial additions.

B. ENCRYPTION

Before performing of the encryption process, the message m should be expressed by the elements of the bi-cartesian algebra as:

$$M = (m_0, m_1)(1, 1) + (m_2, m_3)(k, 1).$$

We choose $\phi \in L_\phi$ which is called the blinding value, to encrypt the message $m \in L_m$:

$$E = pH * \phi + M * K \text{ (mod } q)$$

BCTRU encryption needs sixteen convolution multiplications and eight polynomial additions. Therefore, the speed of the key generation is faster than that of encryption.

C. DECRYPTION

After receiving E , we left and right multiply it by F , then

$$\begin{aligned} A &= F * E * F(\text{mod } q) = F * (pH * \phi + M * K) * F(\text{mod } q) \\ &= pF * H * \phi * F + F * M * K * F(\text{mod } q) \\ &= pF * F_q^{-1} * G * \phi * F + F * M * U * F_q^{-1} * F(\text{mod } q) \\ &= pG * \phi * F + F * M * U(\text{mod } q) \end{aligned}$$

Let $B = A(\text{mod } p) = pG * \phi * F + F * M * U(\text{mod } p)$.

Since the first term is equal to zero modulo p (because it contains p), then

$$B = F * M * U(\text{mod } p), \quad F_p^{-1} * B * U_p^{-1} = M(\text{mod } p)$$

and the resulting coefficients are adjusted to lie in the interval $[-p/2, p/2]$.

BCTRU decryption needs thirty two convolution multiplications and twelve polynomial additions. As a result, the speed of encryption is more than twice as fast as that of decryption.

4 PROBABILITY OF SUCCESSFUL DECRYPTION

The successful decryption of BCTRU depends on the probability of all coefficients of $A = pG * \phi * F + F * M * U$ belongs to the interval $\left[\frac{-q+1}{2}, \frac{q-1}{2} \right]$, which are calculated in the following theorem:

Theorem 4.1. $Pr \left(|A_{j,\tau}| \leq \frac{q-1}{2} \right) = 2 \mathcal{N} \left(\frac{q-1}{2\sqrt{\frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}}} \right)$, where \mathcal{N} denotes the normal distribution, and $j, \tau = 0, 1, 2, 3$.

Proof. To compute this probability, A should be written in a BCTRU form, where

$A = pG * \phi * F + F * M * U$, such that,

$A = (A_0, A_1)(1, 1) + (A_2, A_3)(k, 1)$, A_0, A_1, A_2, A_3 which are polynomials of degree N where

$$\begin{aligned} A_0 &= p(g_0\phi_0f_0 + g_0\phi_2f_2 + g_2\phi_0f_2 + g_2\phi_2f_0) + (f_0m_0u_0 + f_0m_2u_2 + \\ &\quad f_2m_0u_2 + f_2m_2u_0) = [A_{0,0}, A_{0,1}, A_{0,2}, \dots, A_{0,N-1}], \\ A_1 &= p(g_1\phi_1f_1 + g_1\phi_3f_3 + g_3\phi_1f_3 + g_3\phi_3f_1) + (f_1m_1u_1 + f_1m_3u_3 + \\ &\quad f_3m_1u_3 + f_3m_3u_1) = [A_{1,0}, A_{1,1}, A_{1,2}, \dots, A_{1,N-1}], \\ A_2 &= p(g_0\phi_0f_2 + g_0\phi_2f_0 + g_2\phi_2f_0 + g_2\phi_2f_2) + (f_0m_0u_0 + f_0m_2u_0 + \\ &\quad f_2m_2u_0 + f_2m_2u_2) = [A_{2,0}, A_{2,1}, A_{2,2}, \dots, A_{2,N-1}], \\ A_3 &= p(g_1\phi_1f_3 + g_1\phi_3f_1 + g_3\phi_1f_1 + g_3\phi_3f_3) + (f_1m_1u_1 + f_1m_3u_1 + \\ &\quad f_3m_1u_1 + f_3m_3u_3) = [A_{3,0}, A_{3,1}, A_{3,2}, \dots, A_{3,N-1}], \end{aligned}$$

Based on the definition of L_f, L_m , and L_ϕ , the following is obtained:

$$\begin{aligned} f_j &= [f_{j,0}, f_{j,1}, f_{j,2}, \dots, f_{j,N-1}] \\ g_j &= [g_{j,0}, g_{j,1}, g_{j,2}, \dots, g_{j,N-1}] \\ \phi_j &= [\phi_{j,0}, \phi_{j,1}, \phi_{j,2}, \dots, \phi_{j,N-1}] \\ Pr(f_{j,k} = 1) &= \frac{d_f}{N}, \text{ and } Pr(f_{j,k} = -1) = \frac{d_f-1}{N} \approx \frac{d_f}{N} \\ Pr(f_{j,k} = 0) &= 1 - \frac{2d_f}{N} \\ Pr(u_{j,k} = 1) &= \frac{d_u}{N}, \text{ and } Pr(u_{j,k} = -1) = \frac{d_u-1}{N} \approx \frac{d_u}{N} \\ Pr(u_{j,k} = 0) &= 1 - \frac{2d_u}{N} \end{aligned}$$

$$Pr(g_{j,k} = 1) = Pr(g_{j,k} = -1) = \frac{d_g}{N}, Pr(g_{j,k} = 0) = 1 - \frac{2d_g}{N},$$

$$Pr(\phi_{j,k} = 1) = Pr(\phi_{j,k} = -1) = \frac{d_\phi}{N}, Pr(\phi_{j,k} = 0) = 1 - \frac{2d_\phi}{N},$$

$$Pr(m_{j,k} = \gamma) = \frac{1}{p} \quad \gamma \in \left[-\frac{p}{2}, \frac{p}{2}\right], \quad j, k = 0, 1, 2, 3.$$

We assume that all $f_{j,\alpha}$, $g_{k,\beta}$ and $\phi_{t,\lambda}$ are pairwise independent random variables.

For $\alpha, \beta, \lambda = 0, 1, \dots, N-1$,

$$\gamma = -\frac{p-1}{2}, \dots, \frac{p-1}{2}, \quad \text{and } j, k, t = 0, 1, 2, 3.$$

Therefore,

$$\begin{aligned} Pr(g_{j,\alpha} \cdot \phi_{k,\beta} \cdot f_{t,\lambda} = \mp 1) &= \frac{8d_g d_\phi d_f}{N^3}, \\ Pr(g_{j,\alpha} \cdot \phi_{k,\beta} \cdot f_{t,\lambda} = 0) &= 1 - \frac{8d_g d_\phi d_f}{N^3}, \\ Pr(f_{j,\alpha} \cdot m_{k,\beta} \cdot u_{t,\lambda} = \gamma) &= \frac{4d_f d_u}{pN^2}. \end{aligned}$$

Based on the preceding assumptions and after a number of computations, the following is obtained:

$$\begin{aligned} Var(g_{j,\alpha} \cdot \phi_{k,\beta} \cdot f_{t,\lambda})_y &= Var\left(\sum_{\alpha+\beta+\lambda=y \pmod{N}} \sum_{\alpha+\beta+\lambda=y \pmod{N}} g_{j,\alpha} \cdot \phi_{k,\beta} \cdot f_{t,\lambda}\right) \\ &= \frac{8d_g d_\phi d_f}{N}, \\ Var(f_{j,\alpha} \cdot m_{k,\beta} \cdot u_{t,\lambda})_y &= Var\left(\sum_{\alpha+\beta+\lambda=y \pmod{N}} \sum_{\alpha+\beta+\lambda=y \pmod{N}} f_{j,\alpha} \cdot m_{k,\beta} \cdot u_{t,\lambda}\right) \\ &= \frac{d_f d_u (p-1)(p+1)}{3}, \\ Var(A_0, \tau) &= \frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}. \end{aligned}$$

Moreover, $Var(A_1, \tau) = Var(A_2, \tau) = Var(A_3, \tau)$ are equal to $\frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}$ obtained in a similar manner when the probabilities of all coefficients $A_{0,i}$, $A_{1,i}$, $A_{2,i}$, $A_{3,i}$ belong to $[-\frac{q+1}{2}, \frac{q+1}{2}]$. Therefore, successful decryption is performed. Thus,

$$Pr\left(|A_{i,\tau}| \leq \frac{q-1}{2}\right) = Pr\left(-\frac{q-1}{2} \leq A_{j,\tau} \leq \frac{q-1}{2}\right) = 2\mathcal{N}\left(\frac{q-1}{2\sigma}\right),$$

$$\text{where } \sigma = \sqrt{\frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}},$$

$$i = 0, 1, 2, 3 \text{ and } \tau = 0, 1, \dots, N-1$$

□

Corollary 4.1. 1. The probability for any of the messages M_0 , M_1 , M_2 and M_3 to be successfully decrypted is

$$\left(2\mathcal{N}\left(\frac{q-1}{2\sqrt{\frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}}}\right) - 1\right)^N$$

2. The probability for both of the messages M_0 , M_1 , M_2 and M_3 to be successfully decrypted is

$$(2\mathcal{N} \left(\frac{q-1}{2 \sqrt{\frac{32p^2 d_g d_\phi d_f}{N} + \frac{32d_f d_u (p-1)(p+1)}{3}}} \right) - 1)^{4N}.$$

5 SECURITY ANALYSIS

The majority of some attacks to threaten BCTRU has been investigated to prove its security. The most powerful for such type of cryptosystem that based on polynomial algebra is the lattice attack, in which the shortest vector in the lattice vector space of the proposed cryptosystem represents the private key, which can be found by approximate solution for the corresponding vector matrix. Some of these attacks are discussed as follows.

A. ALTERNATE KEY ATTACK

The main objective of this attacker is to find the alternate private keys in order to decrypt the received encrypted media. Therefore, the attacker task is an attempt to find the following alternate keys:

$$\acute{F} = (\acute{f}_0, \acute{f}_1)(1, 1) + (\acute{f}_2, \acute{f}_3)(k, 1)$$

$$\acute{G} = (\acute{g}_0, \acute{g}_1)(1, 1) + (\acute{g}_2, \acute{g}_3)(k, 1)$$

$$\acute{U} = (\acute{u}_0, \acute{u}_1)(1, 1) + (\acute{u}_2, \acute{u}_3)(k, 1)$$

to F , G and U respectively, such that \acute{F} must have multiplicative inverse modulo p and q also, \acute{U} must have multiplicative inverse modulo p .

Thus, an attacker to BCTRU needs twelve polynomials $\acute{f}_0, \acute{f}_1, \acute{f}_2, \acute{f}_3, \acute{g}_0, \acute{g}_1, \acute{g}_2, \acute{g}_3, \acute{u}_0, \acute{u}_1, \acute{u}_2, \acute{u}_3$, with the same properties of polynomials $f_0, f_1, f_2, f_3, g_0, g_1, g_2, g_3, u_0, u_1, u_2, u_3$ respectively. However, an attacker to NTRU only needs extra attempts to find the private key (in this case twelve) than those to decrypt NTRU, which needs only one polynomial in L_f with the same properties of the private key.

B. BRUTE FORCE ATTACK

An attacker to BCTRU that knows the public parameters as well as the public key

$$H = F_q^{-1}G \text{ mod } (q), K = UF_q^{-1} \text{ mod } (q),$$

which are equivalent to the following hidden equations:

$$FH = G \text{ mod } (q), \tag{1}$$

$$KF = U \text{ mod } (q), \quad (2)$$

We tested all the polynomials $F \in L_f$ (hard mathematical problem) and determine if Eqs. (1) and (2) turn into bi-cartesian algebra with small coefficients until the private key is found. The size of the subset L_f is calculated as follows:

$$|L_f| = \left(\frac{N!}{(d_f!)^2(N - 2d_f)!} \right)^4.$$

Accordingly, the number of all attempts to find the private keys F, G and U is equal to

$$\frac{N!^{12}}{(d_f!d_g!d_u!)^2((N - 2d_f)!(N - 2d_g)!(N - 2d_u)!)^4}.$$

6 DISCUSSION

To discuss the lattice attack, the idea lies in finding the length of the shortest nonzero vector in the given vector space. It is calculated by the Gaussian heuristics as $\delta(\mathcal{L}_{\text{BCTRU}}^H) = \sqrt{\frac{2N}{\pi e}} \sqrt{q} \approx 0.48\sqrt{Nq}$.

Also $\frac{\|(f_0, f_1, f_2, f_3, g_0, g_1, g_2, g_3)\|}{\delta} = \frac{2.309\sqrt{N}}{0.48\sqrt{Nq}} \approx \frac{4.8104}{\sqrt{q}}$, However, it gives an expected vector longer than the vector proposed by $\mathcal{L}_{\text{BITRU}}^H$. Moreover, $\mathcal{L}_{\text{BITRU}}^H$ dimension is doubling of $\mathcal{L}_{\text{NTRU}}$ dimension with the same N . The calculation of the shortest nonzero vector length is accomplished using the same way, such as $\delta(\mathcal{L}_{\text{BCTRU}}^K) \approx 0.48\sqrt{Nq}$ and $\frac{\|(f_0, f_1, f_2, f_3, u_0, u_1, u_2, u_3)\|}{\delta} \approx \frac{4.8104}{\sqrt{q}}$, which led us to conclude that BCTRU shows more resistance to lattice, the most serious threaten upon NTRU like cryptosystem. Since BCTRU utilized two public keys H, K with twelve polynomial private keys $f_0, f_1, f_2, f_3, g_0, g_1, g_2, g_3, u_0, u_1, u_2$, and u_3 , therefore, its security is twelve times than that of NTRU.

7 CONCLUSIONS

In this paper, we introduced BCTRU public key cryptosystem that depends on new generated bi-cartesian algebra to enhance the security through discussing of some attacks. We demonstrated that, the security of BCTRU is four times mor than NTRU, and it shows certain resistance against attacks. Also, BCTRU has the ability to encrypt four messages of length N in each round, which granted it a good speed facility that is important for many application.

REFERENCES

Al-Saidi, N. M. and Yassein, H. R. (2017). A New Alternative to NTRU cryptosystem based on Highly Dimensional Algebra with Dense Lattice Structure. *MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES*, 11:29–43.

- Alsaidi, N., Saed, M., Sadiq, A., and Majeed, A. A. (2015). An improved ntru cryptosystem via commutative quaternions algebra. In *Proceedings of the International Conference on Security and Management (SAM)*, page 198. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- Alsaidi, N. M. and Yassein, H. R. (2016). BITRU: Binary Version of the NTRU Public Key Cryptosystem via Binary Algebra. *International Journal of Advanced Computer Science and Applications*, 7(11):1–6.
- Atici, A. C., Batina, L., Fan, J., Verbauwhede, I., and Yalcin, S. B. O. (2008). Low-cost implementations of NTRU for pervasive security. In *Application-Specific Systems, Architectures and Processors, 2008. ASAP 2008. International Conference on*, pages 79–84. IEEE.
- Coglianesi, M. and Goi, B.-M. (2005). MaTRU: A new NTRU-based cryptosystem. In *International Conference on Cryptology in India*, pages 232–243. Springer.
- Gaborit, P., Ohler, J., and Solé, P. (2002). *CTRU, a polynomial analogue of NTRU*. PhD thesis, INRIA.
- Hoffstein, J., Pipher, J., and Silverman, J. H. (1998). Ntru: A ring-based public key cryptosystem. In *International Algorithmic Number Theory Symposium*, pages 267–288. Springer.
- Jarvis, K. and Nevins, M. (2015). ETRU: NTRU over the Eisenstein Integers. *Designs, Codes and Cryptography*, 74(1):219–242.
- Kumar, S., Pal, S. K., et al. An Improved Post-Quantum Cryptographic Scheme Based on NTRU. *International Journal of Computer Applications Technology and Research*, 2(4):499–meta.
- Malekian, E. and Zakerolhosseini, A. (2010). OTRU: A non-associative and high speed public key cryptosystem. In *Computer Architecture and Digital Systems (CADS), 2010 15th CSI International Symposium on*, pages 83–90. IEEE.
- Malekian, E., Zakerolhosseini, A., and Mashatan, A. (2009). QTRU: A Lattice Attack Resistant Version of NTRU PKCS Based on Quaternion Algebra. *preprint, Available from the Cryptology ePrint Archive: <http://eprint.iacr.org/2009/386.pdf>*.
- Pan, Y. and Deng, Y. (2011). A general NTRU-Like framework for constructing lattice-based public-key cryptosystems. In *International Workshop on Information Security Applications*, pages 109–120. Springer.
- Rivest, R. L., Shamir, A., and Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126.
- Schoof, R. (1985). Elliptic curves over finite fields and the computation of square roots mod . *Mathematics of computation*, 44(170):483–494.
- Suri, P. and Puri, P. (2007). Application of LFSR with NTRU Algorithm. In *Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications*, pages 369–373. Springer.
- Thakur, K. and Tripathi, B. (2016). BTRU, A Rational Polynomial Analogue of NTRU Cryptosystem. *International Journal of Computer Applications, Foundation of Computer Science (FCS), NY, USA*, 145(12).

- Vats, N. (2009). NNRU, a noncommutative analogue of NTRU. *arXiv preprint arXiv:0902.1891*.
- Yassein, H. R. and Al-Saidi, N. M. (2016). HXDTRU Cryptosystem Based On Hexadecnicion Algebra. In *Proceeding of the 5th International Cryptology and Information Security Conference, Kota Kinabalu, Malaysia*.
- Yassein, H. R. and Al-Saidi, N. M. (2017). A comparative performance analysis of NTRU and its variant cryptosystems. In *Current Research in Computer Science and Information Technology (ICCIT), 2017 International Conference on*, pages 115–120. IEEE.

On The Unsolvability of the Diophantine Equation

$$x_1^{a_1} + x_2^{a_2} + \cdots + x_m^{a_m} = ny^b \text{ and Its Cryptographic Consequences}$$

Abdulrahman Balfaqih^{*1} and Hailiza Kamarulhaili²

¹Department of Mathematics, Faculty of Education-Seyoun, Hadhramout University
P.O.BOX:(50511-50512), Al-Mukalla, Yemen

²School of Mathematical Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia

E-mail: abdulrahman-balfaqih@student.usm.my

^{*}Corresponding author

ABSTRACT

In this paper, we consider the Diophantine equation $x_1^{a_1} + x_2^{a_2} + \cdots + x_m^{a_m} = ny^b$, where $a_i, (i = 1, 2, 3 \cdots, m), b, n, m$ are positive integers. If p is arbitrary prime such that $p - 1 > m, a_i \equiv 0 \pmod{p - 1}, b \equiv 0 \pmod{p - 1}$ and $n \equiv p - 1 \pmod{p}$, we show that, this equation has no solution in positive integers x_1, x_2, \dots, x_m and y .

Keywords: Diophantine Equation, Infinite Descent, Exponential Diophantine Equation.

1 INTRODUCTION

Diophantine equations has attracted mathematicians for more than 3 centuries. Tracking back a history on Fermat's Last Theorem, indicated in the following equation

$$x^n + y^n = z^n \tag{1}$$

where it states that if n is an integer greater than 2, equation (1) has no solution in positive integers x, y and z . This equation has given rise to many variations of similar forms and of course with restrictions and conditions. (Taylor and Wiles (1995), Wiles (1995)), established the modularity of a large class of curves in 1995 that has led to the accomplishment of this Fermat's problem (Narkiewicz (2012), Cornell et al. (1997)). Apart from the ideas and results on solving equation (1), many has been established on equations related to it. With different exponent on the right hand side of equation (1), Wong and Kamarulhaili (2016), partially solved the Diophantine equation

$$x^4 + y^4 = p^k z^7, \tag{2}$$

where p is a prime and k is a positive integer for when $x = y$. Wong and Kamarulhaili (2017), studied a more general form of (2) indicated as the following equation,

$$x^a + y^a = p^k z^b \quad (3)$$

where p is a prime number, with $\gcd(a, b) = 1$ and $k, a, b \in \mathbb{Z}^+$. They solved this equation parametrically by considering different cases of x and y for when $x = y, x = -y$ and either x or y is zero (not both zero). They also considered for the case when, $|x| \neq |y|$ and both x and y nonzero, but in this case only partial solutions of (x, y, z) were given. The work by Wong and Kamarulhaili has triggered an idea to extend equation (3) to a summation form. In view of the summation form of (3), together with several results Bérczes et al. (2016) and Bérczes et al. (2018) indicated below, has culminated in our main result in this paper. Bérczes et al. (2016), provided all solutions of equation

$$S_k(x) = 1^k + 2^k + \dots + x^k = y^n \quad (4)$$

in positive integers x, k, y, n with $1 \leq x < 25$ and $n \geq 3$. Bérczes et al. (2018), gave upper bounds for n on the Diophantine equation

$$(x+1)^k + (x+2)^k + \dots + (2x)^k = y^n \quad (5)$$

and showed that for $2 \leq x \leq 13, k \geq 1, y \geq 2$ and $n \geq 3$ the equation (5) has no solutions. Putting all these works together, we give a slightly different form of equation, given as, $x_1^{a_1} + x_2^{a_2} + \dots + x_m^{a_m} = ny^b a_i, (i = 1, 2, 3, \dots, m), b, n, m$ positive integers and if p is arbitrary prime such that $p - 1 > m, a_i \equiv 0 \pmod{p-1}, b \equiv 0 \pmod{p-1}$ and $n \equiv p-1 \pmod{p}$, this equation has no solution in positive integers x_1, x_2, \dots, x_m and y . We also show the consequence of our results related to equation (4). This paper is organized as follows: In the next Section we show our main result and give the proof of our main result with give the consequences. In the third Section, we give some examples and finally Section 4 provides the conclusion.

2 MAIN RESULT

Our main result is shown in Theorem 2.1. Before we can provide the proof of the theorem in this Section, we start by showing the following lemma and its proof.

Lemma 2.1. *Suppose that p is arbitrary prime such that $p - 1 > m$ for m is positive integer. Then the Diophantine equation*

$$x_1^{p-1} + x_2^{p-1} + \dots + x_m^{p-1} = (pl - 1)y^{p-1} \quad (6)$$

Has no solution in positive integers x_1, x_2, \dots, x_m and y , where l is positive integer.

Proof. Assume that equation (6) has a solution in nonzero integers x_1, x_2, \dots, x_m and y satisfy the given equation. Now, we will consider two cases as follows:

First case If $\gcd(y, p) = 1$, then from the Fermat's little Theorem we obtain that $y^{p-1} \equiv 1 \pmod{p}$, thus

$$\sum_{i=1}^m x_i^{p-1} = x_1^{p-1} + x_2^{p-1} + \dots + x_m^{p-1} = (pl - 1)y^{p-1} \equiv pl - 1 \equiv p - 1 \pmod{p} \quad (7)$$

But this impossible because, $x_i^{p-1} \equiv 0 \text{ or } 1 \pmod{p}, \forall i, (i = 1, 2, \dots, m)$ and $p - 1 > m$.

Second case If $\gcd(y, p) = p$, then there exist an integer y_1 such that $y = py_1$, and thus we get

$$\sum_{i=1}^m x_i^{p-1} = x_1^{p-1} + x_2^{p-1} + \dots + x_m^{p-1} \equiv 0 \pmod{p} \quad (8)$$

This implies that, $x_i^{p-1} \equiv 0 \pmod{p}, (i = 1, 2, \dots, m)$. Therefore, there exists integers x'_i such that $x_1 = px'_1, x_2 = px'_2, \dots, x_m = px'_m$. This shows that, the original equation (6) can be written in the following form:

$$\sum_{i=1}^m (x'_i)^{p-1} = (px'_1)^{p-1} + (px'_2)^{p-1} + \dots + (px'_m)^{p-1} = (pl - 1)(py_1)^{p-1} \quad (9)$$

And this is equivalent to

$$(x'_1)^{p-1} + (x'_2)^{p-1} + \dots + (x'_m)^{p-1} = (pl - 1)(y_1)^{p-1} \quad (10)$$

Thus $(x'_1, x'_2, x'_3, \dots, x'_m, y_1)$ satisfy the given equation (10) and $y^{p-1} > y_1^{p-1}$. Similarly, we obtain the sequence

$$y^{p-1} > y_1^{p-1} > y_2^{p-1} > y_3^{p-1} > \dots \quad (11)$$

And from the Fermat's method of infinite descent (see Andreescu et al. (2010), AT&T Laboratories (2005), Mordell (1969) and Herman et al. (2000)) we get a contradiction and thus completes the proof. \square

Theorem 2.1. *Suppose that p is arbitrary prime such that:*

1. $p - 1 > m$
2. $a_i \equiv 0 \pmod{p - 1}$
3. $b \equiv 0 \pmod{p - 1}$
4. $n \equiv p - 1 \pmod{p}$

For $a_i, b, n, m \in \mathbb{Z}^+, (i = 1, 2, 3 \dots, m)$. Then, the Diophantine equation

$$x_1^{a_1} + x_2^{a_2} + \dots + x_m^{a_m} = ny^b \quad (12)$$

Has no solution in positive integers x_1, x_2, \dots, x_m and y .

Proof. We can write the equation (12) in the equivalent form

$$\sum_{i=1}^m x_i^{k_i(p-1)} = x_1^{k_1(p-1)} + x_2^{k_2(p-1)} + \dots + x_m^{k_m(p-1)} = (pl - 1)y^{k(p-1)} \quad (13)$$

where $k_i, l \in \mathbb{Z}^+$. We assume that the positive integers x_1, x_2, \dots, x_m and y satisfy the given equation (13). Then, the positive integers $x_1^{k_1}, x_2^{k_2}, \dots, x_m^{k_m}$ and y^k satisfy the equation (6), but from lemma 2.1, the Diophantine equation (6) has no solutions in positive integers. Therefore, the Diophantine equation (12) has no solutions in positive integers x_1, x_2, \dots, x_m and y . Where p is arbitrary prime such that $p - 1 > m, a_i \equiv 0(\text{mod } p - 1), b \equiv 0(\text{mod } p - 1)$, and $n \equiv p - 1(\text{mod } p)$. \square

Corollary 2.1. *Every integer n such that $n \equiv p - 1(\text{mod } p)$, for p an arbitrary prime greater than 2, Then n cannot be the sum of primes as the following*

$$p_1^{(p-1)k_1} + p_2^{(p-1)k_2} + p_3^{(p-1)k_3} + \dots + p_m^{(p-1)k_m} = n \quad (14)$$

Where, $m < p - 1, p_i$ are primes for $i = 1, 2, 3 \dots, m$ and $k_i, m \in \mathbb{Z}^+$.

Proof. Let $p_1, p_2, p_3, \dots, p_m$ are primes such that equation (14) is satisfied. As, $n \equiv p - 1(\text{mod } p)$, Then, $(p_1, p_2, p_3, \dots, p_m, 1)$ is solution of diophantine equation (12). But, this contradicts theorem 2.1. Therefore, n cannot be the sum of primes indicated by equation (14) \square

Corollary 2.2. *Suppose that p is arbitrary prime such that:*

1. $p - 1 > x$
2. $k \equiv 0(\text{mod } p - 1)$
3. $n \equiv 0(\text{mod } p - 1)$
4. $b \equiv p - 1(\text{mod } p)$

For $k, b, n, x \in \mathbb{Z}^+$. Then, the Diophantine equation

$$S_k(x) = 1^k + 2^k + \dots + x^k = by^n \quad (15)$$

Has no solution.

Proof. Since, p is arbitrary prime such that:

1. $p - 1 > x$
2. $k \equiv 0(\text{mod } p - 1)$
3. $n \equiv 0(\text{mod } p - 1)$
4. $b \equiv p - 1(\text{mod } p)$

For $k, b, n, x \in \mathbb{Z}^+$, Then by theorem 2.1, the Diophantine equation (15) has no solution. \square

3 EXAMPLES

Example 3.1. Every integer n such that $n \equiv 12 \pmod{13}$, cannot be the sum of primes for the following equations.

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36} + p_7^{48} + p_8^{48} + p_9^{60} + p_{10}^{60} + p_{11}^{72}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36} + p_7^{48} + p_8^{48} + p_9^{60} + p_{10}^{60}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36} + p_7^{48} + p_8^{48} + p_9^{60}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36} + p_7^{48} + p_8^{48}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36} + p_7^{48}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36} + p_6^{36}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24} + p_5^{36}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24} + p_4^{24}$$

$$n = p_1^{12} + p_2^{12} + p_3^{24}$$

$$n = p_1^{12} + p_2^{12}$$

Example 3.2. All the following Diophantine equations have no solutions in nonzero integers.

1. $x^{12} + y^{24} = 25^{2k-1}z^{36}$

2. $x^4 + y^4 + z^4 = 9^{2k-1}u^4$

3. $x^6 + y^6 + z^6 + w^6 = 13^{2k-1}u^6$

4. $x_1^{10} + x_2^{20} + x_3^{30} + x_4^{40} + x_5^{50} + x_6^{60} + x_7^{70} = 21^{2k-1}y^{80}$

Where k is positive integer.

4 CONCLUSION

Diophantine equations has been known for many years as hot research topics in the study of number theory and cryptography and most popular problem discussed amongst mathematicians

and computer scientists. As the concept of cryptography evolved around solving computational hard problem, therefore solving equations to find rational and integral solutions is one of many preferred problems that cryptographers dwelled themselves in. Most importantly, in a broader sense, solving Diophantine equations can be perceived as Diophantine geometry problems that lead to the idea of solving elliptic curve equations in elliptic curve cryptography (ECC). Another cryptographic idea of Diophantine equation problems is solving multivariate equations that is regarded as multivariate functions or polynomials defined over algebraic field. In view of this, our proposed method in solving the said Diophantine equation shown in this paper, is an initiative to promote and to establish fundamental ideas of such hard mathematical problems that is useful in cryptography.

ACKNOWLEDGMENTS

This work is supported by the Research University (RU) funding, account number 1001 / PMATHS / 811337, Universiti Sains Malaysia.

REFERENCES

- Andrescu, T., Andrica, D., and Cucurezeanu, I. (2010). *An introduction to Diophantine equations*. Birkhäuser Verlag, New York. A problem-based approach.
- AT&T Laboratories, Rosen, K. H. (2005). *Elementary number theory and its applications*. Greg Tobin, fourth edition.
- Bérczes, A., Hajdu, L., Miyazaki, T., and Pink, I. (2016). On the equation $1^k + 2^k + \dots + x^k = y^n$ for fixed x . *J. Number Theory*, 163:43–60.
- Bérczes, A., Pink, I., Sava, G., and Soydan, G. (2018). On the Diophantine equation $(x + 1)^k + (x + 2)^k + \dots + (2x)^k = y^n$. *J. Number Theory*, 183:326–351.
- Cornell, G., Silverman, J. H., and Stevens, G., editors (1997). *Modular forms and Fermat's last theorem*. Springer-Verlag, New York. Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, August 9–18, 1995.
- Herman, J., Kucera, R., and Simsa, J. (2000). *Equations and inequalities: elementary problems and theorems in algebra and number theory*, volume 1. Springer Science & Business Media.
- Mordell, L. J. (1969). *Diophantine equations*, volume 30. Academic Press.
- Narkiewicz, W. (2012). *Rational number theory in the 20th century: from PNT to FLT*. Springer Science & Business Media.
- Taylor, R. and Wiles, A. (1995). Ring-theoretic properties of certain hecke algebras. *Annals of Mathematics*, 141(3):553–572.
- Wiles, A. (1995). Modular elliptic curves and fermat's last theorem. *Annals of mathematics*, 141(3):443–551.

Wong, K. Y. and Kamarulhaili, H. (2016). On the diophantine equation $x^4 + y^4 = p^k z^7$. *International Journal of Pure and Applied Mathematics*, 107:1063–1072.

Wong, K. Y. and Kamarulhaili, H. (2017). On the diophantine equation $x^a + y^a = p^k z^b$. *Journal of Mathematics and Statistics*, 13:38–45.

Elliptic Net Scalar Multiplication using Generalized Equivalent Elliptic Divisibility Sequence

Norliana Muslim^{*1,2} and Mohamad Rushdan Md Said¹

¹*Institute for Mathematical Research, Universiti Putra Malaysia*

²*Faculty of Engineering and Life Sciences, Universiti Selangor*

E-mail: norliana_muslim@unisel.edu.my

**Corresponding author*

ABSTRACT

Elliptic Net is a powerful method to compute cryptographic pairings or scalar multiplication. The elliptic net rank one originated from the nonlinear recurrence relations, also known as the elliptic divisibility sequence. In this paper, a generalization of equivalent sequences is defined. Combining the new generalization with a few restrictions on the initial value, the paper further proposes and discusses an elliptic net scalar multiplication of rank one for Weierstrass equation and non-singular elliptic curve.

Keywords: Equivalence, Net, Divisible, Polynomials

1 INTRODUCTION

Elliptic net scalar multiplication was first introduced by Japanese cryptographer (Kanayama et al., 2014). His method adapts Stange's net theory (Stange, 2007b) and some research directions of elliptic net can be seen in previous year (Muslim and Said, 2017). The rich structure of elliptic net and its scalar multiplication resulted in cryptography field, in which it is used to solve elliptic curve discrete logarithm problem (Lauter and Stange, 2008), compute Ate pairing (Matsuda et al., 2009), and optimize pairing (Tang et al., 2014). Continuous contributions in cryptosystem and net developments are achieved since the discrete log problem on elliptic curve was successfully reduced to a finite field.

In this paper, we begin by reviewing elliptic divisibility sequence with its equivalent properties and division polynomials of the elliptic curve. Next, we propose the elliptic net scalar multiplication of rank one by using new properties. Finally, we discuss the simplification of elliptic net initial values.

2 ELLIPTIC DIVISIBILITY SEQUENCE

Morgan Ward introduced an elliptic divisibility sequence in the form of h_{m+n}
 $h_{m-n}h_1^2 = h_{m+1}h_{m-1}h_n^2 - h_{n+1}h_{n-1}h_m^2$ as a special sequence with the initial value of $h_0 = 0$, $h_1 = 1$, $h_2 \neq 0$ and $h_3 \neq 0$ (Ward, 1948). Meanwhile, the first cryptographic applications of these sequences have been discussed by Shipsey (2000) while the applications were extended by Stange (2007a) and Kanayama et al. (2014). By considering $n = 2$ and $h_1^2 = 1$, two frequently used equations are $h_{2n}h_2 = h_{n+2}h_nh_{n-1}^2 - h_nh_{n-2}h_{n+1}^2$ and $h_{2n+1} = h_{n+2}h_n^3 - h_{n-1}h_{n+1}^3$. Some important topics of elliptic divisibility sequence for cryptographers are the indices (Silverman and Stange, 2011), rank of apparition Gezer and Bizim (2009) and equivalence (Bizim, 2009, Shipsey, 2000). The equivalence theory will be discussed in the next section.

2.1 Equivalent elliptic divisibility sequences

The term of equivalent sequences only can be used for proper elliptic divisibility sequences, in which the $h_0 = 0$, $h_1 = 1$, $h_2 \cdot h_3 \neq 0$ and h_4 divides h_2 . Now, we will show how the equivalent sequences satisfy the nonlinear recurrence relations.

Proposition 2.1. *Consider p, u and v as proper elliptic divisibility sequences and satisfy the nonlinear recurrence relations, $p_{m+n}p_{m-n}p_1^2 = p_{m+1}p_{m-1}p_n^2 - p_{n+1}p_{n-1}p_m^2$, $u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2$ and $v_{m+n}v_{m-n}v_1^2 = v_{m+1}v_{m-1}v_n^2 - v_{n+1}v_{n-1}v_m^2$. Let c_1, c_2 and c_3 be any constant integers and there are equivalent elliptic divisibility sequences $\{j_n\}, \{k_n\}, \{l_n\}$ such that $j_n = c_1^{n^2-1}p_n$, $k_n = c_2^{n^2}u_n$ and $l_n = c_3^n v_n$. Then, $j_{m+n}j_{m-n} = j_{m+1}j_{m-1}j_n^2 - j_{n+1}j_{n-1}j_m^2$, $k_{m+n}k_{m-n} = k_{m+1}k_{m-1}k_n^2 - k_{n+1}k_{n-1}k_m^2$ and $l_{m+n}l_{m-n} = l_{m+1}l_{m-1}l_n^2 - l_{n+1}l_{n-1}l_m^2$.*

Proof. Proof for $j_n = c_1^{n^2-1}p_n$ with $j_{m+n}j_{m-n} = j_{m+1}j_{m-1}j_n^2 - j_{n+1}j_{n-1}j_m^2$ is similar to Shipsey (2000). We will continue to prove for k_n and l_n . Since p, u and v are proper elliptic divisibility sequences, i.e $p_1 = u_1 = v_1 = 1$ then the nonlinear recurrence relations can be simplified to $p_{m+n}p_{m-n} = p_{m+1}p_{m-1}p_n^2 - p_{n+1}p_{n-1}p_m^2$, $u_{m+n}u_{m-n} = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2$ and $v_{m+n}v_{m-n} = v_{m+1}v_{m-1}v_n^2 - v_{n+1}v_{n-1}v_m^2$.

For

$$k_{m+n}k_{m-n} = c_2^{(m+n)^2} u_{m+n} c_2^{(m-n)^2} u_{m-n} \tag{1}$$

$$= c_2^{2(m^2+n^2)} (u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2) \tag{2}$$

$$= c_2^{m^2} u_{m+1} c_2^{m^2} u_{m-1} c_2^{2n^2} u_n^2 - c_2^{n^2} u_{n+1} c_2^{n^2} u_{n-1} c_2^{2m^2} u_m^2 \tag{3}$$

$$= c_2^{(m+1)^2} u_{m+1} c_2^{(m-1)^2} u_{m-1} (c_2^{n^2} u_n)^2 - c_2^{(n+1)^2} u_{n+1} c_2^{(n-1)^2} u_{n-1} (c_2^{m^2} u_m)^2 \tag{4}$$

$$= k_{m+1}k_{m-1}k_n^2 - k_{n+1}k_{n-1}k_m^2 \tag{5}$$

and for

$$l_{m+n}l_{m-n} = c_3^{m+n}v_{m+n}c_3^{m-n}v_{m-n} \quad (6)$$

$$= c_3^{2m}v_{m+n}v_{m-n} \quad (7)$$

$$= c_3^{2m}(v_{m+1}v_{m-1}v_n^2 - v_{n+1}v_{n-1}v_m^2) \quad (8)$$

$$= c_3^{m+1}v_{m+1}c_3^{m-1}v_{m-1} \cdot v_n^2 - v_{n+1}v_{n-1}(c_3^m v_m)^2 \quad (9)$$

$$= l_{m+1}l_{m-1}l_n^2 - l_{n+1}l_{n-1}l_m^2 \quad (10)$$

□

The above steps complete the proof. From Proposition 2.1, we can say that any elliptic divisibility sequences are equivalent if there exist integers c_1, c_2 and c_3 such that for all n , $c_1^{n^2-1}p_n = c_2^{n^2}u_n = c_3^n v_n$. The sequence of $c_3^n v_n$ is a generalization form that will be further used to construct elliptic net scalar multiplication.

3 ELLIPTIC CURVE

The general Weierstrass equation (Silverman, 1986) can be defined as $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ where an elliptic curve E is the set of algebraic solutions of $y^2 = x^3 + ax + b$ whereby a and b are real numbers with the following expression:

$$b_2 = a_1^2 + 4a_2 \quad (11)$$

$$b_4 = 2a_4 + a_1a_3 \quad (12)$$

$$b_6 = a_3^2 + 4a_6 \quad (13)$$

$$b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2 \quad (14)$$

$$D = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \quad (15)$$

The auxiliary polynomials denoted by ϕ_n, ω_n are as follow:

$$\phi_n = x\psi_n^2 - \psi_{n+1}\psi_{n-1} \quad (16)$$

$$4y\omega_n = \psi_{n+2}\psi_{n-1}^2 - \psi_{n-2}\psi_{n+1}^2 \quad (17)$$

Then, for the curve E of the polynomials $\phi_n(P), \psi_n, \omega_n$ can be written as,

$$[n]P = \left(\frac{\phi_n(P)}{\psi_n(P)^2}, \frac{\omega_n(P)}{\psi_n(P)^3} \right) \quad (18)$$

The division polynomials ψ_n in x, y and the first four division polynomials are

$$\psi_1 = 1, \quad \psi_2 = 2y + a_1x + a_3, \quad (19)$$

$$\psi_3 = 3x^4 + b_2x^3 + 3b_4x^2 + 3b_6x + b_8, \quad (20)$$

$$\psi_4 = (2y + a_1x + a_3)(2x^6 + b_2x^5 + 5b_4x^4 + 10b_6x^3 + 10b_8x^2 + (b_2b_8 - b_4b_6)x + b_4b_8 - b_6^2) \quad (21)$$

Therefore, the nonlinear recurrence relations for division polynomial ψ_n when $n \geq 2$ are

$$2y\psi_{2n} = \psi_n(\psi_{n+1}l_{n-1}^2 - \psi_{n-2}\psi_{n+1}^2) \quad (22)$$

$$\psi_{2n+1} = \psi_{n+2}\psi_n^3 - \psi_{n-1}\psi_{n+1}^3 \quad (23)$$

4 ELLIPTIC NET SCALAR MULTIPLICATION OF RANK ONE

The first theory on elliptic net scalar multiplication was proposed by Kanayama et al. (2014) and followed by Chen et al. (2017), with both methods depend on $\hat{W}(i) = \theta^{i^2-1}W(i)$ in their net. It is important to clarify that not all equivalences of proper elliptic divisibility sequences can be used to construct the rank one elliptic net. In our method, the equivalence theory in elliptic net lies on $\hat{W}(2) = 1$ and θ^j . We propose the following definition and lemmas for elliptic net scalar multiplication of rank one using the generalized equivalent elliptic divisibility sequence.

Definition 4.1. *Let $\{W(j)\}$ be the proper elliptic divisibility sequence over a finite field K and $\gcd(2m + 1, 3) = 1$. Then $\hat{W}(j) = \theta^j W(j)$ is a sequence defined over K and $\hat{W}(2) = 1$ with $\theta^2 = W(2)^{-1}$.*

Lemma 4.1. *Consider $\{W(j)\}$ from Definition 4.1, and point $P = (x_1, y_1)$ on elliptic curve of the type $y^2 = Ax + B$ with $\text{Char}(K) \geq 5$. The elliptic net scalar multiplication of rank one $[k]P = (x_k, y_k)$ can be derived as,*

$$x_k = x_1 - \frac{\hat{W}(k-1)\hat{W}(k+1)}{\hat{W}(k)^2} \tag{24}$$

$$y_k = \frac{\hat{W}(k-1)^2\hat{W}(k+2) - \hat{W}(k+1)^2\hat{W}(k-2)}{4y_1\hat{W}(k)^3} \tag{25}$$

Proof. Since $\hat{W}(j) = \theta^j W(j)$, this implies that $\hat{W}(j) = \theta^{-j}W(j)$. Then,

$$x_k = x_1 - \frac{W(k-1)W(k+1)}{W(k)^2} \tag{26}$$

$$= x_1 - \frac{\theta^{-(k-1)}\hat{W}(k-1)\theta^{-(k+1)}\hat{W}(k+1)}{[\theta^{-k}\hat{W}(k)]^2} \tag{27}$$

$$= x_1 - \frac{\theta^{-(k-1)-(k+1)}\hat{W}(k-1)\hat{W}(k+1)}{\theta^{-2k}\hat{W}(k)^2} \tag{28}$$

$$x_k = x_1 - \frac{\hat{W}(k-1)\hat{W}(k+1)}{\hat{W}(k)^2} \tag{29}$$

and for

$$y_k = \frac{W(k-1)^2W(k+2) - W(k+1)^2W(k-2)}{4y_1W(k)^3} \quad (30)$$

$$= \frac{\left(\theta^{-(k-1)}\hat{W}(k-1)\right)^2\theta^{-(k+2)}\hat{W}(k+2) - \left(\theta^{-(k+1)}\hat{W}(k+1)\right)^2\theta^{-(k-2)}\hat{W}(k-2)}{4y_1\left(\theta^{-k}\hat{W}(k)\right)^3} \quad (31)$$

$$= \frac{\theta^{-2(k-1)}\hat{W}(k-1)^2\theta^{-(k+2)}\hat{W}(k+2) - \theta^{-2(k+1)}\hat{W}(k+1)^2\theta^{-(k-2)}\hat{W}(k-2)}{4y_1\left(\theta^{-k}\hat{W}(k)\right)^3} \quad (32)$$

$$= \frac{\theta^{-2k+2-k-2}\hat{W}(k-1)^2\hat{W}(k+2) - \theta^{-2k-2-k+2}\hat{W}(k+1)^2\hat{W}(k-2)}{4y_1\left(\theta^{-3k}\hat{W}(k)^3\right)} \quad (33)$$

$$y_k = \frac{\hat{W}(k-1)^2\hat{W}(k+2) - \hat{W}(k+1)^2\hat{W}(k-2)}{4y_1\hat{W}(k)^3} \quad (34)$$

□

Lemma 4.2. Consider $\{W(j)\}$ from Definition 4.1, and point $P = (x_1, y_1)$ on a non-super singular elliptic curve of type $y^2 + xy = x^3 + a_2x^2 + a_6$ with $\text{Char}(K) = 2$. The elliptic net scalar multiplication of rank one $[k]P = (x_k, y_k)$ can be derived as,

$$x_k = x_1 + \frac{\hat{W}(k-1)\hat{W}(k+1)}{\hat{W}(k)^2} \quad (35)$$

$$y_k = x_1 + y_1 + \left(1 + x_1 + \frac{y_1}{x_1}\right) \frac{\hat{W}(k-1)\hat{W}(k+1)}{\hat{W}(k)^2} + \frac{x_1\hat{W}(k+1)^2\hat{W}(k-2)}{\hat{W}(k)^3} \quad (36)$$

Proof. The derivation for x_k is similar to Lemma 4.1. We will proceed to prove for y_k as follows,

$$y_k = x_1 + y_1 + \left(1 + x_1 + \frac{y_1}{x_1}\right) \frac{W(k-1)W(k+1)}{W(k)^2} + \frac{x_1W(k+1)^2W(k-2)}{W(k)^3} \quad (37)$$

$$= x_1 + y_1 + \left(1 + x_1 + \frac{y_1}{x_1}\right) \frac{\theta^{-(k-1)}\hat{W}(k-1)\theta^{-(k+1)}\hat{W}(k+1)}{\left(\theta^{-k}\hat{W}(k)\right)^2} +$$

$$\frac{x_1\left(\theta^{-(k+1)}\hat{W}(k+1)\right)^2\theta^{-(k-2)}\hat{W}(k-2)}{\left(\theta^{-k}\hat{W}(k)\right)^3} \quad (38)$$

$$= x_1 + y_1 + \left(1 + x_1 + \frac{y_1}{x_1}\right) \frac{\theta^{-2k}\hat{W}(k-1)\hat{W}(k+1)}{\theta^{-2k}\hat{W}(k)^2} +$$

$$\frac{x_1\theta^{-2(k+1)}\hat{W}(k+1)^2\theta^{-(k-2)}\hat{W}(k-2)}{\theta^{-3k}\hat{W}(k)^3} \quad (39)$$

$$y_k = x_1 + y_1 + \left(1 + x_1 + \frac{y_1}{x_1}\right) \frac{\hat{W}(k-1)\hat{W}(k+1)}{\hat{W}(k)^2} + \frac{x_1\hat{W}(k+1)^2\hat{W}(k-2)}{\hat{W}(k)^3} \quad (40)$$

□

Significantly, the factor of θ^{-k} is in the simplest form by using the generalized sequence.

4.1 Discussion

The initial values in the elliptic net scalar multiplication of rank one are as follow:

$$\begin{aligned}\hat{W}(0) &= 0, & \hat{W}(1) &= 1, & \hat{W}(2) &= 1, & \hat{W}(3) &= \hat{p}, & \hat{W}(4) &= \hat{q}, \\ \hat{W}(5) &= \hat{W}(3+2)\hat{W}(3-2) \\ &= \hat{W}(4)\hat{W}(2)[\hat{W}(2)]^2 - \hat{W}(3)\hat{W}(1)[\hat{W}(2)]^2 \\ &= \hat{q} - \hat{p}^3\end{aligned}\tag{41}$$

Meanwhile, the required initial values for Stange method are

$$\begin{aligned}\hat{W}(0) &= 0, & \hat{W}(1) &= 1, & \hat{W}(2) &= \hat{p}, & \hat{W}(3) &= \hat{q}, & \hat{W}(4) &= \hat{r}, \\ \hat{W}(5) &= \hat{W}(3+2)\hat{W}(3-2) \\ &= \hat{W}(4)\hat{W}(2)[\hat{W}(2)]^2 - \hat{W}(3)\hat{W}(1)[\hat{W}(2)]^2 \\ &= \hat{r}\hat{p}^3 - \hat{q}\hat{p}^2\end{aligned}\tag{42}$$

$\hat{W}(5)$ is the last initial value required in the net. The next term for $\hat{W}(6)$ can be calculated using nonlinear recurrence relation of

$$\hat{W}(m+n)\hat{W}(m-n) = \hat{W}(m+1)\hat{W}(m-1)[\hat{W}(n)]^2 - \hat{W}(n+1)\hat{W}(n-1)[\hat{W}(m)]^2\tag{43}$$

such that

$$\hat{W}(6) = \hat{W}(5)\hat{W}(3)[\hat{W}(2)]^2 - \hat{W}(3)\hat{W}(1)[\hat{W}(4)]^2 = \hat{p}[(\hat{q} - \hat{p}^3) - \hat{q}^2]\tag{44}$$

and in Stange method,

$$\hat{W}(6) = \hat{W}(5)\hat{W}(3)[\hat{W}(2)]^2 - \hat{W}(3)\hat{W}(1)[\hat{W}(4)]^2 = \hat{q}(\hat{r}\hat{p}^5 - \hat{q}\hat{p}^4) - \hat{r}^2\tag{45}$$

In order to reduce computation time, it is important to choose an initial values of $\hat{W}(2) = 1$ rather than $\hat{W}(2) \neq 1$. Therefore, the elliptic net scalar multiplication (Chen et al., 2017) and our restriction are shown to provide better simplification.

5 CONCLUSION

This research proposes a generalization of the equivalent elliptic divisibility sequences and uses the generalization form to derive the elliptic net scalar multiplication of rank one. Furthermore, by choosing $\hat{W}(2) = 1$ the term in the proposed elliptic net scalar multiplication was found to be simpler than Stange method. Future research may consider other equivalence theory that satisfies the elliptic net.

ACKNOWLEDGMENTS

The author wishes to thank Universiti Selangor for providing complete research facilities as well as financial support for the project. The author also wishes to thank Miss Edora for formatting paper in Latex.

REFERENCES

- Bizim, O. (2009). On the elliptic divisibility sequences over finite. *World Academy of Science, Engineering and Technology*, 35:1011–1015.
- Chen, B., Hu, C., and Zhao, C. (2017). A note on scalar multiplication using division polynomials. *IET Information Security*, 11(4):195–198.
- Gezer, B. and Bizim, O. (2009). Elliptic divisibility sequences in certain ranks over finite fields. *Hacettepe Journal of Mathematics and Statistics*, 38(2):161–171.
- Kanayama, N., Liu, Y., Okamoto, E., Saito, K., Teruya, T., and Uchiyama, S. (2014). Implementation of an elliptic curve scalar multiplication method using division polynomials. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E97A(1):300–302.
- Lauter, K. E. and Stange, K. E. (2008). The elliptic curve discrete logarithm problem and equivalent hard problems for elliptic divisibility sequences. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5381 LNCS:309–327.
- Matsuda, S., Kanayama, N., Hess, F., and Okamoto, E. (2009). Optimised versions of the ate and twisted ate pairings. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E92A(7):1660–1667.
- Muslim, N. and Said, M. R. M. (2017). Elliptic net and its cryptographic application. In *Proceedings Of The 13th IMT-GT International Conference On Mathematics, Statistics And Their Applications (ICMSA2017)*, volume 1905, Kedah, Malaysia.
- Shipsey, R. (2000). *Elliptic Divisibility Sequences*. PhD thesis, University of London.
- Silverman, J. H. (1986). *The arithmetic of elliptic curve*. Springer-Verlag, New York.
- Silverman, J. H. and Stange, K. E. (2011). Terms in elliptic divisibility sequences divisible by their indices. *Acta Arithmetica*, 146(4):355–378.
- Stange, K. E. (2007a). Elliptic nets and points on elliptic curves. *Algorithmic Number Theory*, (1):1–4.
- Stange, K. E. (2007b). The tate pairing via elliptic nets. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4575 LNCS(1):329–348.
- Tang, C. M., Ni, D. M., Xu, M. Z., Guo, B. A., and Qi, Y. F. (2014). Implementing optimized pairings with elliptic nets. *Science China Information Sciences*, 57(5):1–10.
- Ward, M. (1948). Memoir on elliptic divisibility sequences. *American Journal of Mathematics*, 70(1):31.

Construction of Endomorphisms with J-Invariant 1728 for ISD Method

Siti Noor Farwina Mohamad Anwar Antony^{*1} and Hailiza Kamarulhaili¹

¹*School of Mathematical Sciences, Universiti Sains Malaysia*

E-mail: sitinoorfarwina@yahoo.com, hailiza@usm.my

**Corresponding author*

ABSTRACT

In this paper, we present the idea of constructing the efficiently computable endomorphisms to compute scalar multiplication on elliptic curves with j-invariant 1728 for the Integer Sub-Decomposition (ISD) method. The original ISD method works on integer number field and solves integer scalar multiplication. However, curves with j-invariant 1728 define over the imaginary quadratic field which allows complex multiplication. Hence, the ISD method needs to be extended to work on the imaginary quadratic field to solve the complex multiplication problem. The curve with j-invariant 1728 has a unique discriminant of the imaginary quadratic field. This discriminant of quadratic field yields a unique efficiently computable endomorphism, which later managed to speed up the computations on this curve. However, the ISD method requires three endomorphisms to be accomplished. Hence, we construct all three endomorphisms to be from the same imaginary quadratic field as the first endomorphism.

Keywords: Efficiently computable endomorphism, elliptic scalar multiplication, j-invariant 1728, quadratic field.

1 INTRODUCTION

Let E be an ordinary elliptic curve defined as

$$E : y^2 = x^3 + Ax + B$$

over prime field F_p . The order of elliptic curve define over F_p is denoted as $\#E(F_p) = nh$. For cryptographic purpose $h \leq 4$ where h is the elliptic curve's cofactor. The points on elliptic curve $E(F_p)$ form an abelian group. Let $P = (x, y)$ be points on $E(F_p)$ with prime order n , which forms a subgroup in $E(F_p)$, which is also known as p-subgroup. And there exist only one p-subgroup in $E(F_p)$.

Point multiplication, kP , is one of the most important operation in elliptic curve cryptography (ECC) where $k \in [1, n]$ and $P = (x, y)$ is an element in the p -subgroup. This operation remains to be the most dominant operation in ECC (Park et al., 2002). Researchers have developed approaches such as the Gallant-Lambert-Vanstone (GLV) method and the ISD method to reduce the computation cost.

The GLV method was proposed in 2001 where the scalar k is decomposed into two decomposed scalars k_1 and k_2 such that $k_1, k_2 \leq \sqrt{n}$ (Sica et al., 2002). The general form of GLV method given as

$$kP = k_1P + k_2\Phi(P) \quad (1)$$

such that $\Phi(P) = \lambda P$, where λ is the roots of the minimal polynomial of degree two for the endomorphism, Φ (Gallant et al., 2001). Since the coefficient of the minimal polynomial is in the ring of integer, hence, λ is an algebraic number (Ribenoim, 2001). The GLV method is defined over the complex quadratic field, and it allows complex multiplication on elliptic curves. With the help of efficiently computable endomorphism, the GLV method was able to accelerate the computation by 50%.

However, not all scalars k can be decomposed into scalars $k_1, k_2 \leq \sqrt{n}$. Therefore, the ISD method was proposed to complement the GLV method. The ISD method sub-decomposed the GLV scalars $k_1, k_2 > \sqrt{n}$ into four different scalars $k_{1,1}, k_{1,2}, k_{2,1}, k_{2,2}$ where each scalars fall within \sqrt{n} . The general form of the ISD method is given as

$$kP = k_{1,1}P + k_{1,2}\Phi_1(P) + k_{2,1}P + k_{2,2}\Phi_2(P) \quad (2)$$

where three endomorphisms denoted by Φ, Φ_1, Φ_2 is needed (Ajeena and Kamarulhaili, 2014). By following the same concept in the GLV, the ISD method is capable to increase the percentage of successful computations (Ajeena and Kamarulhaili, 2013). However, the endomorphisms in the ISD method are trivial endomorphisms defined by $X - \lambda = 0$ (Ajeena and Kamarulhaili, 2015). This result in high cost of computation. The mapping for these endomorphisms in the ISD method have not being discussed before. Other than that, these endomorphisms are defined over integer number field which only works on integer multiplications. Hence, the ISD method is not applicable for curves that have complex multiplication such as curves with j -invariant 1728.

In this study, we focus on extending the ISD method to be defined over the imaginary quadratic field which allows it to solve complex multiplication on elliptic curves and able to solve curves with j -invariant 1728. In Section 2, we give some definition and essential theorems related to this study. In Section 3, we describe the first efficiently computable endomorphism acted on curves with j -invariant 1728, and we construct the second and third endomorphism for the ISD method. We also present the cost of computing the second and third endomorphism, and we show an example on how does this endomorphism acted on points in an elliptic curve. Finally, the last section concludes the paper.

2 PRELIMINARIES

In this section, we give some important concepts which are used throughout this study.

Theorem 2.1. (Washington, 2008) Let E be an elliptic curve which allows complex multiplication. Then, $\text{End}(E)$ is isomorphic either to \mathbb{Z} or an order in an imaginary quadratic field.

Definition 2.1. (Washington, 2008) Let $d > 0$ be a square free integer and let

$$K = \mathbb{Q}(\sqrt{-d}) = a + b\sqrt{-d} | a, b \in \mathbb{Q}.$$

Then, K is called an imaginary quadratic field.

Definition 2.2. (Washington, 2008) The discriminant of quadratic field, D is the discriminant of the quadratic polynomial where

$$D = \begin{cases} -f^2d, & \text{if } d \equiv 3 \pmod{4} \\ -4f^2d, & \text{if } d \equiv 1, 2 \pmod{4} \end{cases}$$

where d is the square free integer and f is the conductor of the ring generated by an order in the complex or imaginary quadratic field, $K = \mathbb{Q}(\sqrt{-d})$.

Proposition 2.1. (Cohen, 1996) Let $K = \mathbb{Q}(\sqrt{-d})$ with d a square free integer. Let $\{1, \mathcal{O}_K\}$ be the integral basis of K . Then, the largest subring of K denoted by \mathcal{O}_K , is finitely generated by an abelian group which is defined as

$$\mathcal{O}_K = \begin{cases} \mathbb{Z} \left[\frac{1+\sqrt{-d}}{2} \right], & \text{if } d \equiv 3 \pmod{4} \\ \mathbb{Z} [\sqrt{-d}], & \text{if } d \equiv 1, 2 \pmod{4}. \end{cases}$$

Theorem 2.2. (Washington, 2008) Let $P = (x, y) \in F_p$ be a point on $E : y^2 = x^3 + Ax + B$. Let λ be a positive integer. Then,

$$\lambda P = \left(\frac{A_m(x)}{\psi_m(x, y)^2}, \frac{B_m(x, y)}{\psi_m(x, y)^3} \right). \quad (3)$$

where $A_m(x)$ and $B_m(x, y)$ are rational function and ψ_m is the division polynomial.

Theorem 2.3. (Washington, 2008) Let $E : y^2 = x^3 + Ax$ be an elliptic curve defined over F_p with order, $\#E(F_p)$. Let p be a prime and $A \not\equiv 0 \pmod{p}$. Then, we have

1. If $p \equiv 3 \pmod{4}$, then $\#E(F_p) = p + 1$.
2. If $p \equiv 1 \pmod{4}$, write $p = a^2 + b^2$ where a, b are integers with b is even and $a + b \equiv 1 \pmod{4}$. Then,

$$\#E(F_p) = \begin{cases} p + 1 - 2a, & \text{if } A \text{ is a fourth power mod } p \\ p + 1 + 2a, & \text{if } A \text{ is a square mod } p, \text{ but not a fourth power mod } p \\ p + 1 \pm 2b, & \text{if } A \text{ is not a square mod } p. \end{cases}$$

3 CURVES WITH J-INVARIANT 1728

The elliptic curves with j-invariant 1728 are curves in the form of $E : y^2 = x^3 + Ax$ which is defined over F_p . From the list of the imaginary quadratic field with class number one, Ref.

Cohen (1996), this curve have discriminant of the complex quadratic field, $D = -4$. From Definition 2.2, the discriminant $D = -4$ is correspond to the imaginary quadratic field, $K = \mathbb{Q}(i)$. And according to Theorem 2.3, this curve is ordinary when $p \equiv 1 \pmod{4}$. Moreover, there exists a unique endomorphism that acted on the curve defined over the same imaginary quadratic field. The following proposition describes the mapping of the unique efficiently computable endomorphism acting on curves with j-invariant 1728.

Proposition 3.1. *Let $p \equiv 1 \pmod{4}$ and $P \in E(F_p)$ with prime order n where $E : y^2 = x^3 + Ax$. Let $\beta \in F_p$ be an element of order four. Let $\Phi(P) = \lambda P$ such that $\Phi^2 + 1 = 0$. Then, the map Φ is defined as*

$$\begin{aligned} \Phi : E(F_p) &\rightarrow E(F_p) \\ (x, y) &\mapsto (-x, \beta y) \\ \mathcal{O} &\rightarrow \mathcal{O} \end{aligned}$$

is an endomorphism where $\beta^2 + 1 \equiv 0 \pmod{p}$.

Proof. An element $u \in F_p$ of order four is chosen such that $u^4 \equiv 1 \pmod{p}$. This implies $u^4 - 1 \equiv 0 \pmod{p}$ which can be reduce into $(u + 1)(u - 1)(u^2 + 1) \equiv 0 \pmod{p}$. This corresponds to $u \equiv 1 \pmod{p}, u \equiv -1 \pmod{p}$ and $u \equiv \pm\sqrt{-1} \pmod{p}$. Apparently, the only algebraic number is $u \equiv \pm\sqrt{-1} \pmod{p}$, which obeys a minimal polynomial of the form $u^2 + 1 = 0$. And from Theorem 2.1, Φ is isomorphic to u , since u is an order in imaginary quadratic field. Hence, it obeys the same form of minimal polynomial which is $\Phi^2 + 1 = 0$ and this implies $\lambda^2 + 1 = 0$. From the automorphism definition, the only change of variable that will preserve the equation from $E \rightarrow E$ is given by $x = u^2x, y = u^3y$. Since $u^2 \equiv -1 \pmod{p}$, we then have

$$\begin{aligned} \Phi(x, y) &= (u^2x, u^3y) \\ &= ((-1)x, \beta y) \\ &= (-x, \beta y) \end{aligned}$$

Supposed $\beta \equiv u^3 \pmod{p}$, this implies $\beta^4 \equiv (u^3)^4 \equiv (u^4)^3 \equiv 1 \pmod{p}$. Thus, β is also an element of order four. Therefore, β satisfy the equation $\beta^2 + 1 \equiv 0 \pmod{p}$. \square

Since we need another two endomorphisms to carry out ISD method with complex multiplication, we choose the ring of integer for the second endomorphism and the third endomorphism to be the subring of the endomorphism ring for the first endomorphism. The following lemma describes the existence of the other two non-maximal orders such that they belong to the same complex quadratic field as the maximal order.

Lemma 3.1. *Let $E : y^2 = x^3 + Ax$ defined over a field $K = \mathbb{Q}(i)$ such that $\Phi^2 + 1 = 0$ be the minimal polynomial for the curve with j-invariant 1728, where the maximal order is given as $\mathcal{O}_k = \mathbb{Z}[i]$. Then, there exist two other non-maximal order which is given by $\mathbb{Z}[1 - i]$ and $\mathbb{Z}[1 + i]$ which belong to the same field.*

Proof. From Proposition 2.1, we have the maximal order for the imaginary quadratic field with discriminant, $D = -4$ given by $\mathcal{O}_K = \mathbb{Z}[i]$, and its integral basis as $\{1, i\}$ from Proposition 2.1. This ring of integer generated by the maximal order is isomorphic to the endomorphism ring, which is an abelian group under addition. Any algebraic integer in this abelian group can be written as a linear combination of the basis 1 and i where $z = a(1) + b(i)$ with $a, b \in \mathbb{Z}$. By letting $a = 1, b = 1$, we have $z = 1 + i$. And by letting $a = -1, b = 1$, we have $z = -1 + i$. Both are the element generated by the maximal order $\mathcal{O}_k = \mathbb{Z}[i]$, and both still belong to the field $K = \mathbb{Q}[i]$. \square

From Lemma 3.1, we have the second and third endomorphism in the second layer of decomposition as $\Phi_1^2 - 2\Phi_1 + 2 = 0$ and $\Phi_1^2 + 2\Phi_1 + 2 = 0$ where the endomorphism rings are isomorphic to $End(E) = \mathbb{Z}[\Phi_2] \cong \mathbb{Z}[1 + i]$ and $End(E) = \mathbb{Z}[\Phi_2] \cong \mathbb{Z}[-1 + i]$ respectively. However, we should ensure that the mapping for these endomorphism should be easily computed. The following theorem describes the mapping of the second and third endomorphism.

Theorem 3.1. *Let $p \equiv 1 \pmod{4}$ and $P = (x, y) \in E(F_p)$ with prime order n where $E : y^2 = x^3 + Ax$. Let the second and third endomorphism defined as $\Phi_1^2 - 2\Phi_1 + 2 \equiv 0 \pmod{n}$ and $\Phi_2^2 + 2\Phi_2 + 2 \equiv 0 \pmod{n}$ respectively. Then, the mapping for the second and third endomorphism is given by*

$$\Phi_{1,2}(x, y) = \left(\frac{x^2 + A}{\epsilon_{1,2}^2 x}, y \left[\frac{x^2 - A}{\epsilon_{1,2}^3 x^2} \right] \right)$$

where $\epsilon_{1,2}$ denoted the roots of the minimal polynomials.

Proof. From $E : y^2 = x^3 + Ax$, we have the torsion point $Q = (0, 0)$ of order two. By using Velu's algorithm (Galbraith, 2012), we have

$$\begin{aligned} F &= x^3 + Ax - y^2 \\ F_x &= 3x^2 + A \\ F_y &= -2y \\ U_Q &= F_y(Q) = 0 \quad (\text{points of order two always have } U = 0) \\ V_Q &= F_x(Q) = A. \end{aligned}$$

Then, we have

$$X = x + \frac{V_Q}{x - x_Q} + \frac{U_Q}{(x - x_Q)^2} = x + \frac{A}{x - x_Q} = \frac{x^2 + A}{x}$$

and

$$\begin{aligned} Y &= y - U_Q \frac{2y + a_1x + a_3}{(x - x_Q)^3} + V_Q \frac{a_1(x - x_Q) + y - y_Q}{(x - x_Q)^2} \\ &\quad + \frac{a_1u_Q - F_x(Q)F_y(Q)}{(x - x_Q)^2} \\ &= y - \frac{Ay}{(x - x_Q)^2} = y \frac{x^2 - A}{x^2} \end{aligned}$$

By letting $A_m(x) = \frac{(x^2 + A)}{x}$ and $B_m(x, y) = \frac{y(x^2 - A)}{x^2}$, and $\psi_m(x, y) \equiv \epsilon_{1,2} \pmod{p}$ as the root for the second and third endomorphism and from Theorem 2.2, we have the mapping as

$$\Phi_1(x, y) = \left(\frac{(x^2 + A)}{\epsilon_1^2 x}, \frac{y(x^2 - A)}{\epsilon_1^3 x^2} \right)$$

and

$$\Phi_2(x, y) = \left(\frac{(x^2 + A)}{\epsilon_2^2 x}, \frac{y(x^2 - A)}{\epsilon_2^3 x^2} \right)$$

where $\epsilon_1 \equiv 1 + i \pmod{p}$ and $\epsilon_2 \equiv -1 + i \pmod{p}$. □

Next, we discuss the operations count for the efficiently computable endomorphism defined on elliptic curves with j-invariant 1728. The first endomorphism is define by $\Phi^2 + 1 = 0$ where it maps $\Phi : (x, y) \mapsto (-x, \beta y)$. As can be seen, it only requires one multiplication to obtain ΦP . The following theorem explains the operation counts in the mapping for the second and third endomorphism.

Theorem 3.2. *Let $p \equiv 1 \pmod{4}$ and $P = (x, y) \in E(F_p)$ with prime order n where $E : y^2 = x^3 + Ax$. Given the mapping for the second and third endomorphism as $\lambda_{1,2}P = \left(\frac{x^2 + A}{\epsilon_{1,2}^2 x}, y \left[\frac{x^2 - A}{\epsilon_{1,2}^3 x^2} \right] \right)$ such that λ_1, λ_2 and ϵ_1, ϵ_2 are the root of minimal polynomial for the second and third endomorphism modulo n and p respectively. Then, the cost of computing $\Phi_{1,2}P$ consists of three multiplication, one squaring and two inversion operations.*

Proof. The second and third endomorphisms define by $\Phi_1^2 - 2\Phi_1 + 2 = 0$ and $\Phi_2^2 + 2\Phi_2 + 2 = 0$ respectively. The cost of computing the operation counts involve in that mapping is calculated in the table below:

Multiplication	Squaring	Inversion
$\epsilon_{1,2}^2 \cdot x$	x^2	$\frac{x^2 + A}{\epsilon_{1,2}^2 x}$
$\epsilon_{1,2}^3 \cdot x$		$\frac{x^2 - A}{\epsilon_{1,2}^3 x^2}$
$y \cdot \left[\frac{x^2 - A}{\epsilon_{1,2}^3 x^2} \right]$		
3M	1S	2I

□

The following example illustrates on how these efficiently computable endomorphisms acted on points of an elliptic curve. We let our elliptic curve to be defined over 32-bit length field.

Example 3.1. Let $E : y^2 = x^3 + 3x$ defined over F_p where $p = 1475680177 \equiv 1 \pmod{4}$ and $\#E(F_p) = 1475714276 = 4(368928569)$. So, we have $n = 368928569 \equiv 1 \pmod{4}$ and the cofactor $h = 4$. We choose a random point $P = (1, 2)$ which also have order $\#P = 368928569$. And from Proposition 3.1, we have the first endomorphism as $\Phi^2 + 1 = 0$ where $\lambda = 175354672, 193573897 \pmod{368928569}$ and $\beta = 547721520, 927958657 \pmod{1475680177}$. We then have

$$\lambda P = \begin{cases} 175354672P = (-1, 2 * 547721520) = (1475680176, 1095443040) \\ 193573897P = (-1, 2 * 927958657) = (1475680176, 380237137). \end{cases}$$

And from Theorem 3.1, we have the second endomorphism as $\Phi_1^2 - 2\Phi_1 + 2 = 0$ where $\lambda_1 = 175354673, 193573898 \pmod{368928569}$ and $\epsilon_1 = 547721521, 927958658 \pmod{1475680177}$. Then, we have

$$\lambda_1 P = \begin{cases} 175354673P = \left(\frac{1^2 + 3}{547721521^2 * 1}, 2 \left[\frac{1^2 - 3}{547721521^3 * 1^2} \right] \right) \\ \quad = \left(\frac{4}{299998864566553441}, \frac{-4}{164315834398665656432303761} \right) \\ \quad = (380237137, 547721521) \\ 193573898P = \left(\frac{1^2 + 3}{927958658^2 * 1}, 2 \left[\frac{1^2 - 3}{927958658^3 * 1^2} \right] \right) \\ \quad = \left(\frac{1}{215276817739290241}, \frac{-1}{199767986887862365910856578} \right) \\ \quad = (1095443040, 927958658). \end{cases}$$

And the third endomorphism, $\Phi_2^2 + 2\Phi_2 + 2 = 0$ with $\lambda_2 \equiv 175354671, 193573896 \pmod{368928569}$ and $\epsilon_2 \equiv 547721519, 927958656 \pmod{1475680177}$. By following Theorem 3.1, we have

$$\lambda_2 P = \begin{cases} 175354671P = \left(\frac{1^2 + 3}{547721519^2 * 1}, 2 \left[\frac{1^2 - 3}{547721519^3 * 1^2} \right] \right) \\ \quad = \left(\frac{4}{299998862375667361}, \frac{-4}{164315832598672475605641359} \right) \\ \quad = (1095443040, 547721519) \\ 193573896P = \left(\frac{1^2 + 3}{927958656^2 * 1}, 2 \left[\frac{1^2 - 3}{927958656^3 * 1^2} \right] \right) \\ \quad = \left(\frac{1}{215276816811331584}, \frac{-1}{199767985596201462258991104} \right) \\ \quad = (380237137, 927958656). \end{cases}$$

The following table shows the cost of computing each scalar multiplication from all three endomorphisms using the right-to-left algorithm and using efficiently computable endomorphism. The value in the third column is obtain from Theorem 3.1. Note that, M, S, I are refer to multiplication, squaring and inversion operation respectively.

From Table 1, it is clearly that using efficiently computable endomorphism can actually reduce the cost of computing scalar multiplication.

jP	Operations count using right-to-left algorithm	Operations count using endomorphism
$175354672P$	$80M + 67S + 42I$	$1M$
$193573897P$	$76M + 65S + 39I$	$1M$
$175354673P$	$82M + 69S + 43I$	$3M + 1S + 2I$
$193573898P$	$76M + 65S + 39I$	$3M + 1S + 2I$
$175354671P$	$86M + 71S + 45I$	$3M + 1S + 2I$
$193573896P$	$74M + 64S + 37I$	$3M + 1S + 2I$

Table 1: Comparison of Operations Count Using Right-to-Left Algorithm and Endomorphisms.

4 CONCLUSION

One of the most basic arithmetic in ECC is scalar multiplication. Approaches have been proposed to speed up the computation such as the GLV and ISD method. By extending the ISD method on the imaginary quadratic field, we can solve elliptic curve with complex multiplication such as curve with j-invariant 1728. This curve has a unique discriminant of the imaginary quadratic field, $D = -4$ which correspond to the imaginary quadratic field, $K = \mathbb{Q}(i)$ where $\mathbb{Z}(i)$ is the largest ring of integer in this field. This will result in a unique endomorphism, $\Phi^2 + 1 = 0$, where the mapping only needs one multiplication operation. Since the ISD method needs two more endomorphisms to be able to accomplish, we compute the second and third endomorphism to be as $\Phi_2^2 - 2\Phi_2 + 2 = 0$ and $\Phi_2^2 + 2\Phi_2 + 2 = 0$, which form subrings in $\mathbb{Z}(i)$. Instead of computing the λP , $\lambda_1 P$ and $\lambda_2 P$ using repeated addition and doubling process of right-to-left algorithm, one can use the mapping of efficiently computable endomorphisms directly to obtain the outcomes. The second and third endomorphism only need three multiplication, one squaring and two inversion regardless on how large the field might be. Eventually, these efficiently computable endomorphisms able to reduce the cost of computing kP .

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Ministry of Higher Education and the Universiti Sains Malaysia for the Research University Grant Scheme (FRGS), account no. 1001/PMATHS/811337.

REFERENCES

- Ajeena, R. and Kamarulhaili, H. (2013). Analysis on the elliptic scalar multiplication using integer sub decomposition method. *International Journal of Pure and Applied Mathematics*, 87(1):95–114.

- Ajeena, R. and Kamarulhaili, H. (2014). Glv-isd method for scalar multiplication on elliptic curves. *Australian Journal of Basic and Applied Sciences*, 8(15):1–14.
- Ajeena, R. and Kamarulhaili, H. (2015). On the distribution of scalar k for elliptic scalar multiplication. *AIP Conf. Proc. Journal of Applied Mathematics and Information Sciences*, 1682(020052):1–9.
- Cohen, H. (1996). *A Course in Computational Algebraic Number Theory*.
- Galbraith, S. (2012). *Mathematics of Public Key Cryptography*.
- Gallant, R., Lambert, R., and Vanstone, S. (2001). Faster point multiplication on elliptic curve with efficient endomorphism. *CRYPTO 2001, Advances in Cryptology*, pages 190–200.
- Park, Y., Jeong, S., Kim, C., and Lim, J. (2002). An alternative decomposition of an integer for faster point multiplication on certain elliptic curves. *PKC, LNCS*, 2274:323–334.
- Ribenboim, P. (2001). *Classical Theory of Algebraic Numbers*.
- Sica, R., Ciet, M., and Quisquater, J.-J. (2002). Analysis of the gallant-lambert-vanstone method based on efficient endomorphisms: Elliptic and hyperelliptic curves. *SAC 2002, Selected Areas in Cryptography, 9th Annual International Workshop*, 2595:21–36.
- Washington, L. C. (2008). *Elliptic Curves Number Theory and Cryptography*, 2nd edition.

Garbage-man-in-the-middle (type 2) Attack on the Lucas Based El-Gamal Cryptosystem in the Elliptic Curve Group Over Finite Field

Izzatul Nabila bin Sarbini^{1,5}, Wong Tze Jin^{*1,2}, Koo Lee Feng^{1,2}, Mohamed Othman^{2,3}, Mohd. Rushdan Md. Said^{2,4}, and Yiu Pang Hung¹

¹*Department of Basic Science and Engineering, Universiti Putra Malaysia, Bintulu Campus.*

²*Institute for Mathematical Research, Universiti Putra Malaysia.*

³*Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.*

⁴*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia.*

⁵*Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak.*

E-mail: w.tzejin@upm.edu.my

**Corresponding author*

ABSTRACT

Garbage-man-in-the-middle (type 2) attack is one of the various homomorphic attacks based on its homomorphism nature of cryptosystem. The idea of this attack relies on the possibility to access to the “bin” of recipient. This type of attack requires an access to the “bin” in order to recover the original plaintext. In this paper, an investigation was carried out to evaluate the nature of a homomorphic attack on the Lucas based El-Gamal Cryptosystem in the Elliptic Curve Group over finite field. The result shows that the cryptanalyst is able to obtain the plaintext without knowing the secret number, a , b and R , providing that the receiver decrypts the ciphertexts with a faulty decryption key.

Keywords: bin, decryption, encryption, elliptic curve, Lucas sequence

1 INTRODUCTION

The concept of public key cryptography was proposed by Diffie and Hellman (1976). This is a cryptosystem making use of a public key and a private key. The public key is used to encrypt the plaintext and the private key is used to decrypt the ciphertext. El-Gamal (1985) introduced a signature scheme which is based on Diffie-Hellman key exchange method. This technique is

now referred as El-Gamal Cryptosystem. The security problem of this cryptosystem is based on discrete logarithm. Further, Koblitz (1987) and Miller (1985) individually proposed the public key cryptosystem using elliptic curve group over finite field. The security of the elliptic curve cryptography depends on its ability to compute a point multiplication and the inability of the attacker to calculate the multiplicand using the given, the original and the product points. The size of the elliptic curve determines the difficulty of the problem. Recently, Singh and Singh (2015) proposed a new technique where the classic technique of mapping the characters to affine points in the elliptic curve has been removed to perform the text cryptography using elliptic curve cryptography whilst Thangarasu and Selvakumar (2018) advocated modified Elliptical Curve Cryptography and Abelian group theory to solve linear system problem in sensor-cloud cluster computing.

In mathematics, the second order of Lucas sequences are certain constant-recursive integer sequences that satisfy the recurrence relation, $T_n = PT_{n-1} - QT_{n-2}$, where P and Q are integers. Generally, the second order of Lucas sequences represent the quadratics polynomials in which P and Q are integer coefficients. Due to the recurrence characteristics, Lucas sequences are used to develop the cryptosystem in order to increase its security or efficiency. In this manner, LUCELG proposed by Smith and Skinner (1994), LUC proposed by Smith and Lennon (1993), had been developed based on second order of Lucas sequence to increase their efficiency or security. The LUC_3 proposed by Said (1997) had been developed based on third order Lucas sequence to increase its security. The $LUC_{4,6}$ proposed by Wong et al. (2007) and Wong (2011) had been developed based on the fourth and sixth order of Lucas sequences to increase its security.

The garbage-man-in-the-middle (type 1) and (type 2) attack are the attacks came from Davida's attack and improved by Joye (1997). These attacks rely on the possibility of the cryptanalyst to access the "bin" of the recipient. The garbage-man-in-the-middle attack has three steps - the Lagranges modification step, the step of recovering corresponding plaintext, and the non-trivial relation step. In this paper, garbage-man-in-the-middle (type 2) attack had been selected to strike against the cryptosystem which is based on Lucas sequence and in the elliptic curve group over finite field. This cryptosystem had been proposed by Wong et al. (2014), in which the garbage-man-in-the-middle (type 1) attack was chosen to verify the security. The security of the cryptosystem was then assessed by using Wiener's attack Wong et al. (2018b) and Lenstra's attack Wong et al. (2018a). Hence, it is necessary to select the other types of mathematical attack to analysis further the security capability of the cryptosystem.

2 THE CRYPTOSYSTEM

Let \mathbb{F}_p denotes a finite field of p . Define two points as $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, respectively. An elliptic curve E defined over \mathbb{F}_p is defined as

$$y^2 = x^3 + ax + b \quad (1)$$

where $a, b \in \mathbb{F}_p$ and $4a^3 + 27b^2 \neq 0$. For every field K containing \mathbb{F}_p , one considers the set:

$$E(K) = \{(x, y) \in K \times K \mid y^2 = x^3 + ax + b\} \cup \{\infty\} \quad (2)$$

In the cryptosystem, a general group G will be defined based on elliptic curve and the order of the group G , n is the modulus of system, which is the product of two prime number, r and t .

Suppose the sender and the receiver intend to communicate by using Lucas based El-Gamal cryptosystem in the elliptic curve over finite field with order $n = rt$, then they will choose a secret number, R , which is an element of group G . The sender will choose his own private number, a , whilst the receiver will choose his own private number, b . Both a and b are elements in the group G . Subsequently, the receiver will generate the public key, defined as

$$Q = bR \in R \quad (3)$$

The sender will encrypt the plaintexts (m_1, m_2) using the public key, Q ; and sends three ciphertexts (c_1, c_2, c_3) which defined as

$$\begin{aligned} c_1 &= aR \\ c_2 &= V_{aQ}(m_1, 1) \pmod n \quad \text{and} \\ c_3 &= V_{aQ}(m_2, 1) \pmod n \end{aligned} \quad (4)$$

to the receiver where $V_{aQ}(m_1, 1)$ and $V_{aQ}(m_2, 1)$ are second order Lucas sequence. The second order of Lucas sequence can be defined as

$$V_k(P, 1) = PV_{k-1}(P, 1) - V_{k-2}(P, 1) \quad \text{for } n \geq 2 \quad (5)$$

with initial values, $V_0(P, 1) = 2$ and $V_1(P, 1) = P$. Now, the receiver need to recover the original plaintext by compute the encryption key,

$$e = bc_1 \quad (6)$$

and generates the decryption keys,

$$\begin{aligned} d_1 &= e^{-1} \pmod \left[\left(r - \left(\frac{c_2^2 - 4}{r} \right) \right) \left(t - \left(\frac{c_2^2 - 4}{t} \right) \right) \right] \\ d_2 &= e^{-1} \pmod \left[\left(r - \left(\frac{c_3^2 - 4}{r} \right) \right) \left(t - \left(\frac{c_3^2 - 4}{t} \right) \right) \right] \end{aligned} \quad (7)$$

where $\left(\frac{c_i^2 - 4}{r} \right)$ and $\left(\frac{c_i^2 - 4}{t} \right)$ are Legendre symbol. Employ composite and reverse of Lucas sequences, it's not hard to obtain

$$\begin{aligned} V_{d_1}(c_2, 1) &\equiv m_1 \pmod n \\ V_{d_2}(c_3, 1) &\equiv m_2 \pmod n \end{aligned} \quad (8)$$

In fact, the receiver uses the ciphertext, c_2 to compute the Legendre symbol. Therefore the both quadratic polynomials, $g_1(x) = x^2 - c_2x + 1$ and $f_1(x) = x^2 - m_1x + 1$ must be same type such that the Legendre symbols are $\left(\frac{c_2^2 - 4}{r} \right) = \left(\frac{m_1^2 - 4}{r} \right)$ and $\left(\frac{c_2^2 - 4}{t} \right) = \left(\frac{m_1^2 - 4}{t} \right)$. Similar situation is also applied to the ciphertext c_3 . Thus, the values of a, b and R must be relative to r and t such that the plaintext can be recovered by the receiver correctly.

Consider the following cryptosystem using elliptic curve, $y^2 = x^3 + 13x + 21$ with the modulus $n = 10807$. The sender and receiver agreed to choose $R = 7$. Then, the sender and receiver choose their secret number, $a = 13$ and $b = 49$, respectively; and generates a public key, $Q = 343$. Now, the sender wants to encrypt a set of plain text, $(m_1, m_2) = (20, 91)$ and send the cipher text (c_1, c_2, c_3) to the receiver, where $(20, 91)$ is a point on the elliptic curve. Therefore, the sender computes

$$\begin{aligned}c_1 &= aR = 91 \\c_2 &= V_{4459}(20, 1) \pmod{10807} \equiv 5933 \quad \text{and} \\c_3 &= V_{4459}(91, 1) \pmod{10807} \equiv 1164\end{aligned}$$

When the receiver receives the ciphertext, he will recover the original plaintext as follow steps:

1. Computes Legendre symbol:

$$\begin{aligned}\left(\frac{5933^2 - 4}{101}\right) &= 1 \\ \left(\frac{5933^2 - 4}{107}\right) &= -1 \\ \left(\frac{1164^2 - 4}{101}\right) &= -1 \quad \text{and} \\ \left(\frac{1164^2 - 4}{107}\right) &= 1\end{aligned}$$

2. Computes Encryption Key:

$$e = c_1 \times b = 4459$$

3. Generates Decryption Key:

$$\begin{aligned}d_1 &\equiv 4459^{-1} \pmod{(101 + 1)(107 - 1)} \equiv 3271 \pmod{10812} \quad \text{and} \\ d_2 &\equiv 4459^{-1} \pmod{(101 - 1)(107 + 1)} \equiv 3139 \pmod{10800}\end{aligned}$$

4. Recover the Original Plaintext:

$$\begin{aligned}m_1 &= V_{3271}(5933, 1) \pmod{10807} = 20 \quad \text{and} \\ m_2 &= V_{3139}(1164, 1) \pmod{10807} = 91.\end{aligned}$$

3 THE ATTACK

The garbage-man-in-the-middle (type 2) attack is an attack straightly extended from chosen plaintext attack. Similar to the garbage-man-in-the-middle (type 1) attack, garbage-man-in-the-middle (type 2) consists following three steps:

- i. Lagrange's modification step

- ii. Recovering step and
- iii. Non-trivial relation step.

Theorem 3.1. *Lucas based El-Gamal Cryptosystem in the Elliptic Curve Group, G , over finite field is unsecure if the receiver decrypts the ciphertxts using faulty decryption key.*

Proof. Suppose that n is the product of two prime numbers, r and t . Let R be an element of G which is only known to the sender and receiver. Both of sender and receiver chooses their secret number, $a \in G$ and $b \in G$, respectively; and generates the public key, $Q = bR \in G$. Suppose that (m_1, m_2) be the plaintexts and (c_1, c_2, c_3) be the ciphertxts where

$$\begin{aligned} c_1 &= aR \\ c_2 &\equiv V_e(m_1, 1) \equiv V_{abR}(m_1, 1) \pmod{n} \quad \text{and} \\ c_3 &\equiv V_e(m_2, 1) \equiv V_{abR}(m_2, 1) \pmod{n} \end{aligned}$$

with the encryption key, $e = abR$, decryption key, $d \equiv e^{-1} \pmod{\Phi(n)}$ and $\Phi(n)$ is Euler function.

Step 1 Lagrange's modification step

The cryptanalyst chooses a integers, $k \in G$ and $\gcd(k, Q) = 1$. Then, the cryptanalyst modifies the first ciphertxt, $c'_1 = kc_1$.

Step 2 Recovering step

The receiver will generates the decryption key by

$$\begin{aligned} d'_1 &= (bc'_1)^{-1} \pmod{\left(r - \left(\frac{c_2^2 - 4}{r}\right)\right) \left(t - \left(\frac{c_2^2 - 4}{t}\right)\right)} \\ d'_2 &= (bc'_1)^{-1} \pmod{\left(r - \left(\frac{c_3^2 - 4}{r}\right)\right) \left(t - \left(\frac{c_3^2 - 4}{t}\right)\right)}. \end{aligned}$$

To decrypt the ciphertxts, c_2 and c_3 , the receiver computes

$$\begin{aligned} m'_1 &\equiv V_{d'_1}(c_2, 1) \pmod{n} \\ m'_2 &\equiv V_{d'_2}(c_3, 1) \pmod{n}. \end{aligned}$$

Step 3 Non-trivial relation step.

The cryptanalyst is able to obtains the original plaintexts by evaluate

$$\begin{aligned} V_k(m'_1, 1) &\equiv V_k(V_{d'_1}(c_2, 1), 1) \equiv V_{kd'_1}(c_2, 1) \equiv V_{k(abkR)^{-1}}(c_2, 1) \\ &\equiv V_{(abR)^{-1}}(c_2, 1) \equiv m_1 \pmod{n} \\ V_k(m'_2, 1) &\equiv V_k(V_{d'_2}(c_3, 1), 1) \equiv V_{kd'_2}(c_3, 1) \equiv V_{k(abkR)^{-1}}(c_3, 1) \\ &\equiv V_{(abR)^{-1}}(c_3, 1) \equiv m_2 \pmod{n} \end{aligned}$$

□

In the Lagrange's modification step, the cryptanalyst intercepts the ciphertext. Subsequently, he chooses a random number, k , to modify the first ciphertext and sends the modified ciphertexts, (c'_1, c_2, c_3) , to the receiver.

During the recovering step, the receiver receives the ciphertexts, (c'_1, c_2, c_3) , and generates the decryption keys, d_1 and d_2 using the modified ciphertext, c'_1 . Later, he decrypts the ciphertexts c_2 and c_3 in order to obtain the faulty plaintexts, m'_1 and m'_2 . Since the faulty plaintexts are meaningless, then the receiver will throw the faulty plaintexts into the bin.

If the cryptanalyst is able to obtain the faulty plaintexts in the receiver's bin in time, then he can recover the original plaintext in Non-trivial relation step.

4 CONCLUSION

In this study, an investigation was carried out to evaluate the nature of a homomorphic attack on the Lucas based El-Gamal Cryptosystem in the Elliptic Curve Group over finite field. Result shows that the cryptanalyst is able to obtain the original plaintexts without knowing the secret number, a , b and R , providing that the receiver decrypts the ciphertexts with a faulty decryption key. Thus, the result suggested that the receiver always clean up his bin, especially in time of the receiver decrypts the meaningless plaintexts to avoid any attack from similar forms of garbage-man-in-the-middle (type 2).

ACKNOWLEDGMENT

The authors would like to thank University Putra Malaysia for providing financial support under the Putra-Grant scheme (Vote no: 9588900).

REFERENCES

- Diffie, W. and Hellman, M. (1976). New directions in cryptography. *IEEE Transaction on Information Theory*, 22:644–654.
- El-Gamal, T. (1985). A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transaction on Information Theory*, 31:469–472.
- Joye, M. (1997). *Security Analysis of RSA-type Cryptosystems*. PhD thesis, Universite Catholique de Louvain, Belgium.
- Koblitz, N. (1987). Elliptic curve cryptosystems. *Mathematics of Computation*, 48(177):203–209.

- Miller, V. (1985). Use of elliptic curves in cryptography. In *Advances in Cryptology t CRYPTO 85 (Conference on the Theory and Application of Cryptographic Techniques) Proceedings*, volume 85, pages 417–426.
- Said, M. (1997). *Application of Recurrence Relations to Cryptography*. PhD thesis, Macquarie University, Australia.
- Singh, L. and Singh, K. (2015). Implementation of text encryption using elliptic curve cryptography. *Procedia Computer Science*, 54:73–82.
- Smith, P. and Lennon, M. (1993). Luc: A new public key system. In *Proceedings of the ninth IFIP international Symposium on Computer Security*, pages 103–117.
- Smith, P. and Skinner, C. (1994). A public key cryptosystem and a digital signature systems based on the lucas function analogue to discrete logarithms. In *ASIACRYPT 1994: Advances in Cryptology t ASIACRYPT'94*, pages 298–306.
- Thangarasu, N. and Selvakumar, A. (2018). Improved elliptical curve cryptography and abelian group theory to resolve linear system problem in sensor-cloud cluster computing. *Cluster Computing*, pages 1–10.
- Wong, T. J. (2011). *A RSA-type Cryptosystem Based on Quartic Polynomials*. PhD thesis, Universiti Putra Malaysia, Malaysia.
- Wong, T. J., Koo, L. F., and Yiu, P. H. (2018a). Lucas based el-gamal cryptosystem in the elliptic curve over finite field under lenstras attack. *Asian Journal of Mathematics and Computer Research*, 23(4):207–213.
- Wong, T. J., Koo, L. F., and Yiu, P. H. (2018b). On the wieners attack into lucas based el-gamal cryptosystem in the elliptic curve over finite field. *International Journal of Science and Engineering Investigations vol 7*, 7:37–39.
- Wong, T. J., Said, M. R. M., Atan, K. A. M., and Ural, B. (2007). The quartic analog to the rsa cryptosystem. *Malaysian Journal of Mathematical Sciences*, 1(1):63–81.
- Wong, T. J., Said, M. R. M., Othman, M., and Koo, L. F. (2014). A lucas based cryptosystem analog to the elgamal cryptosystem and elliptic curve cryptosystem. In *AIP Conference Proceedings*, volume 1635, pages 256–259.

On the Underlying Hard Lattice Problems of GGH Encryption Scheme

Arif Mandangan^{*1,2}, Hailiza Kamarulhaili¹, and Muhammad Asyraf Asbullah³

¹*School of Mathematical Sciences, Universiti Sains Malaysia, 11700 USM Penang, Penang, Malaysia*

²*Mathematics, Real Time Graphics and Vizualization Laboratory, Faculty of Science and Natural Resources, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

³*Al-Kindi Cryptography Research Laboratory, , Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

E-mail: arifmandangan@student.usm.my, hailiza@usm.my, ma_asyraf@upm.edu.my, *Corresponding Author*

ABSTRACT

Based on numerous experiments, the inventors of the Goldreich, Goldwasser and Halevi encryption scheme (GGH Scheme) conjectured that the Closest Vector Problem (CVP) instance which arose from the scheme was practically intractable in lattice with a dimension more than 300. However, some attacks had successfully broken the security of the scheme. Instead of solving the conjectured intractable CVP instance, these attacks managed to simplify the instance into its simpler form. Consequently, the security of the GGH Scheme is considered breached. In this paper, we address two most notable attacks on the GGH Scheme. On top of that, we propose a new attack on the GGH Scheme that manages to simplify the underlying CVP instance into a much simpler form. From that, we explicitly define the underlying CVP instance that arises from the GGH Scheme together with its corresponding simplified instances to give further illustration on the weakness points of the scheme. By identifying how these weaknesses are exploited by the attacks, further improvement on the GGH Scheme can be carried out in the future for making the scheme better and stronger.

Keywords: GGH scheme, lattices-based cryptosystem, post-quantum, closest vector problem, algebraic cryptanalysis

1 INTRODUCTION

The first practical lattice-based cryptographic scheme was proposed by Goldreich et al. (1997). The security of Goldreich-Goldwasser-Halevi Encryption Scheme (GGH Scheme) lies on the

hardness of Closest Vector Problem (CVP). In full generality, the CVP is considered as NP-hard problem in lattice. The problem becomes more difficult as the lattice dimension being implemented increases (Hoffstein et al., 2014). Through numerous experiments, the underlying CVP instance was conjectured to be practically intractable when implemented in lattice with a dimension larger than 300 (Goldreich et al., 1997). To build confidence on the scheme, its inventors published five ciphertext in lattice dimensions of 200, 250, 300, 350 and 400 on the internet (Goldreich et al., 1997). Most of the attempts to decrypt the challenge were conducted by using embedding technique to reduce the underlying CVP instance into the Shortest Vector Problem (SVP) instance. By a direct implement of the embedding technique into the underlying CVP instance, the launched attacks only managed to solve the challenge for lattice dimension of 200.

In 1999, Phong Q. Nguyen found a novel way to simplify the underlying CVP instance of the GGH Scheme. Consequently, he completely decrypted four GGH challenges and partially decrypted the challenge for the lattice dimension of 400 (Nguyen, 1999). We refer the attack as Nguyens attack. The final challenge remained unsolved for more than 10 years. Only in 2010, Lee and Hahn completely broke the challenge and decrypted the whole 400 ciphertext (Lee and Hahn, 2010). We refer the attack as Lee-Hahns attack. Since then, there are some notable efforts for improving the scheme can be found in literature (Micciancio, 2001), (Paeng et al., 2003), (Yoshino and Kunihiro, 2012) and (de Barros and Schechter, 2014). According to Plantard and Susilo (2009), the general idea behind the scheme is still viable and worthy to improve. In addition, the GGH Scheme is considered as the most intuitive encryption scheme based on lattices (Micciancio and Regev, 2009). Therefore, the remedy to heal the scheme is still worthy to be explored for keeping the scheme alive and surviving.

This paper aims to address two most notable attacks on the GGH Scheme, namely the Nguyens and Lee-Hahns attacks. We noticed that, both attacks used the same strategy in attacking the scheme. Instead of solving the underlying CVP instance of the scheme, these attacks managed to simplify the instance into its simpler form. On top of that, we propose a new attack which works on the GGH Scheme when it is implemented under certain conditions. Through this attack, we reduced the underlying CVP instance into a much simpler form. By considering this attack, further improvement on the GGH Scheme must avoid the conditions which allow the proposed attack works efficiently. Therefore, stronger countermeasures could be carried out in the future for strengthening the GGH Scheme.

2 GGH ENCRYPTION SCHEME

In this section, we briefly describe the GGH Scheme as proposed by Goldreich et al. (1997). Suppose that Alice and Bob want to communicate and they decide to use GGH Scheme. Alice initiates her keys generation by choosing security parameters $n, \sigma \in \mathbb{N}$ where n is a lattice dimension and σ is a perturbation parameter. Basically, Alices public and private keys are two different bases of the same lattice L , i.e., $\mathcal{L}(R) = L = \mathcal{L}(B)$. The private key R is a good basis with reasonably short and orthogonal basis vectors. For $k, l \in \mathbb{Z}^+$, the private key R is generated as $R = kI + P$ with both I and P are n -by- n matrices where I is an identity matrix and P is a perturbation matrix with integer entries that are uniformly distributed in $\{-l, \dots, l\}$.

The value of k is computed as $k = \lceil \sqrt{n} + 1 \rceil$. On the contrary, the public basis B is a bad basis with long and highly non-orthogonal basis vectors. It is derived from the private basis R in such a way that, deriving the public basis B from R is easy but recovering the private basis R from B is computationally infeasible. Alice sends her public basis B together with her security parameters (n, σ) to Bob and keeping her private basis R and other parameters secretly.

Upon receiving Alices public basis B and security parameters (n, σ) , Bob generates an error vector $\vec{e} \in \mathbb{Z}^n$ with entries $e_i \in \{\pm\sigma\}$ for all $i = 1, \dots, n$. He encodes the secret message into a vector $\vec{m} \in \mathbb{Z}^n$. The encoded message then is encrypted as, $\vec{c} = B\vec{m} + \vec{e}$ where $\vec{c} \in \mathbb{Z}^n$ is a ciphertext vector. Bob sends the ciphertext \vec{c} to Alice. Upon receiving Bobs ciphertext, Alice applies Babais round-off method to recover the vector \vec{m} which contains the encoded message. She computes a vector $\vec{t} \in \mathbb{R}^n$ as $\vec{t} = R^{-1}\vec{c}$. Then, forms an integral vector $\lfloor \vec{t} \rfloor$ by rounding each entries of the vector \vec{t} as $\lfloor t_i \rfloor \in \mathbb{Z}$ for all $i = 1, \dots, n$. She computes a unimodular matrix U such that $U = B^{-1}R$. Finally, decrypts the message as $\vec{m} = U \lfloor \vec{t} \rfloor$ and decodes the secret message from the vector \vec{m} . Obviously, the encryption formula is very simple and consumes low computational cost. In the encryption process, the lattice vector $\vec{v} = B\vec{m}$ is perturbed by the error vector $\vec{e} \in \{\pm\sigma\}^n$ to form the ciphertext vector \vec{c} , which is a non-lattice vector. In this case, the security of the scheme relies on a Closest Vector Problem (CVP) instance which can be explicitly defined as follows:

Definition 2.1. *Let $n, \sigma \in \mathbb{Z}^+$. Given a lattice dimension n , a basis B for a lattice $\mathcal{L}(B) = L$, a perturbation parameter σ and a ciphertext $\vec{c} \in \mathbb{R}^n$ such that $\vec{c} = \vec{v} + \vec{e}$, where $\vec{e} \in \{\pm\sigma\}^n$ is an unknown vector and $\vec{v} \in L$ is an unknown lattice vector. GGH-CVP instance is a problem to find a lattice vector \vec{v} that is closest to ciphertext \vec{c} which minimizes the Euclidean norm $\|\vec{c} - \vec{v}\|$.*

The GGH-CVP instance can be solved by using Babais round-off method which works efficiently only when the basis used in the computation process is a good basis with short and reasonably orthogonal basis vectors. That means, only Alice as an authorized recipient who has the private basis R will be able to execute the Babais round-off method effectively for decrypting the ciphertext sent by Bob. The only available information to Eve as an unauthorized party, is the public basis B which is a bad basis. Executing the Babais round-off method by using the basis B will give an output of undesired closest vector. However, the orthogonality of the public basis B can be enhanced by using lattice reduction methods such as Lenstra-Lenstra-Lovasz (LLL) and Block Korkine-Zolotarev (BKZ) algorithms. In this case, the security of the GGH Scheme relies on a Smallest Basis Problem (SBP) instance which can be explicitly defined as follows:

Definition 2.2. *Let $n, \sigma \in \mathbb{Z}^+$. Given a lattice dimension n , a basis B for a lattice $\mathcal{L}(B) = L$, a perturbation parameter σ and a ciphertext $\vec{c} \in \mathbb{R}^n$ such that $\vec{c} = B\vec{m} + \vec{e}$, where $\vec{e} \in \{\pm\sigma\}^n$ is an unknown vector and $\vec{m} \in \mathbb{Z}^n$ is an unknown vector contains the encoded secret message. GGH-SBP instance is a problem to find a reduced-basis B' where the vector $\vec{w} = B' \lfloor \vec{c} B'^{-1} \rfloor$ minimizes the Euclidean norm $\|\vec{c} - \vec{w}\|$.*

If the GGH-SBP instance can be solved, then the obtained basis would be as good as the private basis R for enabling the Babais round-off method works effectively to decrypt the ciphertext \vec{c} . To make the GGH-SBP instance harder, the lattice dimension to be implemented should be large enough to make lattice reduction algorithms inefficient.

3 NGUYENS ATTACK

The Nguyens attack consists of three main stages. The first stage is a simplification of the GGH-CVP instance. Then, followed by a reduction stage to reduce the simpler CVP instance into a Shortest Vector Problem (SVP) instance. In the final stage, the SVP instance will be solved by using lattice reduction methods. In the simplification stage, a vector $\vec{s} \in \{\sigma\}^n$ is inserted into the encryption formula as follows,

$$\vec{c} + \vec{s} \equiv B\vec{m} + \vec{e} + \vec{s} \pmod{2\sigma}$$

Since $\vec{e} \in \{\pm\sigma\}^n$, then the congruence $\vec{e} + \vec{s} \equiv \vec{0} \pmod{2\sigma}$ holds. Thus,

$$\vec{c} + \vec{s} \equiv B\vec{m} \pmod{2\sigma} \quad (1)$$

The only unknown value in congruence (1) is the vector \vec{m} which contains the encoded message. Nguyen (1999) showed that the congruence (1) is easy to solve with very few solutions. Suppose that the congruence (1) is solved and the value of $\vec{m} \pmod{2\sigma}$ is known and denoted as $\vec{m}_{2\sigma}$. The known $\vec{m}_{2\sigma}$ now is inserted into the encryption formula as follows:

$$\begin{aligned} \vec{c} - B\vec{m}_{2\sigma} &= B\vec{m} - B\vec{m}_{2\sigma} + \vec{e} \\ &= B(\vec{m} - \vec{m}_{2\sigma}) + \vec{e} \end{aligned} \quad (2)$$

Note that $\vec{m} \equiv \vec{m}_{2\sigma} \pmod{2\sigma}$, which means there exist $\vec{m}' \in \mathbb{Z}^n$ such that

$$\vec{m} - \vec{m}_{2\sigma} = 2\sigma\vec{m}' \quad (3)$$

By inserting equation (3) into equation (2), we have

$$\begin{aligned} \vec{c} - B\vec{m}_{2\sigma} &= B(2\sigma\vec{m}') + \vec{e} \\ \frac{\vec{c} - B\vec{m}_{2\sigma}}{2\sigma} &= B\vec{m}' + \frac{\vec{e}}{2\sigma} \end{aligned} \quad (4)$$

Referring to equation (4), the left side is a known non-lattice vector, $B\vec{m}'$ is an unknown lattice vector and the right-most vector is a new unknown error vector. Since $\vec{e} \in \{\pm\sigma\}^n$, therefore

$$\frac{\vec{e}}{2\sigma} \in \left\{ \pm \frac{1}{2} \right\}^n$$

This new error vector is much smaller compared to its initial form. It is more than half shorter than $\vec{e} \in \{\pm\sigma\}^n$ that was used in the GGH internet challenges. From equation (4), a simpler CVP instance can be explicitly defined as follows:

Definition 3.1. Let $n \in \mathbb{N}$ and $L \subset \mathbb{R}^n$ be a lattice. Given a vector $\vec{c}' \in \mathbb{R}^n$ such that:

$$\vec{c}' = \vec{y} + \left\{ \pm \frac{1}{2} \right\}^n$$

where $\vec{y} \in L$ is an unknown lattice vector. The simplified GGH-CVP1 instance is a problem to find a lattice vector $\vec{y} \in L$ that is closest to \vec{c}' which minimizes the Euclidean norm $\|\vec{c}' - \vec{y}\|$.

Note that, the non-lattice vector \vec{c}' is now located much closer to the lattice vector \vec{y} that contains partial information about the vector \vec{m} . This makes the effort for correcting the error becomes easier than before. Consequently, Nguyens attack succeeded for completely decrypting the GGH Internet Challenges in lattice dimensions of 200, 250, 300 and 350. By using the available computing power and technology during that time, the solution of the challenge for lattice dimension of 400 was unreachable by lattice reduction algorithm.

4 LEE-HAHNS ATTACK

Lee-Hahns attack requires some exact value of the secret message \vec{m} to succeed. Experimentally, the more exact value of the plaintext they know, the more effective the attack they can perform (Lee and Hahn, 2010). To decrypt the GGH challenge for the lattice dimension of 400, the Lee-Hahn's attack required the initial stage of Nguyens attack for gaining partial information of the vector \vec{m} . They used the partially decrypted ciphertext, $\vec{m}_{2\sigma} \in \mathbb{Z}^{400}$ published in (Nguyen, 1999) to guess the exact value of some entries of $\vec{m} \in \mathbb{Z}^{400}$ of the challenge which consequently led the attack to completely decrypt the whole ciphertext. In this section, we explain how the Lee-Hahns attack works and why it succeeds. Suppose that the first k entries out of n entries of the vector $\vec{m} \in \mathbb{Z}^n$ are known. The vector \vec{m} is now represented as $\vec{m} = \begin{pmatrix} \vec{m}_1 \\ \vec{m}_2 \end{pmatrix} \in \mathbb{Z}^n$ where \vec{m}_1 represents the known first k entries of \vec{m} and \vec{m}_2 represents the remaining unknown entries. Similarly, the public basis B also has a new representation $B = (B_1 \ B_2)$ where $\vec{b}_i \in B_1$ for $i = 1, \dots, k$ and $\vec{b}_i \in B_2$ for $i = k + 1, \dots, n$. From the encryption formula $\vec{c} = B\vec{m} + \vec{e}$, we have

$$\begin{aligned} \vec{c} &= (B_1 \ B_2) \begin{pmatrix} \vec{m}_1 \\ \vec{m}_2 \end{pmatrix} + \vec{e} \\ \vec{c} - B\vec{m}_1 &= B_2\vec{m}_2 + \vec{e} \end{aligned} \quad (5)$$

Referring to equation (5), the left side is a known non-lattice vector, $B_2\vec{m}_2$ is an unknown lattice vector and the right-most vector is an unknown error vector. Hence, a new simpler CVP instance can be explicitly derived from equation (5) as follows:

Definition 4.1. Let $n, \sigma, k \in \mathbb{N}$ where $k < n$, $\vec{e} \in \{\pm\sigma\}^n$, $\vec{m} \in \mathbb{Z}^n$ and $B = (B_1 \ B_2)$ is a basis for a lattice $L \subset \mathbb{R}^n$ where the sub-basis B_1 consists the first k basis vectors and the sub-basis B_2 consists the remaining basis vectors. Suppose that, the vector \vec{m} can be represented as $\vec{m} = \begin{pmatrix} \vec{m}_1 \\ \vec{m}_2 \end{pmatrix} \in \mathbb{Z}^n$ where the sub-vector \vec{m}_1 consists of the first k entries of \vec{m} and the sub-vector \vec{m}_2 consists of the remaining entries. Given $\vec{c} \in \mathbb{R}^n$, $\vec{m}_1 \in \mathbb{Z}^k$ and a basis B for a lattice L , the simplified GGH-CVP2 instance is a problem to find a lattice vector $\vec{z} \in L$ such that $\vec{c} - B_1\vec{m}_1 = \vec{z} + \vec{e}$ minimizes the Euclidean norm $\|\vec{c}' - \vec{z}\|$ where $\vec{c}' = \vec{c} - B_1\vec{m}_1$.

Compared to GGH-CVP instance, the lattice vector $\vec{z} = B_2\vec{m}_2$ is smaller than the lattice vector $\vec{v} = B\vec{m}$. This is because the lattice vector $B_2\vec{m}_2$ is a linear combination of smaller basis B_2 and a smaller integer vector \vec{m}_2 . The only missing information to complete Lee-Hahns attack on the GGH challenge is the entries of the vector $\vec{m}_1 \in \mathbb{Z}^k$. Lee-Hahns attack used the published $\vec{m}_{2\sigma} \in \mathbb{Z}^{400}$ in (Nguyen, 1999) for guessing some entries of the vector $\vec{m} \in \mathbb{Z}^{400}$ to form the vector \vec{m}_1 . By having the vector $\vec{m}_1 \in \mathbb{Z}^k$, the simplified GGH-CVP2 instance becomes complete and it can be reduced into an SVP instance by using embedding technique. According to Lee and Hahn (2010), the shortest vector in the sub-lattice $\mathcal{L}(B_2)$ could be longer than the shortest vector in the lattice $\mathcal{L}(B)$. This may increase the lattice gap in the derived SVP instance and made the instance solvable by lattice reduction algorithm in reasonable amount of time. Consequently, the Lee-Hahn's attack successfully decrypted the whole 400 ciphertext of the challenge.

5 A NEW ATTACK ON THE GGH SCHEME

The security parameters of GGH Scheme consist of $n, \sigma \in \mathbb{N}$. Thus, the public basis B is a known n -by- n non-singular matrix, $\vec{c} \in \mathbb{Z}^n$ is a known ciphertext vector, $\vec{e} \in \{\pm\sigma\}^n$ is an unknown error vector and $\vec{m} \in \mathbb{Z}^n$ is an unknown vector which contains the encoded message. Consider the following lemma:

Lemma 5.1. *Let S be a 1-by- n matrix with entries $s_{1,j} = +\sigma$ for all $j = 1, \dots, n$ and T be a n -by-1 matrix with entries $t_{i,1} = \pm\sigma$ for all $i = 1, \dots, n$ where $n, \sigma \in \mathbb{N}$ and n is an even number. If the number of $+\sigma$ entries is equal to the number $-\sigma$ entries in the matrix T (in any positions), then $ST = 0$.*

Proof

Given that,

$$S = (s_{1,1} \quad s_{1,2} \quad \cdots \quad s_{1,n})$$

and

$$T = \begin{pmatrix} t_{1,1} \\ t_{2,1} \\ \vdots \\ t_{n,1} \end{pmatrix}$$

Thus, ST is a 1-by-1 matrix computed as follows

$$ST = (s_{1,1}t_{1,1} + s_{1,2}t_{2,1} + \cdots + s_{1,n}t_{n,1})$$

Note that $s_{1,j} = \sigma$ and $t_{i,1} = \pm\sigma$ for all $i, j = 1, \dots, n$. Thus,

$$ST = (\sigma(\pm\sigma) + \sigma(\pm\sigma) + \cdots + \sigma(\pm\sigma))$$

and

$$ST = (\sigma^2 + \sigma^2 + \cdots + \sigma^2 - \sigma^2 - \sigma^2 - \cdots - \sigma^2)$$

Since there are equal numbers of $+\sigma$ and $-\sigma$ in T and also n is an even number, therefore

$$ST = \left(\frac{n}{2}(\sigma^2) - \frac{n}{2}(\sigma^2)\right) = 0$$

To launch the proposed attack, we treat the encryption formula as the following:

$$C = BM + E \tag{6}$$

where C is a known n -by-1 ciphertext matrix, B is a known n -by- n non-singular basis matrix and M is an unknown n -by-1 message matrix. Suppose that the dimension n of the lattice $L \subset \mathbb{R}^n$ is an even number and E is an unknown n -by-1 error matrix with entries $e_{i,1} \in \{\pm\sigma\}$ for all $i = 1, \dots, n$ with the same number of positive entries and negative entries regardless the position of the entries, i.e. $n/2$ entries are $+\sigma$ and the remaining $n/2$ entries are $-\sigma$. Set a new 1-by- n integer matrix X with entries $x_{1,j} = +\sigma$ for all $j = 1, \dots, n$. Multiply the matrix X to the both sides of equation (6) yields

$$XC = XBM + XE$$

According to Lemma 5.1, we have $XE = 0$. Thus,

$$XC = XBM \tag{7}$$

Clearly, the error matrix E has been eliminated as done by Nguyens attack for attacking the GGH Scheme. But this time, there is no modular reduction required. However, the challenging part now is to find a way to solve the equation (7) for the matrix M . We found that the solution for equation (7) which gives exactly the entries as M (the encrypted message) is challenging to compute. Further investigation and development on this attack are still in progress.

6 DISCUSSION

Basically, both Nguyen's and Lee-Hahn's attacks used the same strategy in attacking the GGH Scheme. Instead of solving the conjectured intractable GGH-CVP instance, the attacks managed to simplify the instance into its simpler forms. The simpler instances then were reduced into SVP instances by using embedding technique. Finally, the derived SVP instances were solved by using lattice reduction methods and consequently yielded the solution of the GGH-CVP instance. The success factor of these attacks was the simplification of the GGH-CVP instance. In its original form, the instance was experimentally shown to be resistant against embedding attack and lattice reduction attacks (Goldreich et al., 1997). But in its simpler form, these attacks were able to work effectively even in larger lattice dimensions. Compared to Nguyens and Lee-Hahns attacks, we expect that the proposed attack could be more efficient since it does not require any modular reduction operations. We expect that, the ability of this attack to eliminate the error matrix E could be potentially used to justify another weakness of the GGH Scheme which should be avoided by any improved-variants of the scheme. Once this attack is completely works, then any improved-variants should avoid implementing the parameter setup which permit this attack. That means, the implemented lattice dimension should be an odd natural number and the number of $+\sigma$ and $-\sigma$ entries in the error vector \vec{e} must not the same.

7 CONCLUSION

We addressed the most devastating attacks on the GGH Scheme. On top of that, we propose a new attack which is able to eliminate the error vector \vec{e} . Although the proposed attack works only on certain conditions, but it potentially discovered another weakness point of the GGH Scheme which should be avoided by any improved-variants of the scheme. Improving the GGH Scheme is still worth to be worked on. Great simplicity that is offered by this scheme should be appreciated. It could be a highly competitive scheme in the post-quantum era. Therefore, more effort to improve the scheme should be carried out and we are working into it.

ACKNOWLEDGMENTS

The authors would like to thank Universiti Putra Malaysia and Universiti Sains Malaysia for financial support on this paper to participate in CRYPTOLOGY 2018.

REFERENCES

- de Barros, C. F. and Schechter, L. M. (2014). Ggh may not be dead after all. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 3(1).
- Goldreich, O., Goldwasser, S., and Halevi, S. (1997). Public-key cryptosystems from lattice reduction problems. In Kaliski, B. S., editor, *Advances in Cryptology — CRYPTO '97*, pages 112–131, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Hoffstein, J., Pipher, J., and Silverman, J. H. (2014). Lattices and cryptography. In *An Introduction to Mathematical Cryptography*, pages 373–470. Springer.
- Lee, M. S. and Hahn, S. G. (2010). Cryptanalysis of the ggh cryptosystem. *Mathematics in Computer Science*, 3(2):201–208.
- Micciancio, D. (2001). Improving lattice based cryptosystems using the hermite normal form. In *Cryptography and lattices*, pages 126–145. Springer.
- Micciancio, D. and Regev, O. (2009). Lattice-based cryptography. In Bernstein, D. J., Buchmann, J., and Dahmen, E., editors, *Post-Quantum Cryptography*, pages 147–191. Springer Berlin Heidelberg.
- Nguyen, P. (1999). Cryptanalysis of the goldreich-goldwasser-halevi cryptosystem from crypto97. In *Annual International Cryptology Conference*, pages 288–304. Springer.
- Paeng, S., Jung, B. E., and Ha, K. (2003). A lattice based public key cryptosystem using polynomial representations. In Desmedt, Y. G., editor, *Lecture Notes in Computer Science 2567-PKC 2003*, pages 292–308. Springer, Berlin, Heidelberg.

- Plantard, T. and Susilo, W. (2009). Broadcast attacks against lattice-based cryptosystems. In Abdalla, M., editor, *Lecture Notes in Computer Science 5536- ACNS 2009*, pages 456–472. Springer, Berlin, Heidelberg.
- Yoshino, M. and Kunihiro, N. (2012). Improving ggh cryptosystem for large error vector. In *Information Theory and its Applications (ISITA), 2012 International Symposium on*, pages 416–420. IEEE.

Detecting General Algebraic Manipulation Attacks

Kim Ramchen¹

¹*Department of Computing and Information Systems, The University of Melbourne*

E-mail: kim.ramchen@unimelb.edu.au

ABSTRACT

Algebraic manipulation detection codes are a class of error detecting codes which have found numerous applications in cryptography. In this paper we extend these codes to defeat general algebraic attacks - we call such codes general algebraic manipulation detection (GAMD) codes. Positive results are shown for the existence of GAMDs for the families of tampering functions corresponding to point additions and polynomial functions over a finite field. Compared to non-malleable codes, we demonstrate both positive and negative results regarding the existence of GAMDs for arbitrary families of tampering functions.

1 INTRODUCTION

Fault injection attacks are a class of attacks involve the deliberate introduction of errors into the circuitry or memory modules of a cryptographic device in attempt to deduce some secret state. Algebraic manipulation detection codes (Cramer et al., 2008) are a class of error detecting codes that can thwart such attacks when the class of induced faults corresponds to additions on code-words over a finite space. More precisely let s be a message supplied by an adversary, and suppose c , an element of an abelian group \mathcal{G} , is the corresponding code-word. If for any $\Delta \in \mathcal{G}$ it holds that $c + \Delta$ decodes to s' for any $s' \neq s$, with probability bounded by ϵ , the scheme is said to be an AMD code with error probability ϵ .

Even though AMD codes provide an elegant, keyless alternative to the widely used message authentication codes for robust transmission over an error-prone channel, they cannot defeat some types of powerful adversaries. Suppose that an AMD code is used to protect the output of a one time pad scheme. Let $\mathcal{E}(K \oplus M)$ be the output on ciphertext $c = K \oplus M$. If it happens that \mathcal{E} possesses a linear homomorphism ϕ , then we have $\Delta M \circ_{\phi} \mathcal{E}(c) = \Delta M \circ_{\phi} \mathcal{E}(K \oplus M) = \mathcal{E}(K \oplus (M \oplus \Delta M)) = \mathcal{E}(K \oplus M')$, where M' is the message to be substituted. It is therefore desirable to consider a more powerful adversarial model in which an attacker can choose, in addition to the source message, a tampering function F from a rich class of tampering functions \mathcal{F} . In this work, we consider precisely this model, when the class \mathcal{F} corresponds to algebraic functions over some finite field or the rationals corresponding to the co-domain of the AMD code. We call such a code a generalised algebraic manipulation detection code (GAMD code).

Following previous works on algebraic manipulation detection, we distinguish the case when the source message is assumed to be uniformly distributed over the message space, from the usual (which provides tampering detection with bounded error probability for any message). These are called weak generalised algebraic manipulation detection (weak GAMD) and generalised algebraic manipulation detection (GAMD) respectively.

1.1 Our Contributions

We formally introduce the model of generalised algebraic manipulation detection, in which tamperings corresponding to algebraic functions over the ambient field of the encoding function. In this model we review the previous constructions for manipulation detection against point additions. We show that such constructions translate directly to our new model, leading to direct instantiations of weak GAMDs and GAMDs for this class. Additionally we present a new construction for weak GAMDs in the case of encoding over \mathbb{F}_2 based upon the probabilistic method, leading to the following result (we actually construct a GAMD for a more general class of tampering functions, this is discussed in Section 3.1.1)

Theorem 1.1. (*Probabilistic construction of addition evasive GAMDs - Informal*) *Let n be a power of two. There exists a n^{c-1} -GAMD against the class of point additions on \mathbb{F}_n with rate $c - o(1)$, for any constant $0 < c < 1$.*

We also consider attacks corresponding to the class of polynomial functions. Such attacks in the affine case have been considered in the context of non-malleable cryptography by (Aggarwal et al., 2014, Kiayias et al., 2016). We demonstrate an explicit construction of a GAMD secure against the class of polynomial functions of bounded degree.

Theorem 1.2 (Construction of GAMDS for bounded degree polynomials - Informal). *Fix a positive integer d . There exists an explicit weak ϵ -GAMD secure against the class of polynomials of degree bounded by d of rate $2/\Theta(d^2)$ and error probability $\frac{O(k)}{d} \cdot 2^{-\frac{k}{\Theta(d^2)}}$ where k is the prime bit-length.*

We show that exact constructions imply corresponding weak GAMD codes with inverse polynomial rate and low error-probability. We present a black-box transformation of any weak GAMD to a GAMD. This construction is quite efficient, implying in view of the above results, the existence of GAMDs with constant rate and low error probability for the classes of point additions and polynomial functions respectively. Compared to the celebrated non-malleable codes (Dziembowski et al., 2010) we also establish some separations. Our first result is negative and states that there exists a class of tampering functions for which non-malleable codes but not GAMD codes exist. This may be summarised by

Theorem 1.3 (Non-existence of GAMDs for all functions - Informal). *There exists a family of tampering functions for which non-malleable codes exist with constant rate and negligible simulation error but ϵ -GAMD codes with constant rate do not exist, for any choice of non-negligible ϵ .*

Our second result is a positive one and states that for any non-malleable code there exists a class of tampering functions which violates non-malleability, but for which an efficient GAMD code exists, leading to

Theorem 1.4 (Existence of GAMDs breaking non-malleability - Informal). *For any non-malleable code \mathcal{C} there exists a family of tampering functions such that \mathcal{C} is malleable with respect to this family but there exists a GAMD for this family with constant rate and negligible error probability.*

We also show how to extend the construction of non-malleable codes for the class of bounded degree polynomials to *super non-malleable codes* in the split-state model (Dziembowski et al., 2010). The core observation behind this construction is that super non-malleable codes of Faust et al. (Faust et al., 2014b) can be de-randomised by embedding t -wise independent hash functions inside a plain non-malleable code which is appended to the resulting codeword.

Theorem 1.5 (Super non-malleability in two-state model - Informal). *In the two-state model there exists, for any $0 < \epsilon < 1$, an explicit ϵ -super non malleable code for the class of polynomials of degree bounded by d . The rate is $\frac{1}{\Theta(d^2)}$.*

A significant limitation of our results is that they only apply for tampering functions in one variable, while achieving corresponding deterministic results for multi-variate tampering classes seems considerably more challenging.

1.2 Related Work

Cabello et al. constructed AMD codes in the context of robust secret sharing (Cabello et al., 2002). The notion was made explicit by the works of (Cramer et al., 2008, Dodis et al., 2006) and some further applications provided including robust fuzzy extraction and message authentication codes with key manipulation security. In the former one wishes to guarantee recovery of a uniformly random key from biometric or other noisy data with the property that correctness is maintained under addition of errors up to some prior fixed bound even if the public parameters are compromised. In a similar vein the goal of the latter is to prevent forgery of message authentication tags even in the case that the adversary has algebraic manipulation access to the device storing the key. Other applications include robust information dispersal and anonymous message transmission (Cramer et al., 2008). Dziembowski et al. introduced the notion of non-malleable coding schemes and gave existential constructions for arbitrary tampering classes as well as efficient constructions in the random oracle (Dziembowski et al., 2010). Liu et al. constructed computationally secure non-malleable codes for split-state tampering in the CRS model (Liu and Lysyanskaya, 2012). Dziembowski et al. initiated the study of non-malleable codes from two-source extractors (Dziembowski et al., 2013). Aggarwal et al. (Aggarwal et al., 2014) and Chattopadhyay et al. (Chattopadhyay and Zuckerman, 2014) constructed explicit efficient non-malleable codes in the split-state model. Faust et al. constructed asymptotically optimal non-malleable codes for sufficiently small tampering classes in the CRS model (Faust et al., 2014b). Faust et al. constructed non-malleable codes secure against continual leakage (Faust et al., 2014a).

Although non-malleable cryptography is not the major focus of this work we show how to construct non-malleable codes from polynomial evasive GAMDs as well super non-malleable codes (Faust et al., 2014b) for this class in the two-state model.

2 PRELIMINARIES

We describe the preliminary tools and definitions to be used throughout this paper. We begin firstly by reviewing non-malleable codes (Dziembowski et al., 2010), secondly by stating some combinatorial results and finally, in Section 2.3, by stating our generalisation of classical algebraic manipulation detection codes (Cabello et al., 2002, Cramer et al., 2008, Dodis et al., 2006).

2.1 Non-Malleable Codes

We recall the notion of non-malleable codes for a class of tampering functions. Informally a non-malleable code is one which guarantees that after decoding either the original message is recovered or the message that is recovered is completely “unrelated” to the original.

Definition 1 (Non-Malleable Code (Dziembowski et al., 2010)). *Let \mathcal{F} be a family of tampering functions. For each $F \in \mathcal{F}$ and $s \in \{0, 1\}^k$, define the tampering experiment*

$$\text{Tamper}_s^F =: \left\{ \begin{array}{l} c \leftarrow \text{Enc}(s), \tilde{c} \leftarrow F(c), \tilde{s} = \text{Dec}(\tilde{c}) \\ \text{Output } \tilde{s}. \end{array} \right\}$$

defining a random variable over the randomness of the encoding function Enc . Say that a coding scheme (Enc, Dec) is non-malleable w.r.t. \mathcal{F} if for each $F \in \mathcal{F}$, there exists a distribution D_F over $\{0, 1\}^k \cup \{\perp, \text{same}^*\}$, such that, for all $s \in \{0, 1\}^k$, we have:

$$\text{Tamper}_s^F \approx \left\{ \begin{array}{l} \tilde{s} \leftarrow D_F \\ \text{Output } s \text{ if } \tilde{s} = \text{same}^*, \text{ and } \tilde{s} \text{ otherwise.} \end{array} \right\}$$

and D_F is efficiently samplable given oracle access to $F(\cdot)$.

Let \mathcal{F}_{bit} be the family of tampering functions that tamper every bit of a code-word of length n independently. Formally, \mathcal{F}_{bit} contains all functions $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$ defined by n functions $f_i : \{0, 1\} \rightarrow \{0, 1\}$, namely $f(c_1, \dots, c_n) = (f_1(c_1), \dots, f_n(c_n))$. Each f_i is an affine function on \mathbb{Z}_2 . We require the following proposition proved by (Dziembowski et al., 2010), concerning the existence of non-malleable codes against the family of bit-wise independent tampering functions with constant rate and negligible simulation error.

Lemma 2.1 (Theorem 4.2 (Dziembowski et al., 2010)). *For any $\delta > 0$ and $n \in \mathbb{N}$ there exist non-malleable codes w.r.t the family \mathcal{F}_{bit} , with block length n , message size $k \geq (.18 - \delta)n$ and simulation error $2^{-\Omega(n)}$. Moreover there is an efficient procedure which, given k and n , outputs a description of such a code with probability $1 - 2^{-\Omega(n)}$.*

We will also use the notion of super non-malleability (Faust et al., 2014b) in the split-state model (Dziembowski et al., 2010).

Definition 2 (Super Non-Malleability (Faust et al., 2014b)). *Let $\text{Enc} : \{0, 1\}^k \rightarrow \{0, 1\}^n$, $\text{Dec} : \{0, 1\}^n \rightarrow \{0, 1\}^k$ be a coding scheme and \mathcal{F} be a family of functions $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$. We say that the scheme is (\mathcal{F}, ϵ) -super non-malleable if for any $m_0, m_1 \in \{0, 1\}^k$ and any $f \in \mathcal{F}$, we have $\text{Tamper}_{m_0}^f \approx_\epsilon \text{Tamper}_{m_1}^f$ where:*

$$\text{Tamper}_m^f := \begin{cases} c \leftarrow \text{Enc}(x), c' = f(c) \\ \text{Output same}^* \text{ if } c' = c, \text{ output } \perp \text{ if } \text{Dec}(c') = \perp \\ \text{and else output } c'. \end{cases}$$

Theorem 2.1 ((Faust et al., 2014b)). *Let $\mathcal{H}_1 = \{h_1\}$ and $\mathcal{H}_2 = \{h_2\}$ be t -wise independent hashing families where $h_1 : \{0, 1\}^{v_1} \rightarrow \{0, 1\}^k$ and $h_2 : \{0, 1\}^{k+v_1} \rightarrow \{0, 1\}^{v_2}$. Then for any function family \mathcal{F} , consisting of functions $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$ there exists an (\mathcal{F}, ϵ) -super non-malleable code with probability $1 - p$ provided that*

$$\begin{aligned} t &\geq O(\log |\mathcal{F}| + n + \log(1/p)) \\ v_1 &> 3 \log(1/\epsilon) + 3 \log t + O(1) \\ v_2 &> v_1 + 3. \end{aligned}$$

2.2 Combinatorial Tools

We describe some combinatorial tools used in our constructions of GAMDs.

Definition 3 (Balanced Block Design (Colbourn and Dinitz, 2006)). *Let v, c, λ be a positive integers. For point set V a balanced block design is a multiset \mathcal{B} of blocks of points such that*

1. $|V| = v$
2. $|P| = c$ for each $P \in \mathcal{B}$
3. Each pair of points is a subset of exactly λ blocks

if $|\mathcal{B}| = v$ say that the (v, c, λ) -balanced block design is symmetric.

Definition 4 (Trace (Cramer et al., 2015)). *Let K and L be fields. Suppose that L is separable over K and $n := [L : K] > \infty$. Fix some algebraic closure \bar{L} of L . Let $\sigma_1, \dots, \sigma_n$ be the distinct K -embeddings of L into \bar{L} . The trace map $\text{Tr}_{L/K}$ for each $x \in L$ is:*

$$\text{Tr}_{L/K}(x) = \sum_{i=1}^n \sigma_i(x) \in K$$

Definition 5 (Difference Set (Colbourn and Dinitz, 2006)). *Let $(\mathcal{G}, +)$ be an additive abelian group of order v . A subset $D \subseteq \mathcal{G}$ is a (v, c, λ) -external difference set if $|D| = c$ and every non-zero element of \mathcal{G} has exactly λ representations as a difference $d - d'$ for $d, d' \in D$. If every non-zero element of \mathcal{G} has at most λ representations $d - d'$, say that D is a (v, c, λ) -bounded difference set.*

Definition 6 (Authentication Code (Stinson, 1990, 1994)). Let \mathcal{S} be a set of source states, \mathcal{K} a set of authentication keys and \mathcal{A} be a mapping $\mathcal{A} : \mathcal{S} \times \mathcal{K} \rightarrow \mathcal{T}$ where \mathcal{T} is a set of tags. Let Π be a probability distribution on \mathcal{K} . The probability of a successful substitution attack, with respect to family of substitution functions \mathcal{F} , is

$$p_{\mathcal{F}}^{\text{sub}} =: \max_{F \in \mathcal{F}, s \neq s' \in \mathcal{S}} \Pr_{K \leftarrow \Pi} [F(\mathcal{A}(s, K)) = \mathcal{A}(s', K)].$$

Lemma 2.2 (Schwartz-Zippel). Let K be a field and let $P \in K[x_1, \dots, x_n]$ where $(x_i)_{1 \leq i \leq n}$ are indeterminates. Let $S \subseteq K$ be a finite set and let $(u_i)_{1 \leq i \leq n}$ be selected independently and uniformly at random in S . Then

$$\Pr[P(u_1, \dots, u_n) = 0] \leq \frac{\deg(P)}{|S|}$$

Lemma 2.3 (Prime Number Theorem (Rose, 1994)). Let $\pi(x)$ denote the number of primes p which satisfy $2 \leq p \leq x$. Then

$$\lim_{x \rightarrow \infty} \pi(x) \cdot \frac{\ln(x)}{x} = 1$$

2.3 Generalised Algebraic Manipulation Detection Codes

In this section we define a code which is a generalisation of the classical algebraic manipulation detection coding schemes. The main difference is simply that we allow manipulation functions be general algebraic functions over a field, rather than the restriction to point additions on its group considered by (Cabello et al., 2002, Cramer et al., 2008). In this paper K will always be a finite field or number field (finite extension of the rationals), however below we allow K to be arbitrary for completeness.

Definition 7. Let K be a field with associated metric $d : K^2 \rightarrow \mathbb{R}^+ \cup \{0\}$. Let $\mathcal{G} := K$ and let \mathcal{F} be a family of algebraic tampering functions on \mathcal{G} . Let \mathcal{S} be a set of symbols. Let $\mathcal{E} : \mathcal{S} \rightarrow \mathcal{G}$ be a probabilistic encoding and $\mathcal{D} : \mathcal{G} \rightarrow \mathcal{S} \cup \{\perp\}$ be a deterministic decoding procedure such that $\Pr_{\mathcal{E}}[\mathcal{D}(\mathcal{E}(s)) = s] = 1$ for all $s \in \mathcal{S}$.

- The tuple $(\mathcal{E}, \mathcal{D})$ is an ϵ -generalised algebraic manipulation detection (GAMD) code if $\forall s \in \mathcal{S}, \forall F \in \mathcal{F} \Pr_{\mathcal{E}}[\mathcal{D}(F(\mathcal{E}(s))) \notin \{s, \perp\}] \leq \epsilon$.
- The tuple $(\mathcal{E}, \mathcal{D})$ is a weak ϵ -generalised algebraic manipulation detection code if $\forall F \in \mathcal{F} \Pr_{\mathcal{E}, s \in \mathcal{R}\mathcal{S}}[\mathcal{D}(F(\mathcal{E}(s))) \notin \{s, \perp\}] \leq \epsilon$.

Let $B_d(0, \delta)$ be the set of points at distance at most δ from $0_{\mathcal{G}}$. The (information) rate of a GAMD code is defined as $r = \lim_{\delta \rightarrow \infty} \frac{\log_2 |\mathcal{E}(\mathcal{S}) \cap B_d(0, \delta)|}{\log_2 |\mathcal{G} \cap B_d(0, \delta)|}$.

2.3.1 Families of Tampering Functions

In this paper we consider two classes of tampering functions on a GAMD $(\mathcal{E}, \mathcal{D})$ with co-domain $\mathcal{G} = \mathbb{F}_p^n$ for some prime p and positive integer n .

- **Point Additions:** let $\mathcal{F}_{\text{add}} = \{F_{\Delta}\}_{\Delta \in \mathcal{G}}$ where $F_{\Delta} := x \mapsto x + \Delta$ over \mathcal{G} .
- **Polynomial Functions:** let $\mathcal{F}_{\mathcal{P} \leq d} = \{F_{(\bar{a})}\}_{\bar{a} \in \mathcal{G}^{d+1}}$ where $F_{(\bar{a})} := x \mapsto \sum_{i=0}^d a_i x^i$ over \mathcal{G} .

2.4 Notation

Write $f = o(g)$ if $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$. Write $f = \Omega(n)$ if $\exists c > 0$ and $N_0 > 0$ such that for all $n > N_0$, $f(n) \geq c \cdot g(n)$. Let $e(\cdot)$ denote the real-valued exponential function. Let $\text{SD}(\cdot, \cdot)$ denote the statistical distance. For discrete probability distributions with outcome space \mathcal{X} , $\text{SD}(P_0, P_1) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P_0(x) - P_1(x)|$. For probability distributions P_0 and P_1 let $D(P_0 \| P_1)$ denote the KL-divergence. Pinsker's inequality states that $D(P_0 \| P_1) \geq 2\text{SD}(P_0 \| P_1)^2$. Say that random variables X_1, \dots, X_n are k -wise independent if $\Pr[X_{i_1} = a_1, \dots, X_{i_k} = a_k] = \prod_{j=1}^k \Pr[X_{i_j} = a_j]$ for all $\{i_1, \dots, i_k\} \subseteq [1, n]$. A function is algebraic iff it is the root of a polynomial equation. Let \mathbb{Q} be the set of rationals. For field K , let $\mathcal{P}_{\leq d}$ be the space of univariate polynomials of degree at most d over K . For even integer n denote by I_n , the subset of permutations on n objects consisting of involutions with no fixed points. The independence number of a finite graph G is the size of the largest complete graph in the edge complement of G .

2.5 Tail Bounds on Sums of Dependent Variables

Lemma 2.4 (Multiplicative Chernoff Bound). *Let $\{X_i\}_{1 \leq i \leq n}$ be a sequence of independent random variables such that $0 \leq X_i \leq 1$, $E[X_i] = p$ for $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$ and $\mu = E[X] = np$. Fix $0 < \delta < 1$. Then*

$$\Pr[X < \mu(1 - \delta)] \leq e\left(-\frac{\delta^2 \mu}{2}\right)$$

$$\Pr[X > \mu(1 + \delta)] \leq e\left(-\frac{\delta^2 \mu}{3}\right)$$

Lemma 2.5 (Theorem 1.12 (Pelekis and Ramon, 2015)). *Let $\{X_i\}_{1 \leq i \leq n}$ be a sequence of k -wise independent random variables such that $0 \leq X_i \leq 1$, $E[X_i] = p$ for $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$ and $\mu = E[X] = np$. Fix $\delta > 0$. Then*

$$\Pr[X > \mu(1 + \delta)] \leq \frac{1}{(p - p^2)^{n-k}} e(-nD(p(1 + \delta) \| p))$$

Lemma 2.6 (Theorem 1.14 (Pelekis and Ramon, 2015)). *Let $G = (V, E)$ be a finite graph with vertices v_1, \dots, v_n and let α be its independence number. To each v_i , $i = 1, \dots, n$ we associate*

a Bernoulli 0/1 random variable B_i , such that $\Pr[B_i = 1] = p$. Suppose that each random variable $B_i, i = 1, \dots, n$ is independent of the set $\{B_j : (v_i, v_j) \notin E\}$. Let $0 < \delta < 1$ be a constant and $t = np(1 + \delta)$. Then

$$\Pr\left[\sum_{i=1}^n B_i \geq t\right] \leq p^\alpha \cdot e\left(-\frac{\delta^2 n}{2}\right) \cdot 2^n$$

3 CONSTRUCTIONS

In this section we review some constructions for GAMD codes against the class of tampering functions corresponding to point additions and also polynomial functions. Our results show that efficient GAMDs (i.e, one ones with constant rate and low error probability) exist for the former class, while for the latter, the rate degrades quadratically in the degree of the function. For the class of point additions, we present two constructions of GAMDs based upon difference sets. Our first can be seen as a specific instantiation of the AMD codes in Section 4.1 (Cabello et al., 2002). Our second which is based upon the probabilistic method allows the construction of GAMDs for a broader class of functions.

3.1 Point Additions

Cabello et al. (Cabello et al., 2002) constructed a difference set in $\mathbb{F}_{p^l} \times \mathbb{F}_{p^k}$ from any surjective map $\phi : \mathbb{F}_{p^l} \rightarrow \mathbb{F}_{p^k}$. An efficient instantiation of ϕ for arbitrary p can be found using the field trace (Definition 4). Using this construction we can build a weak-GAMD with rate $1 - o(1)$ and arbitrarily low error probability, described in Lemma 3.2.

Lemma 3.1. (Cabello et al., 2002) *Let p be an odd prime and l and k be positive integers such that $l \equiv 0 \pmod{k}$. Let $(\mathcal{G}, +)$ be the product of groups, $\mathbb{F}_{p^l} \times \mathbb{F}_{p^k}$ under addition. Define*

$$D_{k,l} = \{(\alpha, \phi_{\mathbb{F}_{p^l}/\mathbb{F}_{p^k}}(\alpha^2)) : \alpha \in \mathbb{F}_{p^l}\} \subseteq \mathcal{G}$$

Then $D_{l,k}$ is a (p^{l+k}, p^l, p^{l-k}) -external difference set.

Lemma 3.2. *For a prime p and positive integer n let $\mathcal{G} = \mathbb{F}_p^n$. Then there exists a explicit weak (p^{-1}) -GAMD code with respect to the family of point additions, \mathcal{F}_{add} , on \mathcal{G} , with efficient encoding and decoding procedure and rate $1 - o(1)$.*

Proof. See full version. □

3.1.1 A New Construction

We note that so far the constructions of GAMD codes against the class of point additions have followed a similar recipe to the constructions of AMD codes presented in (Cabello et al., 2002,

Cramer et al., 2008). In this section we present a new construction for this class based upon the probabilistic method. The proofs of the following are deferred to the full version.

Lemma 3.3. *Let \mathcal{G} be an abelian group of order n where n is even. Let $0 \leq c < 1$ be arbitrary. Let $I'_n \subset I_n$ be of polynomial size. Then there exists a subset $S \subset \mathcal{G}$ and maps $\mathcal{E} : [|S|] \rightarrow \mathcal{G}$ and $\mathcal{D} : \mathcal{G} \rightarrow [|S|]$ which define a weak n^{c-1} -GAMD with respect to the set I'_n . The rate is ρ is $c - o(1)$. The sampling error is $e(-\frac{1}{4}n^\rho) + |I'_n| \cdot e(-2n^{2\rho-1})$.*

Corollary 3.1. *Let $G = (\mathbb{F}_n, +)$ where n is an arbitrary power of two. Then there exists a weak $(n^{-1/2})$ -GAMD with respect to the family \mathcal{F}_{add} , with rate $\frac{1}{2} - o(1)$. The sampling error is $e(-\frac{1}{4}n^{1/2}) + n^{-0.1}$.*

We remark that the parameters achieved by Lemma 3.3 are essentially optimal - matching those of classical parameter sets modulo two (Colbourn and Dinitz, 2006). In the full version we also prove the following result concerning the class \mathcal{F}_{add} over the cartesian power of a field K corresponding to the finite extensions of K under addition.

Lemma 3.4. *Let $(\mathcal{E}', \mathcal{D}')$ be a weak γ -GAMD over field $(K, +)$ for the class \mathcal{F}_{add} with rate ρ' . Then there exists $(\mathcal{E}, \mathcal{D})$, a weak γ -GAMD for \mathcal{F}_{add} over $(K^m, +)$, with rate $\rho := \rho'$ and $\gamma = 1 - (1 - \gamma')^m$.*

3.2 Polynomial Functions

In this section we show to construct explicit GAMDs secure against the class of all polynomials of finite degree modulo a prime, extending the constructions in (Aggarwal, 2015, Aggarwal et al., 2014). We first present an informal overview of our construction, while the construction itself is described in section 3.2.1.

Our Construction In A Nutshell Aggarwal (Aggarwal, 2015) constructed codes secure against affine functions by constructing affine-evasive sets modulo a prime. The construction uses the reciprocals of all primes less than some inverse power in the underlying modulus. Fix an affine function F and let the reciprocal primes in its domain be denoted a_i and the primes in its range be denoted b_i . In that case an explicit bi-variate quartic relation is derived on the a_i and b_i (Aggarwal, 2015). We follow this principle but instead use Lagrange interpolation to derive a (cyclically) symmetric relation on the a_i and b_i . Unfortunately the setting $d > 1$ necessitates some changes. Firstly there is no longer symmetry between the a_i and b_i which appears to be unique to the affine setting only. This implies divisibility relations appear possible only from the b_i (primes in the range of the polynomial). We are able to utilise these at slight expense (roughly $O(\log \log k)$ in bit-length) by an additive combinatorics-like construction of a set of primes with the property that no difference of elements of the set is divisible by another element. We believe this construction, which Lemma 3.5 is devoted to, may be of independent interest.

3.2.1 Construction of Polynomial Evasive GAMDs

Lemma 3.5. *For any positive integer N there exists a positive integer B , so that N primes lie in the interval $[0, B]$ and such that no prime divides the difference of two others for $B = O(N \ln^{1+o(1)} N)$.*

Proof. By Lemma 2.3 we can find $\Theta(\frac{B}{\ln B})$ primes q_i in the interval $(B/2, B]$. Suppose $q_i \mid q_j - q_k$ for some $q_i \neq q_j \neq q_k$. Then $B/2 < q_i \leq |q_j - q_k| \leq B/2$ which is a contradiction. \square

For positive integer N , denote the above set D_N .

Theorem 3.2. *Let p a prime of k bits. There exists an explicit weak ϵ -GAMD secure against the class $\mathcal{F}_{\mathcal{P} \leq d}$ modulo p of rate $2/\Theta(d^2)$ and error probability $\epsilon = \frac{O(k)}{d} \cdot 2^{-\frac{k}{\Theta(d^2)}}$ for any positive integer d .*

In the full version we also prove

Theorem 3.3. *Let p be a prime. There exists some constant c so that for any $0 < \epsilon < 1$ there exists a ϵ -non-malleable code (Enc, Dec) for the class $\mathcal{F}_{\mathcal{P} \leq d}$ where $\text{Enc} : \mathbb{Z}_T \rightarrow \mathbb{F}_p$ and $\text{Dec} : \mathbb{F}_p \rightarrow \mathbb{Z}_T$ whenever $p > (\frac{T}{\epsilon})^{c \cdot d^2}$.*

We remark that Theorem 3.2 extends to all finite centred Laurent expansions, i.e., *two-sided polynomial expressions about zero*, as well as to finite fields with similar parameters.

4 A WEAK GAMD TO GAMD TRANSFORMATION

In this section we present a sufficient result for transforming any weak GAMD to a GAMD following a similar idea to that presented in Section 4 (Cramer et al., 2008). Our main result here is Lemma 4.1 which states that if the classes of tampering functions can be represented by a set of polynomials in one or more variable of bounded degree $d \ll |\mathcal{K}|$ then any weak GAMD for this family can be transformed to a GAMD. In particular this implies asymptotically efficient GAMDs for the class of polynomial functions with negligible error probability. The proofs of the following are deferred to the full version.

Prop 4.1. *Suppose that $(\mathcal{E}', \mathcal{D}')$ is a weak ϵ' -GAMD with respect to \mathcal{F} where $\mathcal{E}' : \mathcal{S}' \rightarrow \mathcal{G}'$. Let $\mathcal{A} : \mathcal{S} \times \mathcal{S}' \rightarrow \mathcal{T}$ be an authentication code. Let $\mathcal{G} = \mathcal{S} \times \mathcal{G}' \times \mathcal{T}$. Define $\mathcal{E} : \mathcal{S} \rightarrow \mathcal{G}$ by $\mathcal{E}(s) = (s, \mathcal{E}'(k), \mathcal{A}(s, k))$, where $k \in_R \mathcal{S}'$. Define $\mathcal{D} : \mathcal{G} \rightarrow \mathcal{S} \cup \{\perp\}$ by $\mathcal{D}(s, c', \tau) = s$ iff $\mathcal{D}'(c') \neq \perp$ and $\tau = \mathcal{A}(s, \mathcal{D}'(c'))$. Then $(\mathcal{E}, \mathcal{D})$ is an ϵ -GAMD with respect to \mathcal{F} where $\epsilon = \epsilon' + p_{\mathcal{F}}^{\text{sub}}$.*

Lemma 4.1. *Let ℓ be an arbitrary positive integer and K be a field. Let $\mathcal{K} \subseteq K^2$ be a finite set and $\mathcal{A} : \mathcal{S} \times \mathcal{K} \rightarrow \mathcal{T}$ be the message authentication code defined by $\mathcal{A}((s_1, \dots, s_\ell), (x, y)) = \sum_{i=1}^{\ell} s_i x^i + y$. Then $p_{\mathcal{F}_{\mathcal{P} \leq d}}^{\text{sub}} \leq \frac{\ell d}{|\mathcal{K}|}$.*

Corollary 4.1. *For any $n \in \mathbb{N}$ and large enough prime p there exists an ϵ -GAMD of block length n with respect to the family $\mathcal{F}_{\mathcal{P} \leq d}$ over \mathbb{F}_p where $\epsilon = 2^{-n/\Theta(d^2)}$. The rate is $1 - o(1)$.*

5 SEPARATIONS

We describe some separations regarding non-malleable codes and GAMDs generalising separations noted in previous works (Dziembowski et al., 2013, 2010, Faust et al., 2014b). Although non-malleable codes have already proved a valuable digression from the classical notion of error correction and detection, here we provide evidence that GAMD codes provide a strengthening of classical algebraic manipulation detection distinct to that provided by non-malleable cryptography. Specifically we are able to prove (Theorem 5.2) that any non-malleable code can be broken by some tampering family for which a GAMD with high rate and low error probability exists. This family actually corresponds to a re-coding functionality in which a code-word is decoded, one is added to the message which is then again encoded, so is a natural candidate for this task. On the negative side, however, we show that for at least one family of tampering functions, non-malleable codes exist but GAMDs do not. We also prove that in the two-state model super non-malleable codes exist for arbitrary ϵ against the class $\mathcal{F}_{\mathcal{P} \leq d}$ with inverse polynomial rate. The proofs of the following are deferred to the full version.

Theorem 5.1. *Some family of tampering functions \mathcal{F} exists for which for any $n \in \mathbb{N}$, non-malleable codes of block length n exist with constant rate and simulation error $2^{-\Omega(n)}$, but ϵ -GAMD codes with constant rate do not exist, for any choice of non-negligible (in block-length) ϵ .*

Theorem 5.2. *For any non-malleable code \mathcal{C} of block length n there exists a family of tampering functions $\mathcal{F}_{\mathcal{C}}$ such that \mathcal{C} is malleable with respect to $\mathcal{F}_{\mathcal{C}}$ but there exists a $(2^{-\Omega(n)})$ -GAMD code \mathcal{C}' with respect to $\mathcal{F}_{\mathcal{C}}$ with rate $r \cdot O(1) - o(1)$, where r is the rate of \mathcal{C} .*

Theorem 5.3. *In the two-state model there exists, for any $0 < \epsilon < 1$, an explicit ϵ -super non malleable code for the class $\mathcal{F}_{\mathcal{P} \leq d}$ with negligible sampling error. The rate is $\frac{1}{\Theta(d^2)}$.*

6 CONCLUSION

We have defined a generalisation of algebraic manipulation detection codes to facilitate detection of tampering by algebraic functions over a field. We have demonstrated explicit constructions of these codes for the families of point additions and polynomial functions and randomised constructions for some broader classes over finite fields. In future work it would be interesting to extend these constructions as well as to investigate applications of these codes.

ACKNOWLEDGEMENTS

The author would like to thank anonymous reviewers and Chaitanya Rao for helpful comments and suggestions.

REFERENCES

- Aggarwal, D. (2015). Affine-evasive sets modulo a prime. *Inf. Process. Lett.*, 115(2):382–385.
- Aggarwal, D., Dodis, Y., and Lovett, S. (2014). Non-malleable codes from additive combinatorics. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing, STOC '14*, pages 774–783, New York, NY, USA. ACM.
- Cabello, S., Padró, C., and Sáez, G. (2002). Secret sharing schemes with detection of cheaters for a general access structure. *Des. Codes Cryptography*, 25(2):175–188.
- Chattopadhyay, E. and Zuckerman, D. (2014). Non-malleable codes against constant split-state tampering. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 306–315.
- Colbourn, C. J. and Dinitz, J. H. (2006). *Handbook of Combinatorial Designs, Second Edition (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC.
- Cramer, R., Damgård, I. B., and Nielsen, J. B. (2015). *Secure Multiparty Computation and Secret Sharing*. Cambridge University Press, New York, NY, USA, 1st edition.
- Cramer, R., Dodis, Y., Fehr, S., Padró, C., and Wichs, D. (2008). Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In Smart, N., editor, *Advances in Cryptology – EUROCRYPT 2008: 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*, pages 471–488, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Dodis, Y., Katz, J., Reyzin, L., and Smith, A. (2006). Robust fuzzy extractors and authenticated key agreement from close secrets. In *Proceedings of the 26th Annual International Conference on Advances in Cryptology, CRYPTO'06*, pages 232–250, Berlin, Heidelberg. Springer-Verlag.
- Dziembowski, S., Kazana, T., and Obremski, M. (2013). Non-malleable codes from two-source extractors. In Canetti, R. and Garay, J. A., editors, *Advances in Cryptology – CRYPTO 2013*, pages 239–257, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Dziembowski, S., Pietrzak, K., and Wichs, D. (2010). Non-malleable codes. In *ICS*, pages 434–452.
- Faust, S., Mukherjee, P., Nielsen, J. B., and Venturi, D. (2014a). Continuous non-malleable codes. In Lindell, Y., editor, *Theory of Cryptography*, pages 465–488, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Faust, S., Mukherjee, P., Venturi, D., and Wichs, D. (2014b). Efficient non-malleable codes and key-derivation for poly-size tampering circuits. In Nguyen, P. Q. and Oswald, E., editors, *Advances in Cryptology – EUROCRYPT 2014*, pages 111–128, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Kiayias, A., Liu, F.-H., and Tselekounis, Y. (2016). Practical non-malleable codes from 1-more extractable hash functions. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16*, pages 1317–1328, New York, NY, USA. ACM.

- Liu, F.-H. and Lysyanskaya, A. (2012). Tamper and leakage resilience in the split-state model. In Safavi-Naini, R. and Canetti, R., editors, *Advances in Cryptology – CRYPTO 2012*, pages 517–532, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Pelekis, C. and Ramon, J. (2015). Hoeffding’s inequality for sums of weakly dependent random variables. <https://arxiv.org/abs/1507.06871>.
- Rose, H. E. (1994). *A Course in Number Theory, Second Edition*. Oxford University Press.
- Stinson, D. R. (1990). The combinatorics of authentication and secrecy codes. *J. Cryptol.*, 2(1):23–49.
- Stinson, D. R. (1994). Universal hashing and authentication codes. *Designs, Codes and Cryptography*, 4(3):369–380.

Hierarchical Twin-Schnorr Identity-Based Identification Scheme

Apurva Kiran Vangujar^{*1}, Ji-Jian Chin¹, Syh-Yuan Tan², and Tiong-Sik Ng²

¹*Faculty of Engineering, Multimedia University, Malaysia*

²*Faculty of Information Science and Technology, Multimedia University, Malaysia*

*E-mail: apurva710@gmail.com, jjchin@mmu.edu.my,
sytan@mmu.edu.my, ng.tiong.sik@gmail.com*

ABSTRACT

In 2015, Chin et al. proposed an extension to the Schnorr IBI scheme using two secret keys to tighten the security based on the discrete logarithmic assumption, namely the Twin-Schnorr IBI. Twin-Schnorr IBI works without pairing protocols and this helps to increase the efficiency of the scheme alongside strengthen its security. In this paper, we propose the Hierarchical Identity-based identification (HIBI) version of the Twin-Schnorr IBI scheme, which we refer to as the Hierarchical Twin-Schnorr IBI. Compared to conventional IBI scheme, HIBI is able to reduce the burden of a public key generator (PKG) to handle large number of users.

Keywords: Security attacks, Identity-based identification scheme (IBI), Discrete logarithmic assumption, Twin-Schnorr IBI scheme.

1 INTRODUCTION

1.1 Identification Scheme

The public key cryptography uses the recipient's public key for encryption and the recipient's private key for decryption to recover the original message. The public key cryptography is further sub-divided into digital signature schemes and identification schemes. An identification scheme consists of two parties, namely prover and a verifier in order to perform a challenge response protocol. An identification scheme allows a prover to prove himself to a verifier without revealing any information about himself. The traditional cryptographic scheme which includes identification schemes, requires the use of certificates issued by a certificate authority (CA) in order to authenticate user's public key. Maintaining certificates in large numbers in itself is a major issue.

Identification schemes first proposed in (Shamir, 1984) was built based on three-move protocol using zero-knowledge proof results into higher efficiency. In Identity-based cryptography coined by Shamir, the certificate requirement is abolished by replacing the public key with an identity string Fiat and Shamir (1986). It is the simplest form of cryptographic primitive without relying on certificates. Conventional IBI schemes only allow single user interaction with the verifier.

1.2 Related Work

In 1989, Schnorr described the first scheme based on the discrete logarithm assumption and it is particularly suited for the smart cards. The key generation algorithm is faster and more secure Shamir (1984), proposed an efficient algorithm to pre-process the exponentiation of random numbers.

Boneh and Franklin (2003) proposed the first identity-based encryption scheme, which lead to the booming of identity-based cryptography. Later years, IBI schemes were more secure and efficient formalized in Bellare et al. (2009). Later, Tan et al. (2011) proposed a variant of Schnorr IBI scheme and direct proof with tight security reduction. He described the security against impersonator under passive, active and concurrent attack based on the Decisional Diffie Hellman (DDH) assumption in the random oracle model.

Shortly after that, Chin et al. (2009) extended the Schnorr IBI scheme into the Twin-Schnorr IBI scheme. He proposed to generate two secret keys in key generation algorithm. The key generation algorithm of original Schnorr IBI gives output pair of the secret key and a public key. Henceforth, it tightens the security for twin-Schnorr IBI scheme and more efficient in compare to Schnorr IBI scheme as we have one more extra secret key Tan et al. (2011). The first idea of Hierarchical identity-based encryption (HIBE) was first proposed by Horwitz and Lynn (2002), where they proposed two level HIBE. Inspired by HIBE, Chin et al. (2015) introduced Twin-Schnorr IBI scheme.

This paper focuses on the Hierarchical Twin-Schnorr IBI scheme and it's security proof of passive, active and concurrent attack respectively. HIBI has root PKG as the first-level and n lower-level PKG where n is defined by users. Each node is connected to other node and communicates with each other by three move protocol. The advantages of HIBI are listed as the following.

1. It is an efficient as there is no database needed for identities.
2. It has improved scalability.
3. Solves key escrow problem with delegated key feature.

In Fiat and Shamir (1986), a framework to construct IBI schemes for traditional crypto system was proposed. In Shamir (1984), a basic introduction to identification scheme is defined. The security is stated practically in Bellare et al. (2009) for IBI. The conversion to digital signa-

ture into IBI along with security attack and efficiency analysis is done by Barapatre and Rangan (2013), Kurosawa and Heng (2004).

The security of Twin-Schnorr IBI is more tighter compared to the previously proposed IBI schemes. Twin-Schnorr IBI is highly efficient as it has less communication cost compared to other IBI schemes developed.

In 2012, Fujioka et al. (2012) defined OR-proof technique in identification protocol and enhanced security in static attack model. Fujioka et al. (2014) proposed a purely certificate-less based HIBI scheme which is tied to the RSA and CDH assumption. Fujioka's IBI scheme is proven secure in the standard security model whereas Chin et al. (2009) is proven efficient with random oracle. HIBI without random oracle has increased communication cost and key size as compared to HIBI with random oracle. HIBI random oracle is the extension to IBI follows CDH and OMCDH. It is concluded that HIBI schemes using the random oracle are more efficient. Chin et al. (2015) .

In this paper, we propose a Hierarchical version of the of Twin-Schnorr IBI scheme. The Hierarchical Twin-Schnorr IBI scheme has a Public Key Generator (PKG) which will distribute the secret key once and then partially creates multiple PKG.

1.3 Organization

The paper is organized as follows. In Section 2, we begin with some preliminaries including assumptions, groups, and security definitions for IBI schemes. In Section 3, we define the Hierarchical Twin-Schnorr IBI scheme. We define the security proof against impersonation under active and concurrent attack for the Hierarchical Twin-Schnorr IBI scheme in Section 4. In Section 5, we calculate the efficiency analysis of Hierarchical Twin-Schnorr IBI scheme in comparison with other IBI schemes. We conclude this paper in Section 6.

2 PRELIMINARIES

2.1 Discrete Logarithm Assumption

We adopt the definition of the discrete logarithm assumption from Kurosawa and Heng (2004), Bellare and Palacio (2002) Ioannidis et al. (2005) follows:

Definition 2.1. *Let G be a finite cyclic group of order n . Let α be a generator of G , and let $\beta \in G$. The discrete logarithm of β to the base α , denoted $\log_{\alpha}\beta$, is the unique integer x , $0 \leq x \leq n - 1$, such that $\beta = \alpha^x$.*

2.2 Formal Definition of IBI Schemes

Definition 2.2. An identity-based identification (IBI) scheme is based on the four probabilistic algorithms.

$$ID = (\mathcal{S}, \mathcal{E}, \mathcal{P}, \mathcal{V})$$

- **Key Setup** (\mathcal{S}). It takes the input as 1^k and generates output as (param, masterkey).
- **Extract** (\mathcal{E}). An extract oracle is used to extract the private key. Input (masterkey, ID) and returns the private key d .
- **Identification Protocol** (\mathcal{P} and \mathcal{V}). In this phase, the prover \mathcal{P} and the verifier \mathcal{V} communicates with each other. \mathcal{P} takes input as (param, ID, d) whereas the \mathcal{V} takes input as (param, ID). \mathcal{P} and \mathcal{V} communicates with each other with the help of (CMT, CH, RSP) and gives output in boolean decision 0 (rejects) or 1 (accepts). The canonical protocol acts in four steps as following :
 1. \mathcal{P} sends commitment (CMT) to \mathcal{V} .
 2. \mathcal{V} provides challenge (CH) which is randomly chosen.
 3. \mathcal{P} calculates the response (RSP) to \mathcal{V} as per challenge.
 4. \mathcal{V} verifies (param, ID, CMT, CH, RSP) is DH tuple.

2.3 HIBI Schemes

Definition 2.3. An HIBI scheme is based on the four probabilistic algorithms. Gentry and Silverberg (2002) Chin et al. (2009)

$$ID = (\mathcal{S}, \mathcal{E}, \mathcal{P}, \mathcal{V})$$

- **Root Setup** (\mathcal{S}). This algorithm selects the random generator and a secret key and generates the output pair of param and secret key.
 1. **Lower Level Setup.** This level in which all identities at lower level sets random parameter keeping it secret.
- **Extract** (\mathcal{E}). For any identity, it calculates user secret key with the help of ancestor secret key.
- **Identification Protocol** (\mathcal{P} , \mathcal{V}). In this phase, the prover \mathcal{P} and the verifier \mathcal{V} communicate in three steps as following.
 1. \mathcal{P} chooses random variable to calculate value and send to \mathcal{V} .
 2. \mathcal{V} generates the random challenge and forwards to \mathcal{P} .
 3. \mathcal{P} accepts the challenge and generates response base on the challenge.
 4. \mathcal{V} accepts if and only if, it verifies the final equation.

An impersonator focuses to impersonate an honest user. The following two section states the types of adversary.

- A passive adversary. This is the attack where an adversary obtains the communication transcript between the real prover and a verifier. An adversary only can steal the information but doesn't affect the communication line between the prover and the verifier. This is the weakest attack.
- An active adversary and concurrent adversary. An adversary can directly communicate with the prover playing the role of a cheating verifier actively. The adversary in the active attack can drop, change and configure the information. It threatens authentication and integrity of data. An adversary can concurrently communicate with communication protocol the prover playing the role of the cheating verifier and an adversary can do changes in between ongoing process Katz and Lindell (2014).

We adopt the security model for IBI scheme from Chin et al. (2009). An impersonation attack between an impersonator \mathcal{I} and a challenger \mathcal{C} is described as a two-phased game as follows:

1. **Setup (S).** \mathcal{C} takes input 1 and runs algorithm S . The result of system parameters mpk is given to \mathcal{I} while msk is kept to itself.
2. **Phase 1: Learning Phase.** \mathcal{I} issues some extract queries ID_i to \mathcal{C} . \mathcal{C} responds by running the extract algorithm to generate and return the private key usk corresponding to the identity ID_i to \mathcal{I} . The queries may be asked adaptively. \mathcal{I} issues transcript queries for passive attacks or requests to act as a cheating verifier corresponding to some ID_i for active/concurrent attacks.
3. **Phase 2: Impersonation Phase.** Finally, outputs a challenge identity ID which it wishes to impersonate whereby \mathcal{I} now acts as a cheating prover to convince the verifier \mathcal{C} based on information gathered in Phase 1. \mathcal{I} wins the game if it is successful in convincing the verifier.

2.4 Security Model for HIBI Schemes

We describe the security of a HIBI scheme with the following game between an impersonator \mathcal{I} and a challenger \mathcal{C} . Chin et al. (2009)

1. **Setup (\mathcal{S}).** The challenger first takes in a security parameter 1^k and gives the resulting $params$ to the \mathcal{I} . It keeps $rlmsk$ root-level master secret key to itself.
2. **Phase 1.** \mathcal{I} can issue queries (q_i, \dots, q_m) where q_i is one of:
 - (a) **Extract Key Query (\mathcal{E}).** Upon being queried with the public key of ID_i , returns usk_i to \mathcal{I} .
 - (b) **Transcript/Identification Query (\mathcal{P} and \mathcal{V}).** For passive \mathcal{I} , \mathcal{C} responds with a transcript for the interaction between the prover and a verifier. For active/concurrent, \mathcal{C} acts as the prover while \mathcal{I} takes the role of a cheating verifier.

3. **Challenge (\mathcal{C})**. \mathcal{I} outputs $ID^* \neq ID_i$ wishes to impersonate. ID^* is the targeted identity by impersonator among (ID_1, \dots, ID_i) .
4. **Phase 2**.
 - (a) **Extract Key Query (\mathcal{E})**. \mathcal{I} can continue to query the private keys of ID_i as long as ID_i is not an ancestor of $ID^* \neq ID_i$.
 - (b) **transcripts/Identification Query (\mathcal{P} and \mathcal{V})**. \mathcal{I} can continue to query either transcripts for passive \mathcal{I} or identification interactions for active/concurrent \mathcal{I} for ID^* or any ancestor of ID^* .
5. **Impersonation**. \mathcal{I} takes the role of the cheating prover and tries to convince the verifier. \mathcal{I} wins the game if it succeeds in convincing the verifier to accept with non-negligible probability.

Definition 2.4. We say an HIBI scheme is $(t_{HIBI}, q_{HIBI}, \varepsilon_{HIBI})$ -secure under passive or active/concurrent attacks if for any passive/active/concurrent \mathcal{I} who runs in time t_{HIBI} , $Pr[\mathcal{I} \text{ can impersonate}] \leq \varepsilon_{HIBI}$, where \mathcal{I} can make at most q_{HIBI} extract queries and transcripts/Identification Query.

3 THE HIERARCHICAL TWIN-SCHNORR IBI SCHEME

The Hierarchical Twin-Schnorr IBI scheme which is based on the Twin-Schnorr IBI scheme by Chin et al. (2015).

Root level consists of ID_0 identity. The hierarchy proceed for $level_1$ having identities $(ID_1, ID_2, \dots, ID_i, \dots, ID_m)$ where m represent the last identity of that level. ID_i is the targeted identity which can exist in a such a way that $(ID_1 \leq ID_i \geq ID_m)$. The construction of the Hierarchical Twin-Schnorr IBI scheme is having algorithms $(\mathcal{S}, \mathcal{E}, \mathcal{P}, \mathcal{V})$ are as follows Chin et al. (2009).

1. **Key Setup (\mathcal{S})**. It takes 1^k where k is the security parameter and generates \mathbb{G} the group of order q . It picks random generators $g_1, g_2 \in \mathbb{G}$ and two random integers $x_1, x_2 \in \mathbb{Z}_q$. It sets $X = g_1^{-x_1} g_2^{-x_2}$. It chooses a hash function $H : (0, 1)^* \times \mathbb{G} \times \mathbb{G} \Rightarrow \mathbb{Z}_q$. It publishes pair of (mpk, msk) where $mpk = \langle \mathbb{G}, q, g_1, g_2, X \rangle$ and $msk = \langle x_1, x_2 \rangle$.
2. **Extract (\mathcal{E})**. For ID_0 root level, (mpk, msk, ID_0) is the input. It calculates $R = g_1^{x_1} g_2^{x_2}$ and sets $\alpha_0 = \mathbb{H}(ID_0, R, X)$. later, It picks two random integers $r_{0,1}, r_{0,2} \in \mathbb{Z}_q$ and calculates $S_{0,1} = r_{0,1} + x_1 \alpha_0$, $S_{0,2} = r_{0,2} + x_2 \alpha_0$. Finally, it sets $usk_{ID_0} = \langle S_{0,1}, S_{0,2}, \alpha_0 \rangle$. It passes usk_{ID_0} to next level.

For ID_1 level 1, It picks two random integers $r_{1,1}, r_{1,2} \in \mathbb{Z}_q$ and to calculate $(S_{1,1}, S_{1,2})$ where $S_{1,1} = r_{1,1} + \alpha_1 + S_{0,1}$ and $S_{1,2} = r_{1,2} + \alpha_1 + S_{0,2}$, it uses the $(S_{0,1}, S_{0,2})$ as ancestor user secret key. Therefore, $usk_{ID_1} = \langle S_{1,1}, S_{1,2}, \alpha_1 \rangle$.

For ID_i level i , it takes input $mpk = \langle \mathbb{G}, q, g_1, g_2, X \rangle$, $usk_{ID_{i-1}} = \langle S_{i-1}, S_{i-2} \rangle$ and user identity string (ID_0, \dots, ID_i) . It picks two random integers $r_{i,1}, r_{i,2} \in \mathbb{Z}_q$, calculates $V_i = g_1^{r_{i,1} + S_{i-1,1}} g_2^{r_{i,2} + S_{i-1,2}}$ and sets $\alpha_i = \mathbb{H}(ID_0 || \dots || ID_i, V_i, X)$. Next, calculates $S_{i,1} = r_{i,1} + \alpha_i + S_{i-1,1}$ and $S_{i,2} = r_{i,2} + \alpha_i + S_{i-1,2}$ and sets $usk_{ID_i} = \langle S_{i,1}, S_{i,2}, \alpha_i \rangle$.

3. **Identification Protocol** (\mathcal{P} and \mathcal{V}) in which prover takes in mpk, ID_i and usk_{ID_i} while \mathcal{V} takes in mpk and ID_i . They run an identification protocol as follows.

- \mathcal{P} begins by picking two random integers $y_{i,1}, y_{i,2} \in \mathbb{Z}_q$ and sets $Y = g_1^{y_{i,1}} g_2^{y_{i,2}}$. \mathcal{P} additionally sets $V_i = g_1^{S_{i,1}} g_2^{S_{i,2}} X^{\alpha_i}$ and sends Y, V_i to \mathcal{V} .
- \mathcal{V} picks a random challenge $c \in \mathbb{Z}_q$ and sends it to \mathcal{P} .
- \mathcal{P} responds by setting $z_{i,1} = y_{i,1} + cS_{i,1}$ and $z_{i,2} = y_{i,2} + cS_{i,2}$ and sends $z_{i,1}, z_{i,2}$ to \mathcal{V} as it's response.

\mathcal{V} calculates and accepts if the following equation holds for each i :

$$g_1^{z_{i,1}} g_2^{z_{i,2}} = Y \left(\frac{V_i}{X^{\alpha_i}} \right)^c$$

where $\alpha'_i = H(ID_i, V_i, X)$ VERIFY can calculate $\alpha'_i = H(ID_i, V_i, X)$ by itself since

$$\begin{aligned} g_1^{S_{i,1}} g_2^{S_{i,2}} X^{\alpha_i} &= g_1^{r_{i,1} + \alpha_i} g_2^{r_{i,2} + \alpha_i} g_1^{-\alpha_i} g_2^{-\alpha_i} \\ &= g_1^{r_{i,1}} g_2^{r_{i,2}} \\ &= R \end{aligned}$$

The correctness of the identification protocol can be proven as such:

$$\begin{aligned} Y \left(\frac{V_i}{X^{\alpha_i}} \right)^c &= g_1^{y_{i,1}} g_2^{y_{i,2}} \left(\frac{g_1^{S_{i,1}} g_2^{S_{i,2}} X^{\alpha_i}}{X^{\alpha_i}} \right)^c \\ &= (g_1^{y_{i,1}} g_2^{y_{i,2}}) (g_1^{S_{i,1}} g_2^{S_{i,2}})^c \\ &= (g_1^{y_{i,1}} g_2^{y_{i,2}}) (g_1^{cS_{i,1}} g_2^{cS_{i,2}}) \\ &= g_1^{y_{i,1} + cS_{i,1}} g_2^{y_{i,2} + cS_{i,2}} \\ &= g_1^{z_{i,1}} g_2^{z_{i,2}} \end{aligned}$$

4 SECURITY ANALYSIS

We describe the security of the Hierarchical Twin-Schnorr IBI scheme with the following game between an impersonator \mathcal{I} and a challenger \mathcal{C} .

Theorem 4.1. *Hierarchical Twin-Schnorr IBI scheme is secure against impersonation under active and concurrent attack if the discrete logarithm problem is hard in group \mathbb{G} , where*

$$\epsilon_{Twin-SchnorrHIBI}^{imppa} = \sqrt[l]{\epsilon_{G,C}^{DLOG}(k) + \frac{1}{2^k} + \frac{1}{2^k}}$$

Proof. Let \mathcal{I} be an impersonator who (t, q_i, ϵ) breaks the security of Hierarchical Twin-Schnorr IBI scheme. \mathcal{C} is a simulator that find out the value of a according to discrete logarithm assumption. \mathcal{C} will be given a group G , generators $(g_1 = g, g_2 = g^a) \in G$, \mathcal{C} will simulate for \mathcal{I} as follows.

1. **Setup**(S). \mathcal{C} takes 1^k and returns $mpk = \langle \mathbb{G}, q, g_1, g_2, X \rangle$ to \mathcal{I} .
2. **Phase 1**. \mathcal{I} can issue queries $(q_0, \dots, q_i, \dots, q_m)$ where q_i is for ID_i . There are q_m queries in total as there is m number of queries. In training phase, \mathcal{I} tries to learn from the \mathcal{C} . It will forge the user secret key and runs transcript. It is considered as a hierarchical version of the Twin-Schnorr IBI scheme for (ID_1, \dots, ID_m) , where (ID_1, \dots, ID_i) for $1 \leq i \leq m$ and $(level_1, level_2, \dots, level_l)$ where $(level_1, \dots, level_j)$ for $1 \leq j \leq l$ to define hierarchy.

(a) Case 1.

i. **Extract Query** (\mathcal{E}). For $ID_i \neq ID^*$, \mathcal{C} takes master public key and identity string as the input. Upon being queried with the public key of ID_i and returns $usk_{ID_i} = (S_{i,1}, S_{i,2})$ to \mathcal{I} . To calculate usk_{ID_i} with the help of ancestor $usk_{ID_{i-1}}$ can be done.

ii. **Identification query** (\mathcal{P} and \mathcal{V}). For \mathcal{I} , \mathcal{C} responds with a transcript for the interaction between the prover and a verifier. In the simulation, Prover takes input (mpk, ID_i, usk_{ID_i}) where the verifier takes input (mpk, ID_i) . Prover generates (Y, V_i) . \mathcal{C} generates random challenge $c \in Z_q$. On the basis of challenge prover calculates $z_{i,1}, z_{i,2}$ to \mathcal{V} as its response. Lastly \mathcal{V} verifies

$$g_1^{z_{i,1}} g_2^{z_{i,2}} = Y \left(\frac{V_i}{X^{\alpha_i}} \right)^c$$

(b) Case 2.

i. **Extract Query** (\mathcal{E}). For $ID^* = ID_i$, the ancestor of usk_{ID^*} is unknown. But, the root secret key is known. Therefore, the algorithm aborts. There is ID string where all ID are defined as parent and child node according to hierarchy. Parent helps to generate usk of child node. Child node's usk is generated only in case it has parent usk defined. \mathcal{C} takes master public key and identity string as the input. Upon being queried with the public key of ID^* and returns $usk_{ID^*} = (S_{*,1}, S_{*,2})$ to \mathcal{I} .

ii. **Identification query** (\mathcal{P} and \mathcal{V}). When transcript will create even if not yet queried before as an extract query. Prover participate in transcript and add in the set. We will not able to issue transcript for the already corrupted user. Prover and verifier communicates in this phase. ID^* is targeted identity and verifier needs to verify it. $ID_i = ID_*$, \mathcal{I} act as the cheater \mathcal{V} and \mathcal{C} does not have user secret key of ID_* , however it needs to create it again to run an identification protocol. When \mathcal{I} tries to forge ID^* then he should know the previous (ID_{*-1}) . We can perform transcript as many times as number of queries does not exceed. Prover takes input (mpk, ID^*, usk_{ID^*}) where the verifier takes input (mpk, ID^*) . Prover generates (Y, V_*) . \mathcal{C} generates random challenge $c \in Z_q$ where c corresponds to ID^* . On the basis of challenge prover calculates $z_{*,1}, z_{*,2}$ to \mathcal{V} as its response. Lastly \mathcal{V} verifies $g_1^{z_{*,1}} g_2^{z_{*,2}} = Y \left(\frac{V_*}{X^{\alpha_*}} \right)^c$.

(c) **Challenge** (\mathcal{C}). \mathcal{I} outputs an $ID_i \neq ID^*$ that it wishes to impersonate.

3. **Phase 2.** Breaking phase calculates as follows:

$[y_{i,1}, c_1, V_i, z_{i,1}]$ and $[y_{i,2}, c_2, V_i, z_{i,2}]$ from \mathcal{I} where $c_1 \neq c_2$. From here, \mathcal{C} extracts $\widetilde{S}_{i,1} = (z_{i,1} - z_{i,2})/(c_2 - c_1)$ and $\widetilde{S}_{i,2} = (z_{i,1} - z_{i,2})/(c_2 - c_1)$.

If $S_{i,1} = \widetilde{S}_{i,1}$ and $S_{i,2} = \widetilde{S}_{i,2}$ then \mathcal{C} aborts.

$$\begin{aligned} g^{S_{i,1}} g^{S_{i,2}} &= g^{\widetilde{S}_{i,1}} g^{\widetilde{S}_{i,2}} \\ g^{S_{i,1} + a S_{i,2}} &= g^{\widetilde{S}_{i,1} + a \widetilde{S}_{i,2}} \\ g^{a S_{i,2}} - g^{a \widetilde{S}_{i,2}} &= g^{\widetilde{S}_{i,1}} - g^{S_{i,1}} \\ g^a &= g^{(\widetilde{S}_{i,1} - S_{i,1})(S_{i,2} - \widetilde{S}_{i,2})} \\ a &= -\frac{\widetilde{S}_{i,1} - S_{i,1}}{S_{i,2} - \widetilde{S}_{i,2}} \end{aligned}$$

□

To calculate the probability of \mathcal{C} winning the game to solve the discrete logarithm problem. By the Reset Lemma, will successfully extract 2 valid conversations to derive $(S_{i,1}, S_{i,2})$ and calculating a with the probability $\varepsilon_{\text{TwIn-SchnorrHIBI}}^{\text{imp}_{aa/ca}} - (-\frac{1}{2^k} - \frac{1}{2^k})^l$. Assume \mathcal{C} solves the discrete logarithm assumption. \mathcal{C} which computes correct value of a then event is A and not aborting event is B. Winning probability can be given as following.

$$\begin{aligned} \mathcal{C} &= Pr[A \wedge B] \\ \mathcal{C} &= Pr[A|B]Pr[B] \\ \varepsilon_{G,C}^{\text{DLOG}}(k) &\geq (\varepsilon - \frac{1}{2^k})^l Pr[B] \end{aligned}$$

The probability of \mathcal{C} aborting when event B is $S_{i,1} = \widetilde{S}_{i,1}$ and $S_{i,2} = \widetilde{S}_{i,2}$. Therefore probability of winning \mathcal{C} is,

$$\begin{aligned} \varepsilon_{G,C}^{\text{DLOG}}(k) &\geq (\varepsilon_{\text{TwIn-SchnorrHIBI}}^{\text{imp}_{ca/aa}} - \frac{1}{2^k})^l - \frac{1}{2^k} \\ \varepsilon_{G,C}^{\text{DLOG}}(k) + \frac{1}{2^k} &\geq (\varepsilon_{\text{TwIn-SchnorrHIBI}}^{\text{imp}_{ca/aa}} - \frac{1}{2^k})^l \\ \varepsilon_{\text{TwIn-SchnorrHIBI}}^{\text{imp}_{pa}} &\leq \sqrt[l]{\varepsilon_{G,C}^{\text{DLOG}}(k) + (\frac{1}{2^k} + \frac{1}{2^k})} \end{aligned}$$

5 EFFICIENCY ANALYSIS

In this section, we provide the efficiency cost of the Hierarchical Twin-Schnorr IBI scheme in Table 1. We consider exponentiations (E), multiplications in group \mathbb{G} (MG), multiplications in \mathbb{Z}_q (MZ) and additions in \mathbb{Z}_q (A) in terms to define the efficiency in order.

Algorithm	E	MG	MZ	A
SETUP	2	1	0	0
EXTRACT	4	2	4	4
PROVE	5	3	2	2
VERIFY	4	3	0	0

Table 1: Efficiency analysis for the Hierarchical Twin-Schnorr IBI scheme

We consider other schemes in order to calculate the identification cost in Table 2. The Twin-Schnorr IBI is slightly superior in terms of efficiency and security compared to the HIBI scheme proposed in Fujioka et al. (2014). We are considering the Hierarchical Twin-Schnorr IBI scheme which is efficient scheme in case of targeted identity where E: exponentiation, MG: multiplication in G, MZ: multiplication in Z, G: element in G, and Z: element in Z.

Scheme	E	MG	MZ	A	Assumption
HIBI by Chin et al. (2009)	6	3	2	1	CDH,OMCDH
HIBI by Fujioka et al. (2014)	2	1	1	0	Prime order bilinear group
HIBI by Fujioka et al. (2014)	4	2	2	1	Composite order bilinear group
HIBI Fujioka et al. (2014)	2	2	1	1	RSA
Twin-Schnorr HIBI	9	6	2	2	DLP

Table 2: Comparison of the identification protocol with other HIBI schemes

According to communication cost calculation in Table 3, the Hierarchical Twin-Schnorr-IBI is more efficient and secure compared to the HIBI scheme. In the Hierarchical Twin-Schnorr-IBI, we are using msk in order to generate the user secret key for root level and then parent usk helps to generate the child usk. All child nodes are depended on the parent nodes to calculate usk. Therefore, Hierarchical Twin-Schnorr-IBI is more efficient in comparison with HIBI.

Scheme	E	MG	MZ	Communication costs
HIBI by Chin et al. (2009)	5	4	4	$(2G+2Z)$
Twin-Schnorr HIBI	9	6	2	$(6G+2Z)$

Table 3: Comparison of the identification protocol of Hierarchical Twin-Schnorr IBI and HIBI after precomputation

6 CONCLUSION

In this paper, we upgraded the Twin-Schnorr IBI scheme into the Hierarchical Twin-Schnorr IBI scheme. Our proposed Hierarchical Twin-Schnorr IBI scheme is designed to prove many identification and verification at a time. The proposed scheme is efficient as it is pairing-free and secure based the discrete logarithmic assumption.

REFERENCES

- Barapatre, P. and Rangan, C. P. (2013). Identity-based identification schemes from id-kems. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*, pages 111–129. Springer.
- Bellare, M., Namprempre, C., and Neven, G. (2009). Security proofs for identity-based identification and signature schemes. *Journal of Cryptology*, 22(1):1–61.
- Bellare, M. and Palacio, A. (2002). Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In *Annual International Cryptology Conference*, pages 162–177. Springer.
- Boneh, D. and Franklin, M. (2003). Identity-based encryption from the weil pairing. *SIAM journal on computing*, 32(3):586–615.
- Chin, J.-J., Heng, S.-H., and Goi, B.-M. (2009). Hierarchical identity-based identification schemes. In *International Conference on Security Technology*, pages 93–99. Springer.
- Chin, J.-J., Tan, S.-Y., Heng, S.-H., and Phan, R. C.-W. (2015). Twin-schnorr: a security upgrade for the schnorr identity-based identification scheme. *The Scientific World Journal*, 2015.
- Fiat, A. and Shamir, A. (1986). How to prove yourself: Practical solutions to identification and signature problems. In *Conference on the Theory and Application of Cryptographic Techniques*, pages 186–194. Springer.
- Fujioka, A., Saito, T., and Xagawa, K. (2012). Security enhancements by or-proof in identity-based identification. In *International Conference on Applied Cryptography and Network Security*, pages 135–152. Springer.
- Fujioka, A., Saito, T., and Xagawa, K. (2014). Secure hierarchical identity-based identification without random oracles. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 97(6):1307–1317.
- Gentry, C. and Silverberg, A. (2002). Hierarchical id-based cryptography. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 548–566. Springer.
- Horwitz, J. and Lynn, B. (2002). Toward hierarchical identity-based encryption. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 466–481. Springer.
- Ioannidis, J., Keromytis, A., and Yung, M. (2005). *Applied cryptography and network security*. Springer Berlin/Heidelberg.
- Katz, J. and Lindell, Y. (2014). *Introduction to modern cryptography*. CRC press.
- Kurosawa, K. and Heng, S.-H. (2004). From digital signature to id-based identification/signature. In *International Workshop on Public Key Cryptography*, pages 248–261. Springer.

- Shamir, A. (1984). Identity-based cryptosystems and signature schemes. In *Workshop on the theory and application of cryptographic techniques*, pages 47–53. Springer.
- Tan, S.-Y., Heng, S.-H., Phan, R. C.-W., and Goi, B.-M. (2011). A variant of schnorr identity-based identification scheme with tight reduction. In *International Conference on Future Generation Information Technology*, pages 361–370. Springer.

Revisiting the Invisibility of Yuen et al.'s Undeniable Signature Scheme

Jia-Ch'ng Loh¹, Swee-Huay Heng², and Syh-Yuan Tan³

*Faculty of Information Science and Technology,
Multimedia University, Melaka, Malaysia*

*E-mail: ¹jasonlohjc@gmail.com, ²shheng@mmu.edu.my,
³sytan@mmu.edu.my*

ABSTRACT

Undeniable signature is a special featured signature which should not be verifiable without the help of the original signer. Invisibility is a security property possessed by undeniable signature which implies the inability of the adversary to guess the validity of an undeniable signature. Convertibility was later introduced to the undeniable signature where it allows the signer to transform an undeniable signature into a publicly verifiable digital signature. Yuen et al. proposed the first pairing based provably secure convertible undeniable signature scheme in the standard model. However, it was later revisited by Phong et al. and Zhao. Phong et al. pointed out that the scheme is not strongly unforgeable and its invisibility proof is imperfect and incorrect, and Zhao showed that its invisibility is broken. In this paper, we show that Zhao's attack is infeasible. We then show that Yuen et al.'s scheme is invisible under Galbraith and Mao's model and propose a fix to the invisibility proof.

Keywords: Attack, undeniable signature, convertible, invisibility

1 INTRODUCTION

The notion of undeniable signature was introduced by Chaum and van Antwerpen (1990). Unlike ordinary digital signature, undeniable signature has a distinctive feature, i.e. without the help of the signer, the verifiers will not be able to verify the validity of the undeniable signature. There are various applications of undeniable signature such as licensing software, electronic cash, electronic voting and auctions. There are also some variants of undeniable signature proposed such as convertible undeniable signature, designated verifier signature, and designated confirmer signature.

The notion of convertible undeniable signature as proposed by Boyar et al. (1991) is an extension of undeniable signature that allows the signer to transform an undeniable signature into

a universally verifiable ordinary digital signature. There are two types of convertible undeniable signature, namely, selectively convertible and universally convertible. The selectively convertible undeniable signature allows the signer to convert only a specific undeniable signature into a universally verifiable one by releasing a token. In universally convertible undeniable signature, the signer releases part of his secret to make the undeniable signatures universally verifiable.

The ultimate goal of undeniable signature is to protect the privacy of the signer. Traditionally, the notion of invisibility was introduced by Galbraith and Mao (2003) as the main security property for undeniable signature. Undeniable signature should satisfy invisibility which implies the inability of a user to distinguish a valid undeniable signature from a random element as shown by Galbraith and Mao (2003). The first pairing based provably secure convertible undeniable signature without random oracles was proposed by Yuen et al. (2007a). However, it was later revisited by Phong et al. (2009) in the full version of Phong et al. (2010) and Zhao (2010). Phong et al. (2009) pointed out that the scheme does not satisfy strong unforgeability and its invisibility proof is imperfect and incorrect, and Zhao (2010) showed that the invisibility is actually broken.

Our Contributions. We revisit Yuen et al.'s convertible undeniable signature scheme especially their invisibility proof. We also review the past attacks on the invisibility by Phong et al. (2009) and Zhao (2010) respectively. We point out that Zhao's attack is infeasible. We also show that Yuen et al.'s convertible undeniable signature scheme is invisible under Galbraith and Mao's model upon proposing a fix to the invisibility proof so that that the simulation is perfect and correct.

2 PRELIMINARIES

2.1 Bilinear Pairings (Boneh and Franklin, 2001)

A brief review on the properties of bilinear pairings is presented here. Let \mathbb{G} and \mathbb{G}_T be cyclic groups of prime order p and a generator $g \in \mathbb{G}$. The map $\hat{e} : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ is a bilinear map which satisfies the following three properties: **Bilinearity**: for all $(x, y, z) \in \mathbb{G}$ and $(a, b) \in \mathbb{Z}_p$, we have $\hat{e}(x^a, y^b) = \hat{e}(x, y)^{ab}$. **Non-degeneracy**: if g is a generator of \mathbb{G} , then $\hat{e}(g, g)$ is a generator of \mathbb{G}_T , which also implies $\hat{e}(g, g) \neq 1$. **Computability**: there exists an efficient algorithm to compute $\hat{e}(x, y)$ for all $x, y \in \mathbb{G}$.

Decision Linear Problem. (Boneh et al., 2004) Given $u, u^a, v, v^b, h, h^c \in \mathbb{G}$ for unknown $a, b, c \in \mathbb{Z}_p^*$, output "1" if $c = a + b$ and "0" otherwise.

Definition 2.1. *The (ε, t) -Decision Linear assumption holds in \mathbb{G} if there is no t time algorithm that can have advantage at least ε in solving the Decision Linear problem in \mathbb{G} .*

2.2 Convertible Undeniable Signature

A convertible undeniable signature scheme consists of the following algorithms and protocols (Yuen et al., 2007a):

- *Setup*: On input a security parameter 1^k , it outputs a public parameter $param$.
- *KeyGen*: On input $param$, it outputs a signer public and private key pair (pk, sk) .
- *Sign*: On input $param$, a message m , and a signer private key (m, sk) , it outputs an undeniable signature σ .
- *Confirmation/Disavowal Protocol*: An interactive protocol that runs between the signer and the verifier on common input $(param, pk, m, \sigma)$. The signer uses sk to check the validity of σ , then the signer proves to the verifier that σ is valid on m under pk , and the verifier outputs “1” for acceptance and “0” for rejection.
- *SConvert*: On input $(param, sk, m, \sigma)$, it computes a selective token π_s which can be used to publicly verify (m, σ) on pk .
- *SVerify*: On input $(param, pk, m, \sigma, \pi^s)$, it outputs \perp if π^s is an invalid token on pk . Else, it outputs “1” if (m, σ) is valid on pk and outputs “0” otherwise.
- *UConvert*: On input $(param, sk)$, it computes a universal token π^u which can be used to publicly verify every σ generated by sk .
- *UVerify*: On input $(param, pk, m, \sigma, \pi^u)$, it outputs \perp if π^u is an invalid token on pk . Else, it outputs “1” if (m, σ) is valid on pk and outputs “0” otherwise.

2.2.1 Invisibility

Invisibility requires that given (m, σ) and a possible signer’s public key pk , there is no computational way to decide whether σ is valid on m or not without the help of the signer. Its security model is defined as the following game between an adversary \mathcal{A} and a simulator \mathcal{S} . We review the invisibility models proposed by Phong et al. (2009), Yuen et al. (2007a), and Galbraith and Mao (2003) respectively. We also highlight the differences among them. The models are arranged in the order of the strongest model to the weakest model.

Adversary Game. Let \mathcal{A} be the adversary who is allowed to query sign oracle \mathcal{O}_S , sconvert oracle \mathcal{O}_{SC} , and confirmation/disavowal oracle \mathcal{O}_{CD} . At some point, \mathcal{A} outputs a challenge message \hat{m} which has never been queried to \mathcal{O}_S before. \mathcal{S} then randomly selects $b \in \{0, 1\}$ and generates a challenge signature $\hat{\sigma}$ where $\hat{\sigma} = \text{Sign}(sk, \hat{m})$ if $b = 0$, and $\hat{\sigma}$ is randomly selected from the signature space if $b = 1$. \mathcal{A} can still make queries to \mathcal{O}_S , \mathcal{O}_{SC} , and \mathcal{O}_{CD} with the restrictions based on the following models:

- Phong et al.’s model: \mathcal{A} cannot query $(\hat{m}, \hat{\sigma})$ to \mathcal{O}_{SC} and \mathcal{O}_{CD} .

- Yuen et al.'s model: \mathcal{A} cannot query \hat{m} to \mathcal{O}_S , any (\hat{m}, \cdot) to \mathcal{O}_{SC} , and $(\hat{m}, \hat{\sigma})$ to \mathcal{O}_{CD} .
- Galbraith and Mao's model: \mathcal{A} cannot query any (\hat{m}, \cdot) in the equivalence class of $(\hat{m}, \hat{\sigma})$ to \mathcal{O}_{SC} and \mathcal{O}_{CD} .

At the end of the game, \mathcal{A} outputs its guess b' and wins the game if $b' = b$. The advantage of \mathcal{A} has in the above game is defined as $\text{Adv}(\mathcal{A}) = |\Pr[b = b'] - \frac{1}{2}|$.

Definition 2.2. A convertible undeniable signature is $(\varepsilon, t, q_s, q_{sc}, q_{cd})$ -invisible if there is no (t, q_s, q_{sc}, q_{cd}) - \mathcal{A} who can have success probability more than ε in the game above with at most q_s queries to \mathcal{O}_S , q_{sc} queries to \mathcal{O}_{SC} , and q_{cd} queries to \mathcal{O}_{CD} in time t .

3 REVIEW OF YUEN ET AL.'S CONVERTIBLE UNDENIABLE SIGNATURE

3.1 Yuen et al.'s Convertible Undeniable Signature Scheme

Yuen et al.'s convertible undeniable signature scheme consists of the following algorithms and protocols. Readers may refer to Yuen et al. (2007a) for *UConvert* and *UVerify* algorithms as it will not affect the discussion in this paper.

- *Setup*: Let \mathbb{G} be groups of prime order p . Given a pairing: $\hat{e} : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$, select generators $g, g_2 \in \mathbb{G}$. Generator $u' \in \mathbb{G}$ is selected at random, and a n -length vector $U = (u_i)$ whose elements are chosen at random from \mathbb{G} . Select an integer d as a system parameter. Denote $\ell = 2^d$ and $k = \frac{n}{d}$. Let $H_j : \{0, 1\}^n \rightarrow \mathbb{Z}_\ell^*$ be collision resistant hash functions, where $1 \leq j \leq k$.
- *KeyGen*: Select $\alpha, \beta', \beta_i \in_R \mathbb{Z}_p^*$ for $1 \leq i \leq \ell$. Set $g_1 = g^\alpha, v' = g^{\beta'}$ and $v_i = g^{\beta_i}$. The public keys are $(g_1, v', v_1, \dots, v_\ell)$. The private keys are $(\alpha, \beta', \beta_1, \dots, \beta_\ell)$.
- *Sign*: To sign a message $m = (m_1, \dots, m_n) \in \{0, 1\}^n$, denote $m_j^* = H_j(m)$ for $1 \leq j \leq k$. The signer picks $r \in_R \mathbb{Z}_p^*$ and computes the signature $\sigma = (S_1, S_{2,1}, \dots, S_{2,k})$:

$$S_1 = g_2^\alpha (u' \prod_{i=1}^n u_i^{m_i})^r \quad S_{2,j} = (v' \prod_{i=1}^{\ell} v_i^{m_j^{*i}})^r$$

- *Confirmation/Disavowal Protocol*: On input $(S_1, S_{2,1}, \dots, S_{2,k})$, the signer computes:

$$\begin{aligned} L &= \hat{e}(g, g_2) & M &= \hat{e}(g_1, g_2) \\ N_j &= \hat{e}(v' \prod_{i=1}^{\ell} v_i^{m_j^{*i}}, g_2) & O_j &= \hat{e}(v' \prod_{i=1}^{\ell} v_i^{m_j^{*i}}, S_1) / \hat{e}(S_{2,j}, u' \prod_{i=1}^n u_i^{m_i}) \end{aligned}$$

for $1 \leq j \leq k$ to show that whether (L, M, N_j, O_j) are Diffie-Hellman (DH) tuples by executing the 3-move witness indistinguishable (WI) protocols¹ as in Kurosawa and Heng

¹We refer readers to Kurosawa and Heng (2005) for the detailed description of WI protocols.

(2005) of the equality or the inequality of discrete logarithm $\alpha = \log_L M$ and $\log_{N_j} O_j$ in \mathbb{G}_T .

- *SConvert*: On input the signature $(S_1, S_{2,1}, \dots, S_{2,k})$ on the message m , the signer computes $m_1^* = H_1(m)$ and $S_2' = S_{2,1}^{1/(\beta' + \sum_{i=1}^{\ell} \beta_i m_1^{*i})}$. It outputs a selective token S_2' for message m .
- *SVerify*: On input the signature $(S_1, S_{2,1}, \dots, S_{2,k})$ for the message m and the selective token S_2' , it computes $m_j^* = H_j(m)$ for $1 \leq j \leq k$ and checks $\hat{e}(g, S_{2,j}) = \hat{e}(S_{2,j}', v' \prod_{i=1}^{\ell} v_i^{m_j^{*i}})$. If they are not equal, it outputs \perp . Otherwise, it compares $\hat{e}(g, S_1) = \hat{e}(g_1, g_2) \cdot \hat{e}(S_2', u' \prod_{i=1}^n u_i^{m_i})$. It outputs “1” if the equation holds and outputs “0” otherwise.

3.2 The Invisibility Proof of Yuen et al.’s Convertible Undeniable Signature

Theorem 3.1. (Yuen et al., 2007a) *The scheme is $(\varepsilon, t, q_s, q_{sc}, q_{cd})$ -invisible if the (ε', t') -decision linear assumption holds in \mathbb{G} .*

Proof. Suppose there is a $(\varepsilon, t, q_s, q_{sc}, q_{cd})$ -adversary \mathcal{A} who can break the invisibility, we build a probabilistic polynomial time (PPT) algorithm \mathcal{B} that makes use of \mathcal{A} to solve the decision linear problem with probability at least ε' in time t' . \mathcal{B} is given a decision linear problem instance (u, v, h, u^a, v^b, h^c) as in Definition 2.1.

Setup. An interest reader may refer to Yuen et al. (2007a) for the full setup. At the end, \mathcal{B} constructs such equations:

$$u' \prod_{i=1}^n u_i^{m_i} = g_2^{F(m)} g^{J(m)}, \quad v' \prod_{i=1}^{\ell-1} v_i^{m_j^{*i}} = g^{G(m_j^*)} v^{K(m_j^*)} \quad \text{for } 1 \leq j \leq k$$

where $m_j^* = H_j(m)$ for $1 \leq j \leq k$. \mathcal{B} maintains an empty list \mathcal{L} and passes all public parameters to \mathcal{A} .

Queries. \mathcal{B} simulates the oracles as follows:

- **Signing Oracle:** On input a message query $m_i = \{m_1, \dots, m_n\}$, \mathcal{B} constructs a signature by assuming $F(m_i) \neq 0 \pmod p$ and $K(m_{i,j}^*) = 0 \pmod p$, where $m_{i,j}^* = H_j(m_i)$ for all $1 \leq j \leq k$. It randomly chooses $r_i \in_R \mathbb{Z}_p$ and computes the signature as $\sigma_i = (S_1, S_{2,1}, \dots, S_{2,k})$, where

$$S_1 = g_1^{-\frac{J(m_i)}{F(m_i)}} (g_2^{F(m_i)} g^{J(m_i)})^{r_i}, S_{2,j} = (g_1^{-\frac{1}{F(m_i)}} g^{r_i})^{G(m_{i,j}^*)} \text{ for } 1 \leq j \leq k$$

\mathcal{B} stores (m_i, σ_i) into \mathcal{L} and returns σ_i to \mathcal{A} .

- **Confirmation/Disavowal Oracle:** On input a message and signature pair (m_i, σ_i) , \mathcal{B} will first check whether (m_i, σ_i) is in \mathcal{L} . If so, \mathcal{B} outputs “1” and runs *Confirmation Protocol* with \mathcal{A} to show that (L, M, N_j, O_j) are DH tuples. Since \mathcal{B} knows the discrete logarithm of N_j with base $L = (1/G(m_{i,j}^*))$, it can simulate the interactive proof perfectly. Otherwise, if (m_i, σ_i) is not in \mathcal{L} , it outputs “0” and runs *Disavowal Protocol* with \mathcal{A} .
- **SConvert Oracle:** On input a message and signature pair (m, σ) , \mathcal{B} computes $m_j^* = H_j(m)$ for $1 \leq j \leq k$. If $K(m_j^*) \not\equiv 0 \pmod{p}$ for any j , \mathcal{B} aborts. Otherwise, \mathcal{B} outputs a valid selective token $S_2' = S_{2,1}^{1/G(m_1^*)}$, which can be used to verify (m, σ) .

Challenge. At some point, \mathcal{A} submits a challenge message $\hat{m} = (\hat{m}_1, \dots, \hat{m}_n)$ to \mathcal{B} . Denote $\hat{m}_j^* = H_j(\hat{m})$ for $1 \leq j \leq k$. If $F(\hat{m}_i) = 0 \pmod{p}$, $J(\hat{m}_i) \not\equiv 0 \pmod{p}$ or $G(\hat{m}_j^*) \not\equiv 0 \pmod{p}$ for any j , \mathcal{B} aborts. Otherwise, \mathcal{B} computes the challenge signature as $\hat{\sigma} = (\hat{S}_1, \hat{S}_{2,1}, \dots, \hat{S}_{2,k})$ to \mathcal{A} , where

$$\hat{S}_1 = h^c, \quad \hat{S}_{2,j} = v^{bK(\hat{m}_j^*)/F(\hat{m}_i)} \quad \text{for } 1 \leq j \leq k$$

Output. \mathcal{A} outputs its guess b' . \mathcal{B} returns b' as the solution to the decision linear problem. Note that if $c = a + b$, then $\hat{S}_1 = g_2^{a+b} = g_2^a (g_2^{F(\hat{m}_i)})^{b/F(\hat{m}_i)} = g_2^a (u' \prod_{i=1}^n u_i^{m_i})^{b/F(\hat{m}_i)}$, and $\hat{S}_{2,j} = v^{bK(\hat{m}_j^*)/F(\hat{m}_i)} = (v' \prod_{i=1}^{\ell} v_i^{\hat{m}_j^{*i}})^{b/F(\hat{m}_i)}$ for $1 \leq j \leq k$. \square

4 PAST ATTACKS

In this section, we briefly recall the past attacks on Yuen et al.'s convertible undeniable signature scheme from Phong et al. (2009) and Zhao (2010).

We first recall again a signature on a message m is in the form of $\sigma = (S_1, S_{2,1}, \dots, S_{2,k})$, and $S_1 = g_2^\alpha U^r$ and $S_{2,j} = V_j^r$ for $1 \leq j \leq k$, where α is the private key, r is random salt, and (g_2, U, V_j) are publicly computable values.

4.1 The Attack by Phong et al. (2009)

Phong et al. (2009) showed that the undeniable signature σ is not strongly unforgeable as a new signature in the form $\sigma' = (S_1 U^{r'}, S_{2,1} V_1^{r'}, \dots, S_{2,k} V_k^{r'})$ is also valid on the same message m where r' is chosen, so the random salt is now $r + r'$. They then assumed that there is an adversary \mathcal{A} who requests a challenge signature $\hat{\sigma}$ on \hat{m} which is either a valid signature or a random element. \mathcal{A} constructs a new $\hat{\sigma}'$ again where the validity of both $(\hat{m}, \hat{\sigma})$ and $(\hat{m}, \hat{\sigma}')$ is the same. \mathcal{A} then submits a query to sconvert oracle \mathcal{O}_{SC} with the input of $(\hat{m}, \hat{\sigma}')$. The output of the selective token can then help to verify $(\hat{m}, \hat{\sigma}')$ using the algorithm *SVerify*. If the answer is “0”, it indicates $(\hat{m}, \hat{\sigma}')$ is invalid, hence $(\hat{m}, \hat{\sigma})$ is also invalid. Finally, \mathcal{A} can decide its guess b' and win the game of invisibility under Phong et al.'s model. However, they also mentioned that this attack is infeasible in Yuen et al.'s model as such attack is restricted where \mathcal{A} cannot

query any (\hat{m}, \cdot) to \mathcal{O}_{SC} . This shows that Yuen et al.'s scheme is not invisible in Phong et al.'s stronger model.

Phong et al. (2009) then pointed out that the invisibility proof of Theorem 3.1 is actually imperfect and incorrect, which happens in the confirmation/disavowal oracle \mathcal{O}_{CD} . The algorithm \mathcal{B} as defined in the proof contains a list \mathcal{L} which stores every message and signature pair (m, σ) generated by its sign oracle \mathcal{O}_S . Due to the existence of list \mathcal{L} , \mathcal{O}_{CD} can check the validity of (m, σ) by comparing against the list \mathcal{L} . Recall that σ is not strongly unforgeable, \mathcal{A} can construct a new signature σ' where the validity of σ and σ' is equivalent. Subsequently, if \mathcal{A} submits a valid (m, σ') to \mathcal{O}_{CD} , \mathcal{B} will always return "0" and execute *Disavowal Protocol* with \mathcal{A} because (m, σ') is not in \mathcal{L} which is an incorrect simulation.

4.2 The Attack by Zhao (2010)

Assume there is an adversary \mathcal{A} in Yuen et al.'s invisibility model as in Definition 2.2. \mathcal{A} submits a challenge message \hat{m} to request a challenge signature $\hat{\sigma}$ where $\hat{\sigma}$ is a valid signature on \hat{m} if the challenge bit $b = 0$, otherwise $\hat{\sigma}$ is a random element if $b = 1$. \mathcal{A} chooses a random value r' and constructs a new signature $\hat{\sigma}' = (S_1 U^{r'}, S_{2,1} V_1^{r'}, \dots, S_{2,k} V_k^{r'})$ where the validity of both $(\hat{m}, \hat{\sigma})$ and $(\hat{m}, \hat{\sigma}')$ is the same. \mathcal{A} then queries $(\hat{m}, \hat{\sigma}')$ to the confirmation/disavowal oracle \mathcal{O}_{CD} . If the oracle returns "1" and runs the *Confirmation Protocol* with \mathcal{A} , it means $(\hat{m}, \hat{\sigma}')$ is a valid signature. Otherwise it is a random element if the oracle returns "0" and runs the *Disavowal Protocol* with \mathcal{A} . In either case, \mathcal{A} can decide its guess b' and win the game of invisibility.

5 DISCUSSION

5.1 The Attack by Zhao (2010) is Infeasible

We notice that the attack by Zhao (2010) is infeasible in the invisibility proof as in Theorem 3.1. The algorithm \mathcal{B} in the proof is keeping a list \mathcal{L} which stores every message and signature pair (m, σ) that has been queried to sign oracle \mathcal{O}_S . Hence, if \mathcal{A} constructs a new signature $\hat{\sigma}'$, where the validity of a challenge message and signature pair $(\hat{m}, \hat{\sigma})$ is equivalent to the validity of $(\hat{m}, \hat{\sigma}')$, and queries to confirmation/disavowal oracle \mathcal{O}_{CD} , \mathcal{B} will always output "0" and run *Disavowal Protocol* with \mathcal{A} because $(\hat{m}, \hat{\sigma}')$ is not in the list \mathcal{L} . Thus, this falsifies the attack put forth by Zhao.

5.2 Yuen et al.'s Convertible Undeniable Signature Scheme is Invisible under Galbraith and Mao's Model

Recall again in Section 2.2.1, we reviewed three types of invisibility models which were proposed by Phong et al. (2009), Yuen et al. (2007a), and Galbraith and Mao (2003) respectively. We observe that Yuen et al.'s scheme is invisible under Galbraith and Mao's model which re-

stricted the query of any challenge message pair (\hat{m}, \cdot) in the equivalence class of challenge message and signature pair $(\hat{m}, \hat{\sigma})$ to sconvert oracle \mathcal{O}_{SC} and confirmation/disavowal oracle \mathcal{O}_{CD} . However, the simulation in the proof of Theorem 3.1 is incorrect, specifically in \mathcal{O}_{CD} , where \mathcal{B} will disavow every (m, σ) that is not in the list \mathcal{L} even if it is a valid pair signed by its sign oracle \mathcal{O}_S . In order to ensure that algorithm \mathcal{B} can simulate \mathcal{O}_{CD} correctly, we propose a fix to the simulation. We suggest to remove the list \mathcal{L} and make a verification to the message with the function $K(\cdot) \neq 0$. We then redefine \mathcal{O}_{CD} as follows:

- **Confirmation/Disavowal Oracle:** On input (m, σ) where $\sigma = (S_1, S_{2,1}, \dots, S_{2,k})$, \mathcal{B} first computes $m_j^* = H_j(m)$ for $1 \leq j \leq k$. If $K(m_j^*) \neq 0$, \mathcal{B} aborts. Otherwise, \mathcal{B} generates a selective token, where $S_2' = S_{2,1}^{1/G(m_1^*)}$, and then runs the algorithm *SVerify* to check the validity of σ . If the output is “1”, \mathcal{B} returns “1” and executes *Confirmation Protocol* with \mathcal{A} . Otherwise, \mathcal{B} returns “0” and executes *Disavowal Protocol* with \mathcal{A} .

Since the scheme is only existential unforgeable against chosen message attack, as long as \mathcal{B} receives a valid (m, σ) generated by \mathcal{O}_S or a new (m, σ') which has the same validity with (m, σ) , \mathcal{B} can always claim it as a valid pair from \mathcal{O}_S .

An improved version of Yuen et al. (2007a) was presented by Yuen et al. (2007b) where they improved the scheme to be strongly unforgeable and fixed Phong et al.'s attack as mentioned in Section 4. In Yuen et al.'s improved scheme, their signature consists of an extra element and its invisibility proof still maintains a list \mathcal{L} to keep every message and signature pair (m, σ) signed by \mathcal{O}_S . This indicates that in real life application, the signer needs to always store a list \mathcal{L} of (m, σ) which is not cost effective as the signer is required to maintain a huge storage of (m, σ) . In comparison, the original Yuen et al.'s scheme after our proposed fix in the invisibility proof still assures invisibility under Galbraith and Mao's model without strong unforgeability even though it is not invisible under the stronger models of Yuen et al. and Phong et al. respectively.

6 CONCLUSION

We pointed out that Zhao's attack is infeasible. We showed that Yuen et al.'s convertible undeniable signature scheme is invisible under Galbraith and Mao's model after our proposed fix to its invisibility proof so that the simulation can be perfect and correct.

Acknowledgement. The authors would like to acknowledge the Malaysia government's Fundamental Research Grant Scheme (FRGS/1/2015/ICT04/MMU/03/5) for supporting this work.

REFERENCES

- Boneh, D., Boyen, X., and Shacham, H. (2004). Short group signatures. In Franklin, M., editor, *Advances in Cryptology – CRYPTO 2004*, pages 41–55, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Boneh, D. and Franklin, M. (2001). Identity-based encryption from the weil pairing. In Kilian, J., editor, *Advances in Cryptology – CRYPTO 2001*, pages 213–229, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Boyar, J., Chaum, D., Damgård, I., and Pedersen, T. (1991). Convertible undeniable signatures. In Menezes, A. J. and Vanstone, S. A., editors, *Advances in Cryptology-CRYPTO' 90*, pages 189–205, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Chaum, D. and van Antwerpen, H. (1990). Undeniable signatures. In Brassard, G., editor, *Advances in Cryptology – CRYPTO' 89 Proceedings*, pages 212–216, New York, NY. Springer New York.
- Galbraith, S. D. and Mao, W. (2003). Invisibility and anonymity of undeniable and confirmer signatures. In *Cryptographers Track at the RSA Conference*, pages 80–97. Springer.
- Kurosawa, K. and Heng, S.-H. (2005). 3-move undeniable signature scheme. In Cramer, R., editor, *Advances in Cryptology – EUROCRYPT 2005*, pages 181–197, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Phong, L. T., Kurosawa, K., and Ogata, W. (2009). Provably secure convertible undeniable signatures with unambiguity. Cryptology ePrint Archive, Report 2009/394. <https://eprint.iacr.org/2009/394>.
- Phong, L. T., Kurosawa, K., and Ogata, W. (2010). Provably secure convertible undeniable signatures with unambiguity. In Garay, J. A. and De Prisco, R., editors, *Security and Cryptography for Networks*, pages 291–308, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Yuen, T. H., Au, M. H., Liu, J. K., and Susilo, W. (2007a). (convertible) undeniable signatures without random oracles. In *International Conference on Information and Communications Security*, pages 83–97. Springer.
- Yuen, T. H., Au, M. H., Liu, J. K., and Susilo, W. (2007b). (convertible) undeniable signatures without random oracles. Cryptology ePrint Archive, Report 2007/386. <https://eprint.iacr.org/2007/386>.
- Zhao, W. (2010). On the security of yuan et al.'s undeniable signature scheme. *International Journal of Network Security*, 11(3):177–180.

Searchable Symmetric Encryption: Defining Strength Against Query Recovery Attacks

Moesfa Soeheila Mohamad^{*1,2}, Syh-Yuan Tan³, and Ji-Jian Chin⁴

¹Information Security Lab, Mimos Berhad, Kuala Lumpur.

²Faculty of Computing and Informatics, Multimedia University, Cyberjaya.

³Faculty of Information Science and Technology, Multimedia University, Melaka.

⁴Faculty of Engineering, Multimedia University, Cyberjaya.

E-mail: *soeheila.mohamad@mimos.my, {³sytan,⁴jjchin}@mmu.edu.my

*Corresponding author

ABSTRACT

After many searchable symmetric schemes have been proven secure under existing security models, empirical evidence of successful attacks on the published schemes were published. The attacks indicate current security models do not cover query security. This work compares the \mathcal{L} -security to the count and file injection attacks. Finally, query indistinguishability security is introduced with a game-based definition using which a scheme may be proved to be strong against the query recovery attacks. The new security definition is proven to be consistent with current security definitions.

Keywords: searchable encryption, SSE, security model.

1 INTRODUCTION

Searchable symmetric encryption (SSE) is a category of schemes with ciphertext searching function utilising symmetric cryptographic algorithms. The purpose of SSE is for a user to store a document collection in a storage facility without exposing the documents to the storage owner or co-resident. In addition, the data owner or authorized users may perform searches to find documents of interest, while maintaining security.

In order to achieve sublinear search complexity, SSE scheme designs deploys inverted index where the index key is encoded keywords and the index values are document identifiers. Examples of such design includes Curtmola et al. (2006), Chase and Kamara (2010), Naveed et al. (2014), Cash et al. (2014) and Kamara and Moataz (2017). There are two types of SSE, static and dynamic. A static SSE scheme consists of six algorithms.

- KeyGen** This is a probabilistic algorithm run by the client. From a security parameter 1^k , this algorithm generates a set of symmetric keys, K , including an encryption key.
- BuildIndex** This algorithm is run by the client, taking the keyword-document identifiers mapping DB and keys K and output the index I .
- Encrypt** This algorithm is run by the client, non-interactive, and is usually probabilistic because of probabilistic symmetric encryption. For input a set of document \mathbf{D} and keys K , this algorithm outputs a set of ciphertexts \mathbf{c} of the documents.
- Trapdoor** This algorithm is run by the client and is usually deterministic. It takes as input the keyword w and the key K and outputs a trapdoor t_w .
- Search** This is a deterministic interactive algorithm run by the server. Inputs are t_w sent by the client and the index I stored on the server. This algorithm finds the set of document identifiers corresponding to documents containing the keyword w . The set of document identifiers being output is returned to the client.
- Decrypt** This deterministic algorithm is on the client. Taking input ciphertexts c_1, \dots, c_n and key K to output documents d_1, \dots, d_n .

Dynamic SSE schemes include another algorithm, Update which takes as input a document, the list of keywords and an operation name such as add, remove or modify. The algorithm outputs a new index and ciphertext. An SSE scheme is correct if the Search on a keyword followed by Decrypt produces all documents containing the queried keyword.

The trade off in enabling search on ciphertexts is the disclosure of some information regarding the documents, called leakage. Specifically the leakage function is $\mathcal{L}=(\mathcal{L}^{setup}, \mathcal{L}^{query})$. The first component, \mathcal{L}^{setup} , is the leakage from the index and ciphertexts after KeyGen, BuildIndex and Encrypt algorithms have been executed. The second component, \mathcal{L}^{query} , is leakage from the keyword trapdoor during search queries, after Trapdoor and Search have been executed. The leakage function \mathcal{L} varies from scheme to scheme. Nevertheless, \mathcal{L}^{setup} contains at least the length of ciphertexts, and \mathcal{L}^{query} contains at least the Access Pattern(AP) and the Query Pattern(QP). AP is a record of the document identifiers returned in each of the queries in the history. SP is a record of repeated (and unrepeated) queries in the query sequence in the history.

Despite the leakage, an SSE scheme aims to protect the confidentiality of the stored documents and the queried keyword using symmetric cryptographic schemes. Currently, the \mathcal{L} -security (Chase and Kamara, 2010) is the definition accepted for SSE schemes security. Proving an SSE scheme achieves \mathcal{L} -security, is in fact asserting that the scheme's leakage is safe to be disclosed.

However, attacks by Islam et al. (2012), Zhang et al. (2016) and Cash et al. (2015) were successful in recovering query keywords in published index-based SSE schemes. Wright and Pouloit (2017) generalized the attacks as statistic inference attack and produced a statistical method framework to detect such vulnerabilities in an SSE scheme. Cash et al. (2015) studied the practical attacks and defined the attack goals and adversary capabilities. In addition, they categorize SSE leakage profiles to identify the extent of vulnerability of an SSE design to the different attacks. Here we are only concerned with leakage profile L1 which reveals keyword occurrence pattern only after search queries are performed.

Our Contribution This work takes the provable security perspective by comparing the \mathcal{L} -security definition and the attacks. Then, a security game is proposed to define strength against the distribution-based query recovery attacks.

2 BRIEF ANALYSIS

The current indistinguishability (IND) and semantic security games for SSE security were defined by Curtmola et al. (2006). Then Chase and Kamara (2010) parameterized the leakage in the SSE security definition and named it the \mathcal{L} -security. This definition implies the semantic security of the stored documents and search queries except for the declared \mathcal{L} .

Definition 2.1 (\mathcal{L} -security under adaptive CKA). *Let $\Sigma = (\text{KeyGen}, \text{Encrypt}, \text{Trapdoor}, \text{Search}, \text{Decrypt})$ be an SSE scheme and consider the following probabilistic experiments where \mathcal{A} is an adversary, \mathcal{S} is a simulator and $\mathcal{L}_{\Sigma}^{\text{setup}}$ and $\mathcal{L}_{\Sigma}^{\text{query}}$ are stateful leakage algorithms of Σ :*

Real $_{\Sigma, \mathcal{A}}(k)$: *the challenger begins by running $\text{KeyGen}(1^k)$ to generate a key K . \mathcal{A} outputs a tuple (DB, \mathbf{D}) and receives $(I, \mathbf{c}) \leftarrow \text{Encrypt}_K(DB, \mathbf{D})$ from the challenger. The adversary makes a polynomial number of adaptive queries and, for each query q , receives a trapdoor $t \leftarrow \text{Trapdoor}_K(q)$ from the challenger. Finally \mathcal{A} returns a bit b that is output by the environment.*

Ideal $_{\Sigma, \mathcal{A}, \mathcal{S}}(k)$: *\mathcal{A} outputs a tuple (DB, \mathbf{D}) . Given $\mathcal{L}_{\Sigma}^{\text{setup}}(DB, \mathbf{D})$, \mathcal{S} generates and sends a pair (I, \mathbf{c}) to \mathcal{A} . The adversary makes a polynomial number of adaptive queries and for each query q the simulator is given $\mathcal{L}_{\Sigma}^{\text{query}}(q)$. The simulator returns a trapdoor t . Finally, \mathcal{A} returns a bit b that is output by the experiment.*

We say that Σ is $(\mathcal{L}_{\Sigma}^{\text{setup}}, \mathcal{L}_{\Sigma}^{\text{query}})$ -secure against adaptive chosen query attacks (CKA) if for all PPT adversaries \mathcal{A} , there exists a PPT simulator \mathcal{S} such that

$$|\Pr [\mathbf{Real}_{\Sigma, \mathcal{A}}(k) = 1] - \Pr [\mathbf{Ideal}_{\Sigma, \mathcal{A}, \mathcal{S}}(k) = 1]| \leq \text{negl}(k).$$

Islam et al. (2012) creates the first practical attack and is known as the IKK attack. In this attack, the attacker has the whole document set and some keyword-trapdoor pairs. From the document set, the attacker computed the keyword co-occurrence matrix. Then, by observing a number of search queries, the observed co-occurrence, inferred from AP in $\mathcal{L}^{\text{query}}$ and the estimated co-occurrence are put through simulated annealing to guess the keyword corresponding to the trapdoor. Then, Cash et al. (2015) defined count attack which identifies unique keyword frequency in the known document and finds a count match in AP in $\mathcal{L}^{\text{query}}$, before applying the IKK attack. File injection attack was created by Zhang et al. (2016). In this attack, the attacker introduces documents into the SSE to make its keyword distribution estimation more accurate. The attacker create documents containing intersecting subsets of keywords. If the inserted document appears in a search result disclosed by AP in $\mathcal{L}^{\text{query}}$, the attacker can infer the keyword of the search trapdoor. Less injected files are required in the attack if the attacker has partial knowledge of the stored documents.

		\mathcal{L} -security	Count Attack	File Injection Attack
Security Goals	Documents	Semantic security	Know document keywords	
	Queries	Semantic security	Query keyword recovery	
Adversary Capability	Documents	Chosen document	Known document	Chosen document
	Queries	Adaptively chosen query	Observed queries	

Table 1: Comparing security/attack goals and adversary capabilities of \mathcal{L} -security to count attack and file injection attack.

By comparing the \mathcal{L} -security game to the attacks, as in Table 1, the attacker is at most as powerful as the adversary in the game. However, the attacker is able to break the semantic security of both the documents and the search queries. One gap immediately identified is the usage of the complete knowledge about the keyword distribution from the generated document.

Applying the file injection attack for the count attack, an adversary can perform 100% query recovery with probability 1. If the adversary is given the power to generate the document set, the attacker can generate the document set such that each keyword has a unique number of documents containing it. As such, by the number of document identifiers in $\mathcal{L}^{\text{query}}$, the attacker can immediately identify the correct keyword. However, in the current Real-Ideal game this power is useless. The power is also useless for the **Ind** game in (Curtmola et al., 2006) because it assumes two different histories which produce the same trace (leakage).

Therefore, the gap between the security definitions and the attacks is query security. The Real-Ideal security game does not signify the vulnerability due to the stored documents keyword distribution revealed in $\mathcal{L}^{\text{query}}$.

3 NEW SECURITY DEFINITION

Here, a security game is defined to discriminate SSE schemes whose leakage enables query recovery attacks. The strength against query recovery is defined by providing assurance that even knowing the set of all keywords and access to the trapdoor oracle, the adversary would not be able to identify keywords of other trapdoors.

The adversary's prior knowledge determines the source of documents during the game initiation as follows.

- Without prior knowledge: Document set \mathbf{D} is in the environment.

- **Known document distribution:** \mathbf{D} is set by environment. Challenger gives adversary distribution information.
- **Known document:** \mathbf{D} is set by environment. Challenger gives adversary all or some document.

Definition 3.1 (Query indistinguishability(Q-IND)). *Let SSE be an index-based SSE scheme consisting of (KeyGen, Encrypt, Trapdoor, Search, Decrypt), $k \in \mathbb{N}$ be the security parameter and \mathcal{A} be an adversary.*

Initiation *The document set \mathbf{D} is generated according to the adversary knowledge above. The challenger \mathcal{C} generate the secret keys $\mathbf{K}=\text{KeyGen}(1^k)$ and index I on the document set \mathbf{D} . \mathcal{C} sends I , ciphertexts \mathbf{c} , set of all keywords \mathbf{W} to \mathcal{A} .*

Queries *\mathcal{A} is allowed to make adaptive trapdoor queries by keyword $w_i \in \mathbf{W}$ to obtain $t_i=\text{Trapdoor}(K,w_i)$.*

Challenge *Next, \mathcal{A} chooses two keywords $w_0, w_1 \in \mathbf{W}$ which has not been queried and submit to \mathcal{C} . \mathcal{C} randomly choose $b \in \{0, 1\}$ and give \mathcal{A} the corresponding trapdoor $t_b=\text{Trapdoor}(K,w_b)$. After the challenge is issued, \mathcal{A} can make more trapdoor queries except for w_0, w_1 .*

Response *Finally, \mathcal{A} outputs b' as a guess of b . The adversary \mathcal{A} wins if $b' = b$.*

The advantage of \mathcal{A} is defined as the probability of winning this game beyond guessing, $\text{Adv}_{\mathcal{A}}(k)=|\Pr [b = b'] - \frac{1}{2}|$ where the probability is over \mathcal{A} and \mathcal{C} 's coin tosses. An SSE scheme is said to achieve query indistinguishability if for any \mathbf{D} , $\text{Adv}_{\mathcal{A}}(k) \leq \text{negl}(k)$.

The schemes on which the count attack applies, would not achieve Q-IND because the distinguisher can use the attack to identify b correctly. On the other hand, schemes with less leakage, especially those which obfuscate the keyword distribution, will be able to achieve this.

Since Curtmola et al. (2006) has proven that IND implies semantic security under adaptive attacks, the soundness of the new security definition is demonstrated by proving that Q-IND implies IND and \mathcal{L} -security.

Theorem 3.1. *Adaptive Q-IND under chosen document and keyword attack implies **Ind** under adaptive chosen keyword attack (Curtmola et al., 2006).*

Proof. Assume there exists an adversary \mathcal{A} who has non-negligible advantage in the **Ind** game as defined by Curtmola et al. (2006). We show that there exists an adversary \mathcal{B} who has non-negligible advantage in guessing b correctly in the Q-IND game. Consider the adversary \mathcal{B} who works as below.

Setup

1. \mathcal{B} initiates **Ind** game and receives \mathbf{D}_0 and \mathbf{D}_1 from \mathcal{A} .
2. \mathcal{B} submits $\mathbf{D}=\mathbf{D}_0 \cup \mathbf{D}_1$ to the challenger \mathcal{C} who returns \mathbf{W} , I and \mathbf{c} .
3. \mathcal{B} create ciphertext set \mathbf{c}' by including in \mathbf{c}' exactly half of every set of equal length ciphertexts in \mathbf{c} . Next, index I' is created by generating a random entry such that $|I'[t_i]| = \frac{1}{2}|I[t_i]|$ for every key t_i of I .

4. \mathcal{B} gives \mathcal{A} I' and \mathbf{c}' .

Queries: For $i = 1$ to $q - 1$,

1. \mathcal{A} submits $(w_{0,i}, w_{1,i})$ to \mathcal{B} .
2. \mathcal{B} passes $w_{0,i}$ to \mathcal{C} and obtains trapdoor t_i .
3. \mathcal{B} gives t_i to \mathcal{A} .

Challenge: When \mathcal{A} submits $(w_{0,q}, w_{1,q})$, \mathcal{B} forwards $(w_{0,q}, w_{1,q})$ to \mathcal{C} . \mathcal{C} returns $t_b = \text{Trapdoor}(K, w_{b,q})$ where $b \xleftarrow{R} \{0, 1\}$ as the challenge for \mathcal{B} .

Response: \mathcal{B} passes t_b to \mathcal{A} and obtain a reply b' . If $b' = 0$ then \mathcal{B} submits 0 to \mathcal{C} , otherwise submits 1.

First, we argue that I' and \mathbf{c}' is indistinguishable from the index and ciphertexts for \mathbf{D}_b to \mathcal{A} . Since $\tau(\mathbf{D}_0) = \tau(\mathbf{D}_1)$, the trace of the index and ciphertexts for \mathcal{A} are exactly one half of the trace $\tau(\mathbf{D})$.

Denote the ciphertext size $|d_{i,j}|$ as $\ell_{i,j}$. Then $\tau(\mathbf{D}_0) = (\ell_{0,1}, \ell_{0,2}, \dots, \ell_{0,n})$ and $\tau(\mathbf{D}_1) = (\ell_{1,1}, \ell_{1,2}, \dots, \ell_{1,n})$. Since $\tau(\mathbf{D}_0) = \tau(\mathbf{D}_1)$, $\ell_{0,j} = \ell_{1,j}$ for all $j = 1, \dots, n$, let $\ell_j = \ell_{0,j} = \ell_{1,j}$, and hence $\tau(\mathbf{D}) = (\ell_1, \ell_1, \ell_2, \ell_2, \dots, \ell_n, \ell_n)$ because $\mathbf{D} = \mathbf{D}_0 \cup \mathbf{D}_1$. The constructed \mathbf{c}' consists of one ciphertext for each ℓ_j , and hence indistinguishable from the ciphertext sets for either \mathbf{D}_0 or \mathbf{D}_1 because it produce the same trace as $\tau(\mathbf{D}_0)$ or $\tau(\mathbf{D}_1)$. Similar argument applies to the indistinguishability of I' from I_{D_0} and I_{D_1} .

At the query stage, from \mathcal{A} 's perspective, it is playing the **Ind** game when its challenger chooses $b = 0$. The keywords $w_{0,i}$ passed to \mathcal{C} is input to the Trapdoor algorithm and hence t_i returned to \mathcal{A} is exactly what it will receive in the **Ind** game because it is produced using the correct key. By the construction of I' , $I'[t_i]$ will produce $\tau(w_i) = (\alpha(w_i), \sigma(w_i))$ such that $|\alpha_D(w_i)| = \frac{1}{2} |\alpha_D(w_i)|$ because $\tau(w_{0,i}) = \tau(w_{1,i})$. For the same reason, $\sigma(w_{0,i}) = \sigma(w_{1,i})$.

Secondly, we show that the probability of \mathcal{B} 's response to \mathcal{C} is correct with non-negligible probability. If the challenge given to \mathcal{B} is $\text{Trapdoor}(w_{0,q})$ then in \mathcal{A} 's perspective it has been receiving $(t_{0,1}, t_{0,1}, \dots, t_{0,q-1}, t_{0,q})$ and hence replies $b' = 0$ with probability $\frac{1}{2} + \text{Adv}_{\mathcal{A}}(k)$. On the other hand, if the challenge for \mathcal{B} is $\text{Trapdoor}(w_{1,q})$ then in \mathcal{A} 's perspective it has received $(t_{0,1}, t_{0,1}, \dots, t_{0,q-1}, t_{1,q})$ which is not consistent with the **Ind** challenger choosing $b = 0$ or $b = 1$. We assume here that \mathcal{A} will make a random guess. Therefore, we have that

$$\begin{aligned}
\Pr[\mathcal{B} \text{ wins}] &= \Pr[b = 0] \cdot \Pr[0 \leftarrow \mathcal{B} | b = 0] + \Pr[b = 1] \cdot \Pr[1 \leftarrow \mathcal{B} | b = 1] \\
&= \frac{1}{2} \Pr[b' = 0 | b = 0] + \frac{1}{2} \Pr[b' = 1 | b = 1] \\
&= \frac{1}{2} (1 + \text{Adv}_{\mathcal{A}}(k)) \\
&= \frac{1}{2} + \frac{1}{2} \text{Adv}_{\mathcal{A}}(k)
\end{aligned}$$

which means $\text{Adv}_{\mathcal{B}}(k) = \frac{1}{2} \text{Adv}_{\mathcal{A}}(k)$ and hence non-negligible.

Therefore, if an adversary which can distinguish between document sets exists, then an adversary who can distinguish queries exists. In conclusion, if an SSE scheme achieves adaptive query indistinguishability, then it also achieves adaptive (document set) indistinguishability (**Ind**). \square

Theorem 3.2. *Adaptive query indistinguishability under chosen document and keyword attack implies adaptive \mathcal{L} -security under chosen keyword attack (Chase and Kamara, 2010).*

Proof. Assume for any polynomial-sized simulator there exists an adversary \mathcal{A} and a distinguisher \mathcal{D} such that after \mathcal{A} makes q adaptive trapdoor queries, \mathcal{D} can distinguish the **Real** environment from **Ideal** with non-negligible advantage. We show that there exists an adversary \mathcal{B} who has non-negligible advantage in distinguishing queries. Consider \mathcal{B} below.

Setup

1. \mathcal{B} initiate the \mathcal{L} -security game.
2. \mathcal{A} submits a document set \mathbf{D} and a keyword-documents mapping DB to \mathcal{B} .
3. \mathcal{B} submits \mathbf{D} to \mathcal{C} and obtain $(I, \mathbf{c}, \mathbf{W})$.
4. \mathcal{B} gives (I, \mathbf{c}) to \mathcal{A} .

Queries: For $i = 1$ to $q - 1$

1. \mathcal{A} submits query w_i to \mathcal{B} .
2. \mathcal{B} submits w_i to \mathcal{C} and obtain trapdoor t_i .
3. \mathcal{B} gives t_i to \mathcal{A} .

Challenge

1. When \mathcal{A} submits the last query w_q .
2. \mathcal{B} chooses a keyword $\tilde{w} \in \mathbf{W}$ which has not been queried by \mathcal{A} .
3. \mathcal{B} submits $(w_0 = \tilde{w}, w_1 = w_q)$ to \mathcal{C} .
4. \mathcal{C} returns the challenge $t^* = \text{Trapdoor}(w_b)$ where $b \xleftarrow{R} \{0, 1\}$ to \mathcal{B} .
5. \mathcal{B} passes t^* to \mathcal{A} .

Response: Finally, \mathcal{A} replies b' to \mathcal{B} which is forwarded to \mathcal{C} as response from \mathcal{B} .

The game is played by \mathcal{B} in a way that \mathcal{A} is playing in the **Real** environment because $(I, \mathbf{c}, t_1, t_2, \dots, t_{q-1})$ is computed by \mathcal{C} using the SSE scheme. Hence, \mathcal{A} is receiving expected replies from \mathcal{B} except for the last query.

If \mathcal{C} chooses $b = 1$, \mathcal{A} would have $(I, \mathbf{c}, t_1, t_2, \dots, t_{q-1}, \text{Trapdoor}(w_q))$ which is a consistent **Real** environment replies. Hence, \mathcal{A} would output $b' = 1$ with probability $\frac{1}{2} + \text{Adv}_{\mathcal{A}}$. Otherwise, if \mathcal{C} chooses $b = 0$, \mathcal{A} would have $(I, \mathbf{c}, t_1, t_2, \dots, t_{q-1}, \text{Trapdoor}(\tilde{w}))$ which is not consistent with both **Real** and **Ideal**. In this case, \mathcal{A} may output $b' = 0$ or $b' = 1$ with equal

probability. Thus,

$$\begin{aligned}
 \Pr[\mathcal{B} \text{ wins}] &= \Pr[b = 0] \cdot \Pr[0 \leftarrow \mathcal{B} | b = 0] + \Pr[b = 1] \cdot \Pr[1 \leftarrow \mathcal{B} | b = 1] \\
 &= \frac{1}{2} \Pr[b' = 0 | b = 0] + \frac{1}{2} \Pr[b' = 1 | b = 1] \\
 &= \frac{1}{2} (1 + \text{Adv}_{\mathcal{A}}(k)) \\
 &= \frac{1}{2} + \frac{1}{2} \text{Adv}_{\mathcal{A}}(k)
 \end{aligned}$$

That implies $\text{Adv}_{\mathcal{B}}(k) = \frac{1}{2} \text{Adv}_{\mathcal{A}}(k)$ which is non-negligible.

This contradicts the assumption that Q-IND holds. This means that adversary \mathcal{A} cannot exist. Therefore, adaptive Q-IND implies adaptive \mathcal{L} -security under chosen keyword attack. \square

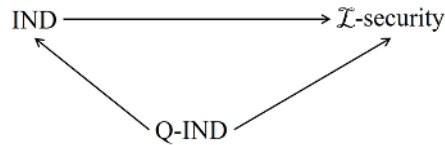


Figure 1: The implication relation of the Q-IND definition to the current IND-CKA and \mathcal{L} -security definitions.

By these theorem, we conclude that the Q-IND security is consistent with the existing SSE security definitions.

4 CONCLUSION

The gap between the SSE security definitions and the practical attacks is the significance of the keyword distribution in an SSE scheme search leakage to recover the query keywords. The query indistinguishability game is proposed to identify schemes which obfuscate keyword distribution information. Query indistinguishability implies both the IND and \mathcal{L} -security games. Nevertheless, the defining game for semantic security of both the documents and the queries which manifest safe leakage is still an open question.

REFERENCES

- Cash, D., Grubbs, P., Perry, J., and Ristenpart, T. (2015). Leakage-Abuse Attacks Against Searchable Encryption. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 668–679. ACM.
- Cash, D., Jaeger, J., Jarecki, S., Jutla, C., Krawczyk, H., Rosu, M.-C., and Steiner, M. (2014). Dynamic Searchable Encryption in Very Large Databases: Data Structures and Implementation. Cryptology ePrint Archive, Report 2014/853. <http://eprint.iacr.org/2014/853>.

- Chase, M. and Kamara, S. (2010). Structured Encryption and Controlled Disclosure. In Abe, M., editor, *ASIACRYPT 2010*, volume 6477 of *LNCS*, pages 577–594. Springer.
- Curtmola, R., Garay, J. A., Kamara, S., and Ostrovsky, R. (2006). Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions. In Juels, A., Wright, R. N., and di Vimercati, S. D. C., editors, *ACM Conference on Computer and Communications Security, CCS 2006*, pages 79–88. ACM.
- Islam, M. S., Kuzu, M., and Kantarcioglu, M. (2012). Access Pattern Disclosure on Searchable Encryption: Ramification, Attack and Mitigation. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012*. The Internet Society.
- Kamara, S. and Moataz, T. (2017). Boolean searchable symmetric encryption with worst-case sub-linear complexity. Cryptology ePrint Archive, Report 2017/126. <http://eprint.iacr.org/2017/126/>.
- Naveed, M., Prabhakaran, M., and Gunter, C. A. (2014). Dynamic Searchable Encryption via Blind Storage. In *2014 IEEE Symposium on Security and Privacy, SP 2014*, pages 639–654. IEEE Computer Society.
- Wright, C. V. and Pouloit, D. (2017). Early detection and analysis of leakage abuse vulnerabilities. Cryptology ePrint Archive, Report 2017/1052. <http://eprint.iacr.org/2017/1052/>.
- Zhang, Y., Katz, J., and Papamanthou, C. (2016). All Your Queries Are Belong To Us: The Power of File-Injection Attacks on Searchable Encryption. Cryptology ePrint Archive, Report 2016/172. <http://eprint.iacr.org/2016/172/>.

Enhanced AA_β Cryptosystem: The Design

Muhammad Asyraf Asbullah ^{*1}, Muhammad Rezal Kamel Ariffin^{1,2}, and Zahari Mahad¹

¹*Al-Kindi Cryptography Research Laboratory, Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

²*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

E-mail: ma_asyraf@upm.edu.my, rezal@upm.edu.my, zaharimahad@upm.edu.my

**Corresponding Author*

ABSTRACT

This paper presents an enhancement of the AA_β cryptosystem. Our approach is to adjust the key generation algorithm and incorporate the Rabin- p decryption techniques upon the AA_β decryption procedure. As a result, in comparison to any versions of AA_β cryptosystem prior to this work, the enhanced AA_β cryptosystem is superior in term of smaller key sizes, and have computational advantageous in decryption procedure. Consequently, we show that breaking the AA_β cryptosystem is computationally reduce to solving the Rabin- p cryptosystem, and (partially) vice versa.

Keywords: AA_β Cryptosystem, Rabin- p Cryptosystem, Internet of Things, integer factorization problem, embedded system

1 INTRODUCTION

The AA_β cryptosystem is a public key cryptosystem that was designed initially in earlier 2012. The objective is to visualize an asymmetric encryption that utilized the security instances by a so-called Bivariate Function Hard Problem (Mahad and Ariffin, 2012). In addition, the earlier design also able to highlight a solution for the decryption failure scenario of Rabin encryption.

A comparative analysis in Ariffin and Mahad (2012) shows that the AA_β cryptosystem is faster than RSA and ECC in term of encryption speed. However, the decryption speed was comparable with those two well known public key cryptosystems. A full-fledged version of the AA_β cryptosystem was introduced later in Ariffin et al. (2013). Furthermore, in this work, the authors conducted a number of mathematical analysis, along with updated experimental results

against ECC and RSA. Since then, the AA_β cryptosystem received a considerable attention from interested researchers.

In 2014, Asbullah and Ariffin (2014) proposed a fast decryption methods for AA_β cryptosystem. In this work, they replaced the computation of Chinese Remaindering Theorem (CRT) during the AA_β decryption procedure with the Garner's algorithm. Consequently, the modification makes the decryption become very fast and reduces the needed computational complexity. A comparative analysis was also provided, against Rabin-Takagi (Takagi, 1998) and HIME(R) cryptosystem (Nishioka et al., 2002). In 2016, Adnan et al. (2016a) proposed a practical implementation of the AA_β as a lightweight asymmetric encryption scheme on an embedded system device. On the other hand, Adnan et al. (2016b) focussed on a timing analysis of the AA_β encryption on embedded Linux for the Internet of Things (IoT). In the meantime, Adnan et al. (2017) give a result for energy analysis of the AA_β cryptosystem. Their results positively indicate that a possibility for AA_β cryptosystem is implemented for a lightweight public key encryption on an embedded device, hence suitable also for IoT.

A number of cryptanalyses were conducted upon the AA_β cryptosystem. For instance, we will survey several results on algebraic cryptanalysis and also side-channel cryptanalysis upon the said cryptosystem. The work in Ghafar and Ariffin (2014) shows that the AA_β cryptosystem is susceptible to a timing attack (i.e. a category of side channel attack). Fortunately, their result only discusses in the theoretical sense of timing attack with the assumption that the attacker able to collect some leaked values of a certain parameter during the decryption process. Furthermore, Asbullah and Ariffin (2016a) shows that there exist inappropriate keys selection that can be manipulated to break the cryptosystem (as analogous as to their prior work in Asbullah and Ariffin (2016b)). These observations due to the algebraic nature implicitly reside in the public and private keys. Hence, they suggest that the parameters chosen during key generation for AA_β cryptosystem need to be scrutinized and selected wisely before implementation. Later on, Ghafar and Ariffin (2016) designed a simple power analysis and show that the secret keys of the AA_β cryptosystem could be retrieved by using such method. We highlight that the above cryptanalytical results yet are far from practical to breaks the cryptosystem, nevertheless the results indeed helpful to the security measures of the AA_β cryptosystem.

Our contribution. In this work, we proposed a design that enhances the AA_β cryptosystem (of any version prior to this work). Our methodology is to adjust the key generation algorithm and incorporate the Rabin- p decryption techniques upon the AA_β decryption procedure. The reason (rationale) by doing so is that our enhanced AA_β cryptosystem will be as efficient as the Rabin- p cryptosystem. Furthermore, we provide the computational reducibility to Rabin- p cryptosystem, and vice versa (partially), in addition to the ones that explained in the original work of Ariffin et al. (2013).

This paper has been divided into five sections, begins with a brief overview of the AA_β cryptosystem in Section 1. Section 2 laying out the background and important materials for this research. Section 3 describes the design of our enhanced AA_β cryptosystem. In Section 4, we explain the design rationale for the enhanced version. In addition, we put forward the relation between our proposed cryptosystem with the security of the Rabin- p cryptosystem. Finally, we conclude the paper in Section 5.

2 PRELIMINARIES

2.1 AA_β Function

First of all, we will review the AA_β function which is proposed earlier by Ariffin et al. (2013). Consider the following definition.

Definition 2.1. (Ariffin et al., 2013). Suppose p, q be two distinct primes satisfies $3 \pmod{4}$ where $2^k < p, q < 2^{k+1}$. Let $A_1 \in \mathbb{Z}_{(2^{3k+4}, 2^{3k+6})}^+$ and $A_2 = p^2q$ such that $\gcd(A_1, A_2) = 1$. Suppose $m^2 \in \mathbb{Z}_{(2^{2k-2}, 2^{2k-1})}^+$ and $t \in \mathbb{Z}_{(2^{4k}, 2^{4k+1})}^+$. Then we define the following equation (1) as the AA_β function.

$$c = A_1m^2 + A_2t \tag{1}$$

Note that the integers (A_1, A_2) are known parameters and (m, t) be unknown integers to be solved.

Theorem 2.1. (Ariffin et al., 2013). Let $c = A_1m^2 + A_2t$ be AA_β function, then it has a unique solution for m and t , respectively.

The above theorem show that the AA_β function once solved will have a unique integers m and t , respectively.

2.2 The AA_β Cryptosystem

The details of the original design of the AA_β cryptosystem is given here, following the description in Ariffin et al. (2013). However, we only give a simplified version of the AA_β decryption algorithm due to Asbullah and Ariffin (2014). We now describe the key generation, encryption and decryption procedure of the original AA_β cryptosystem as follows.

Algorithm 2.1 Key Generation

Input: The security parameter k

Output: Public key tuple (A_1, A_2) and the private key tuple (d', p, q)

1. Select two distinct primes $2^k < p, q < 2^{k+1}$ satisfying $p, q \equiv 3 \pmod{4}$
 2. Calculate $A_2 = p^2q$
 3. Choose an integer $A_1 \in \{(2^{3k+4}, 2^{3k+6})\}$ randomly such that $\gcd(A_1, A_2) = 1$
 4. Determine the integer d' such that $A_1d' \equiv 1 \pmod{pq}$
 5. Output a tuple (A_1, A_2) as the public key and a tuple (d', p, q) as the private key.
-

Remark 2.1. Observe that the encryption algorithm for AA_β cryptosystem only involves the basic multiplications and additions.

Algorithm 2.2 Algorithm for Encryption

Input: The plaintext tuple (m, t) and the public key tuple (A_1, A_2) **Output:** A ciphertext c

1. Select a plaintext $m \in \{(2^{2k-2}, 2^{2k-1})\}$
 2. Select a plaintext $t \in \{(2^{4k}, 2^{4k+1})\}$
 3. Compute the ciphertext $c = A_1m^2 + A_2t$
-

Remark 2.2. Suppose we consider AA_β decryption algorithm as described in Ariffin et al. (2013) which need to solve simultaneous congruence equations using the Chinese Remaindering Theorem (CRT). However, there exists a faster and more efficient method to solve the CRT, namely the Garner's algorithm as appeared in the work of Asbullah and Ariffin (2014). Hence, we remark that the decryption algorithm for the AA_β used here is taken from the work of Asbullah and Ariffin (2014) since it is more efficient than its earlier original version as follows.

Algorithm 2.3 Algorithm for Decryption

Input: The ciphertext c and the private key tuple (d', p, q) **Output:** The plaintext tuple (m, t)

1. Calculate $w \equiv cd' \pmod{pq}$
 2. Determine $m_p \equiv w^{\frac{p+1}{4}} \pmod{p}$
 3. Determine $m_q \equiv w^{\frac{q+1}{4}} \pmod{q}$
 4. Determine $j \equiv p^{-1} \pmod{q}$
 5. Compute $h_1 \equiv (m_q - m_p)j \pmod{q}$
 6. Compute $h_2 \equiv (-m_q - m_p)j \pmod{q}$
 7. Compute $m_1 = m_p + h_1p$
 8. Compute $m_2 = m_p + h_2p$
 9. Compute $m_3 = pq - m_2$
 10. Compute $m_4 = pq - m_1$
 11. For $m_i < 2^{2k-1}$, determine $t_i = \frac{c - A_1m_i^2}{A_2}$ where $i = 1, 2, 3, 4$
 12. Sort the pair (m_i, t_i) for integer t_i , else reject
 13. Output the plaintext tuple (m, t)
-

2.3 Useful Lemmas

In this section we provides two important lemmas that will be useful later for our work in this paper.

Lemma 2.1. (Asbullah and Ariffin, 2016c). Let $p \equiv 3 \pmod{4}$ be a prime number. Let $c \equiv m^2 \pmod{p^2}$ where m is an unknown integer such that $m < p^2$ and $\gcd(m, p) = 1$. Then a solution

to $c \equiv m^2 \pmod{p^2}$ can be determine by $m_1 = m_p + jp$ where $m_p \equiv c^{\frac{p+1}{4}} \pmod{p}$, $j \equiv \frac{i}{2m_p} \pmod{p}$ such that $i = \frac{c-m_p^2}{p}$. Furthermore $m_2 = p^2 - m_1$ is the another solution.

Lemma 2.2. (Asbullah and Ariffin, 2016c). Let m_1 and m_2 be the two solutions from Lemma 2.1. Then one of the two solutions less than $\frac{p^2}{2}$.

3 AA_β CRYPTOSYSTEM: ENHANCED VERSION

This section is dedicated to describe the new design for enhanced version of the AA_β -cryptosystem. Then, we provide the proof of correctness.

3.1 The Proposed Cryptosystem

Algorithm 3.1 Key Generation of The Proposed Cryptosystem

Input: The security parameter k

Output: Public key tuple (A_1, A_2) and the private key tuple (d', p, q)

1. Select two distinct primes $2^k < p, q < 2^{k+1}$ satisfying $p, q \equiv 3 \pmod{4}$
 2. Calculate $A_2 = p^2q$
 3. Choose an integer $A_1 \in \{2^{3k+4}, 2^{3k+6}\}$ randomly such that $\gcd(A_1, A_2) = 1$
 4. Determine the integer d such that $A_1d \equiv 1 \pmod{p^2}$
 5. Output a tuple (A_1, A_2) as the public key and a tuple (d, p) as the private key.
-

Remark 3.1. Note that the following encryption algorithm (Algorithm 3.2) is identical to the original encryption algorithm (Algorithm 2.2).

Algorithm 3.2 Encryption Algorithm of The Proposed Cryptosystem

Input: The plaintext tuple (m, t) and the public key tuple (A_1, A_2)

Output: A ciphertext c

1. Select a plaintext $m \in \{2^{2k-2}, 2^{2k-1}\}$
 2. Select a plaintext $t \in \{2^{4k}, 2^{4k+1}\}$
 3. Compute the ciphertext $c = A_1m^2 + A_2t$
-

Algorithm 3.3 Decryption Algorithm for The Proposed Cryptosystem**Input:** A ciphertext c and the private key tuple (d, p) **Output:** The plaintext tuple (m, t)

- 1: Calculate $w \equiv dc \pmod{p^2}$
- 2: Calculate $m_p \equiv w^{\frac{p+1}{4}} \pmod{p}$
- 3: Calculate $i = \frac{w - m_p^2}{p}$
- 4: Calculate $j \equiv \frac{i}{2m_p} \pmod{p}$
- 5: Calculate $m_1 = m_p + jp$
- 6: Output $m = m_1$ if $m_1 < 2^{2k-1}$. Else return $m = p^2 - m_1$
- 7: Determine $t = \frac{c - A_1 m^2}{A_2}$
- 8: Output the plaintext tuple (m, t)

3.2 Proof of Correctness

Theorem 3.1. *Let $c = A_1 m^2 + A_2 t$ be the ciphertext output by Algorithm 3.2. Then such ciphertext will correctly decrypted by the Algorithm 3.3 and retrieved the plaintext tuple (m, t) .*

Proof. Suppose c be the ciphertext with parameters dictated in the Algorithm 3.2. The ciphertext c then can be decrypted efficiently as follows. Since d such that $dA_1 \equiv 1 \pmod{p^2}$, we proceed to determine w as follows.

$$w \equiv dc \equiv d(A_1 m^2 + A_2 t) \equiv dA_1 m^2 \equiv m^2 \pmod{p^2} \quad (2)$$

Then, by using Lemma 2.1, the equation (2) efficiently solved, which produces two distinct solution m_1 and m_2 . From Lemma 2.2, only one of m_1 and m_2 satisfies an integer less than $\frac{p^2}{2}$. Since we have $m < 2^{2k-1} < \frac{p^2}{2}$, thus we get the unique m . We further to compute $t = \frac{c - A_1 m^2}{A_2}$. Hence, analytically the ciphertext c correctly decrypted by the Algorithm 3.3 and output the unique solution of the tuple (m, t) . \square

4 DISCUSSIONS**4.1 Enhancement from the Original Version**

Table 1 illustrates the comparison between the original AA_β (Ariffin et al., 2013), the fast variant AA_β (Asbullah and Ariffin, 2014) and the proposed enhanced variant.

	Original AA_β	Fast AA_β	Enhanced Version
Public keys	$ A_1 , A_2 = 3k$	$ A_1 , A_2 = 3k$	$ A_1 , A_2 = 3k$
Private keys	$ d' = 2k, p , q = k$	$ d' = 2k, p , q = k$	$ d = 2k, p = k$
Mod Exponent	2	2	1
Mod Inverse	2	1	1
Mod Reduction	5	3	2
Division	4	4	2
Novak's attack	Yes	Yes	No

Table 1: Comparison between three versions of AA_β Cryptosystem in consideration.

Note that the enhanced version only use d, p as the private keys during key generation process while in the original version, in addition to the private keys d', p is another large prime q . Meaning that smaller key size hence leads to less storage and faster computation. Furthermore, this step emphasizes that we only require finding q for creating the public key $A_2 = p^2q$, then the large prime q can be discarded afterward.

The decryption process in our proposed cryptosystem takes advantage from the efficiency of the Rabin- p decryption algorithm, which only required a single prime p coupled with a private integer d instead of two primes p, q as in the original version, with additional private key d' . Such requirements would affect the overall operations in term of computational advantages. For instance, as shown in Table 1, the modular operations that is needed for the proposed enhanced version are minimal in comparison to the original version and the another variant in consideration.

Furthermore, the decryption process in our proposed cryptosystem takes advantage from the security feature provided in the Rabin- p cryptosystem in term of resistant against the Novak's attack; since the enhanced version of the proposed algorithm (i.e. Algorithm 3.3) does not execute the computation of the CRT or the Garner's method. Therefore, we claimed that the proposed enhanced version give additional security feature. Refer to (Asbullah and Ariffin, 2016c) for details.

4.2 Computational Reducibility

In the original version, Ariffin et al. (2013) show that for any efficient algorithm able to factor the modulus $A_2 = p^2q$, then such algorithm also able to solve the AA_β function. Furthermore, they also prove that the AA_β function can be solved if there exists algorithm that can solve the Bivariate Function Hard Problem (BFHP). This section will provide another cases regarding breaking the AA_β function (cryptosystem) with respect to the computational reducibility as follows.

Theorem 4.1. *Breaking the AA_β function (cryptosystem) is reducible to solving the Rabin- p cryptosystem.*

Proof. Suppose we are given a problem satisfies Definition 2.1, i.e. the AA_β function. Notice that there exists an integer x such that $A_1x \equiv 1 \pmod{A_2}$. Thus, we can compute $w' \equiv cx \equiv m^2 \pmod{A_2}$. Suppose there exists an algorithm able to solve the Rabin- p cryptosystem, then the same algorithm eventually able to solve $w' \equiv m^2 \pmod{A_2}$. Since the AA_β function (cryptosystem) has a unique solution for the integer m , hence we proceed to obtained the unique integer t such that $t = \frac{c-A_1m^2}{A_2}$. \square

Theorem 4.2. *Solving the Rabin- p cryptosystem is partially reduce to breaking the AA_β function (cryptosystem).*

Proof. Let $c \equiv m^2 \pmod{p^2q}$ be a ciphertext output by Rabin- p cryptosystem. Suppose there exists an algorithm able to solve the AA_β function, therefore the same algorithm can be used to solve the Rabin- p cryptosystem whenever m satisfies $2^{2k-2} < m < 2^{2k-1}$. Since the integer m from the Rabin- p cryptosystem is taken from $m \in \{(0, 2^{2k-1})\}$, yet such algorithm only efficiently solve for the set of integers in the range $(2^{2k-2}, 2^{2k-1})$. \square

5 CONCLUSION

In conclusion, this paper presents an enhancement of the AA_β cryptosystem of any versions prior to this work. The main idea is to incorporate the Rabin- p decryption method upon the original AA_β design. Consequently, this approach leads to an efficient AA_β cryptosystem in term of smaller key size and efficient decryption procedure. Furthermore, we show that breaking the AA_β function (cryptosystem) is reducible to solving the Rabin- p cryptosystem, and (partially) vice versa. Nevertheless, we plan to compare the enhanced AA_β cryptosystem and all of its predecessor version as a future work; in terms of complexity analysis, running time, memory consumption, hardware and software implementations, etc.

REFERENCES

- Adnan, S., Isa, M., and Hashim, H. (2016a). Implementation of the AA_β lightweight asymmetric encryption scheme on an embedded system device. *Advanced Science Letters*, 22(10):2910–2913.
- Adnan, S., Isa, M., and Hashim, H. (2016b). Timing analysis of the lightweight AA_β encryption scheme on embedded Linux for Internet of Things. In *ISCAIE 2016 - 2016 IEEE Symposium on Computer Applications and Industrial Electronics*, pages 113–116. IEEE.
- Adnan, S., Isa, M., and Hashim, H. (2017). Energy analysis of the AA_β lightweight asymmetric encryption scheme on an embedded device. In *IEACon 2016 - 2016 IEEE Industrial Electronics and Applications Conference*, pages 116–122. IEEE.
- Ariffin, M. R. K., Asbullah, M. A., Abu, N. A., and Mahad, Z. (2013). A New Efficient Asymmetric Cryptosystem Based on the Integer Factorization Problem of $N = p^2q$. *Malaysian Journal of Mathematical Sciences*, 7(S):19–37.

- Ariffin, M. R. K. and Mahad, Z. (2012). AA_β public key cryptosystem - A comparative analysis against RSA and ECC. In *Proceedings - 2012 7th International Conference on Computing and Convergence Technology (ICCIT, ICEI and ICACT), ICCCT 2012*, pages 589–594. Cited By :4.
- Asbullah, M. A. and Ariffin, M. R. K. (2014). Comparative Analysis of Three Asymmetric Encryption Schemes Based Upon the Intractability of Square Roots Modulo $N = p^2q$. In *The 4th International Cryptology and Information Security Conference 2014 (Cryptology2014)*, pages 86–99.
- Asbullah, M. A. and Ariffin, M. R. K. (2016a). Analysis on the AA_β Cryptosystem. In *The 5th International Cryptology and Information Security Conference 2016 (Cryptology2016)*, pages 41–48.
- Asbullah, M. A. and Ariffin, M. R. K. (2016b). Analysis on the Rabin- p cryptosystem. In *The 4th International Conference on Fundamental and Applied Sciences (ICFAS2016)*, pages 080012–1–080012–8. AIP Conf. Proc. 1787.
- Asbullah, M. A. and Ariffin, M. R. K. (2016c). Design of Rabin-like Cryptosystem without Decryption Failure. *Malaysian Journal of Mathematical Sciences*, 10(S):1–18.
- Ghfar, A. H. A. and Ariffin, M. R. K. (2014). Timing Attack Analysis on AA_β Cryptosystem. *Journal of Computer and Communication*, 2:1–9.
- Ghfar, A. H. A. and Ariffin, M. R. K. (2016). SPA on Rabin variant with public key $N = p^2q$. *Journal of Cryptographic Engineering*, 6(4):339–346.
- Mahad, Z. and Ariffin, M. R. K. (2012). AA_β public key cryptosystem - A new practical asymmetric implementation based on the square root problem. In *Proceedings - 2012 7th International Conference on Computing and Convergence Technology (ICCIT, ICEI and ICACT), ICCCT 2012*, pages 584–588. Cited By :1.
- Nishioka, M., Satoh, H., and Sakurai, K. (2002). Design and Analysis of Fast Provably Secure Public-Key Cryptosystems Based on a Modular Squaring. In *Information Security And Cryptology - ICISC 2001*, pages 81–102.
- Takagi, T. (1998). Fast RSA-Type Cryptosystem Modulo p^kq . *Advances in Cryptology-CRYPTO'98*, 1462:318–326.

Extending Pollard Class of Factorable RSA Modulus

Amir Hamzah Abd Ghafar^{*1}, Muhammad Rezal Kamel Ariffin^{1,2}, and
Muhammad Asyraf Asbullah^{1,3}

¹*Al-Kindi Cryptography Research Laboratory, Institute for Mathematical Research,
Universiti Putra Malaysia*

²*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia*

³*Centre of Foundation Studies for Agricultural Science, Universiti Putra Malaysia*

E-mail: amirghafar87@gmail.com

**Corresponding author*

ABSTRACT

Pollard $p - 1$ method is able to solve integer factorization problem if the targeted composite number has small prime factors. This is a reason why implementation in key generation algorithm of RSA cryptosystem requires its primes, p and q not to be constituted by small primes. In this paper, we showed another method which targets $N = pq$ and manipulates it against $p - 1$ and $q - 1$ structures. We remarked here that both $p - 1$ and $q - 1$ do not have small prime factors hence they can be generated without error by RSA libraries in current practice.

Keywords: Integer factorization problem, RSA cryptosystem, Pollard $p - 1$ algorithm

1 INTRODUCTION

Integer factorization problem (IFP) has intrigued scholars for a long time. The problem to find the decomposition of an integer is trivial if the integer is relatively small but becoming more computationally infeasible as the integer increases in value. Trial division is the most arduous approach to solve IFP. Given integer N to be factored, the method works by dividing an integer one by one with N and outputs the integer as the factor of N if the division produces another integer. In 1974, Pollard introduced a special-purpose method that can solve IFP if N constitutes specific types of factors (Pollard, 1974). The significance of IFP appeals cryptanalysts when RSA cryptosystem adopted IFP to be the source of its security strength in 1977 Rivest et al. (1978). Since then, many algorithms are invented either to solve IFP with specific-structured N (similar to Pollard's method) which is called special-purpose algorithms or general-structured of N which is called general-purpose algorithms. The importance of these special-purpose algorithms

is that they become the tools to outline the criterion in choosing safe RSA parameters which can only be attacked by general-purpose factorization algorithms at most in sub-exponential running time. In this paper, we introduce a new special-purpose factorization algorithm which specifies on the structure of both $p - 1$ and $q - 1$ where $N = pq$ is an RSA modulus.

1.1 RSA Cryptosystem

Before we go further, it is important to identify the parameters of RSA cryptosystem. One of the parameter, known as RSA modulus is $N = pq$ where p and q are two distinct primes and a corresponding Euler's function of N is also calculated, that is, $\phi(N) = (p - 1)(q - 1)$. To dismiss the possibility of N can be factored using trial division, p and q are chosen such that $p < q < 2p$. Using the value of $\phi(N)$, one select a random $e < \phi(N)$ and compute $d \equiv e^{-1} \pmod{\phi(N)}$. In RSA cryptosystem, N and e are called public keys while $p, q, d, \phi(N)$ are private keys.

We remark that the secrecy of $\phi(N)$ is crucial to the security of the RSA cryptosystem since knowing such integer leads to factor N in polynomial time. Suppose the integer $\phi(N)$ is known for some reasons, then we explain why this is the case as follows. We already know that $N = pq$. If we know $\phi(N)$, then we will have

$$\begin{aligned} N - \phi(N) + 1 &= pq - (p - 1)(q - 1) + 1 \\ &= p + q. \end{aligned}$$

We can see that

$$(X - p)(X - q) = X^2 - (p + q)X + pq.$$

Thus, by applying formula to find roots in quadratic equation, we can find

$$p, q = \frac{(p + q) \pm \sqrt{(p + q)^2 - 4(1)(pq)}}{(2)(1)}.$$

Thus, if there exists an algorithm that is able to find $\phi(N)$, then such algorithm able to efficiently factor the RSA modulus $N = pq$ (Asbullah and Ariffin, 2016).

1.2 Our Contributions

In this paper, we present a new class of weak primes that fulfill the required conditions in Section 2.3 but still can be factored in polynomial time. This new class have a specific structure that eases the factorization process and it only requires the public modulus N . This special structure may be exposed by current factorization algorithm if the size of N is considerably small, however as the necessity for security keep increasing over time, the size of N will be larger over time. Although current technology may not factor this larger N but we prove that our proposed method still can solve it in polynomial time.

1.3 Outline of This Paper

In Section 2, we introduce several algorithms called special-purpose factoring algorithms that may give initial idea of factoring exercise on special-structured primes used in RSA. In the same section, we also provide current guideline to generate the primes. In Section 3, we propose our first special-purpose factoring method along with its numerical example. Then we introduce the second special-purpose factoring method in Section 4 along with its generalized algorithm and numerical example. We conclude our result in the final section.

2 SOME PREVIOUS SPECIAL-PURPOSE FACTORING ALGORITHMS AND CURRENT STANDARD OF RSA PRIMES

2.1 Pollard $p - 1$ Algorithm

Pollard $p - 1$ algorithm is a special purpose algorithm to factor an integer N which has p as one of its factor where $p - 1$ is a B -smooth number. That is, $p - 1$ can be broken completely into small prime factors that are less than an integer, B . The algorithm manipulates Fermat's Little Theorem that states for all positive integers k and integers x that coprime to p , we have

$$a^{k(p-1)} \equiv 1 \pmod{p}.$$

To utilize the algorithm, an adversary has to choose a suitable B which is the guessed bound of the smoothness of $p - 1$. Then she has to compute

$$L = \prod_{\text{primes } \rho \leq B} \rho^{\lfloor \log_{\rho} B \rfloor}.$$

If $p - 1$ is a B -smooth number then the adversary can factor N when she computes $\gcd(a^L - 1, N) = p$ because

$$a^L - 1 \equiv a^{k(p-1)} - 1 \equiv 0 \pmod{p}$$

for some integer k . The adversary can extend the stages of this algorithm to factor $N = pq$ which $p - 1$ has one prime factor larger than B_1 but less than B_2 while the remaining factors are less than B_1 where $B_2 \gg B_1$. To achieve it, she has to compute

$$M = \prod_{\text{primes } \rho' \in (B_1, B_2]} (a^L)^{\rho'} - 1$$

and compute $\gcd(a^M - 1, N)$ to factor N . The cost of computation for Pollard $p - 1$ algorithm increases when B (for one-stage algorithm) or B_2 (for two-stages algorithm) increases.

2.2 Elliptic Curve Factoring (ECM) Algorithm

This algorithm was introduced by Lenstra Jr (1987) which fundamentally replacing the multiplicative group used in Pollard $p - 1$ algorithm to the group of points in a random elliptic curve.

By doing that, the adversary can re-execute the attack by choosing different elliptic curves if the initial elliptic curve chosen is not suitable, until a random elliptic curves with smooth order of $\mathbb{Z} + p$ is found. This cannot be done in computation of a multiplicative group as in Pollard $p - 1$ algorithm.

2.3 Current Standard of RSA Primes

Up until this point, we can see that if $p - 1$ is a B -smooth number then $N = pq$ can be factored in polynomial time by both Pollard $p - 1$ and ECM algorithms where B is a suitably small and $B_{pollard} < B_{ECM}$. Hence, several papers have suggested method to find ‘strong’ primes including by Williams and Schmid (Williams and Schmid, 1979) and Gordon (Gordon, 1984). However in 1999, Rivest and Silverman Rivest and Silverman (1997) noted that as long as the size of the modulus, N used in RSA is large enough (say, 2048-bit) then the necessity of finding such a strong prime is needless.

The latest standard by US National Institute of Standards and Technology (NIST) for generating RSA primes, FIPS PUB 186-4, stated in Appendix B.3 that p and q of RSA-2048 and above can be generated using random primes as long as the primes are provable or probable primes and satisfy the followings:

1. Size of p must be $\sqrt{2} \cdot 2^{(nlen/2)-1} < p < 2^{(nlen/2)} - 1$ where $nlen$ is the size of N ;
2. Size of q must be $\sqrt{2} \cdot 2^{(nlen/2)-1} < q < 2^{(nlen/2)} - 1$ where $nlen$ is the size of N ; and
3. Size of $|p - q|$ must be greater than $2^{(nlen/2)-100}$ where $nlen$ is the size of N .

More details on the standard can be read here (FIPS, 2013).

3 FIRST ATTACK

We begin our first attack with the next lemma, which shows that r can be a positive value less than 1 if the condition of r with respect to the value of a is satisfied.

Lemma 3.1. *Suppose $a, r \in \mathbb{Z}$. Let $a^2 + r$ with $0 < r < 2a + 1$. Then $\sqrt{a^2 + r} = a + \epsilon$ where $0 < \epsilon < 1$.*

Proof. Let $a^2 + r$ be an integer where $0 < r < 2a + 1$ such that $a, r \in \mathbb{Z}$. We have $a^2 < a^2 + r < a^2 + 2a + 1$. Taking the square roots, we have

$$\begin{aligned} a &< \sqrt{a^2 + r} < \sqrt{(a + 1)^2} \\ a &< \sqrt{a^2 + r} < a + 1. \end{aligned}$$

Hence, there exists a number ϵ such that $\sqrt{a^2 + r} = a + \epsilon$ where $0 < \epsilon < 1$. \square

Next, we present our first attack in the following theorem. Note that the first attack considers $N = pq$ for p, q in the form of $p = a^2 + 1$ and $q = b^2 + 1$, for some integers a, b , respectively.

Theorem 3.1. *Suppose $a < b < 2a$ where a, b are n -bit numbers. Let $\alpha = \frac{1}{2u}$ and $\beta = \frac{1}{2v}$ for $u > a^2$ and $v > b^2$. If $\epsilon_1 = \frac{1}{2a} - \frac{1}{2u}$ and $\epsilon_2 = \frac{1}{2b} - \frac{1}{2v}$ then $1 < \sqrt{(a^2 + 1)(b^2 + 1)} - ab < 1 + \delta$ for some $0 < \delta < 1$.*

Proof. Based on Lemma 3.1, we can see that

$$\begin{aligned} \sqrt{(a^2 + 1)(b^2 + 1)} &= \sqrt{a^2 + 1}\sqrt{b^2 + 1} \\ &= (a + \epsilon_1)(b + \epsilon_2) \\ &= ab + a\epsilon_2 + b\epsilon_1 + \epsilon_1\epsilon_2. \end{aligned} \quad (1)$$

From equation (1), if $\epsilon_1 = \frac{1}{2a} - \frac{1}{2u}$ and $\epsilon_2 = \frac{1}{2b} - \frac{1}{2v}$, we get

$$\begin{aligned} \sqrt{(a^2 + 1)(b^2 + 1)} - ab &= a\epsilon_2 + b\epsilon_1 + \epsilon_1\epsilon_2 \\ &= a\left(\frac{1}{2b} - \beta\right) + b\left(\frac{1}{2a} - \alpha\right) + \left(\frac{1}{2a} - \frac{1}{2u}\right)\left(\frac{1}{2b} - \frac{1}{2v}\right) \quad (2) \\ &= a\left(\frac{1}{2b} - \beta\right) + b\left(\frac{1}{2a} - \alpha\right) + \frac{1}{4}\left(\frac{1}{ab} - \frac{1}{ub} - \frac{1}{va} + \frac{1}{uv}\right) \\ &= a\left(\frac{1}{2b} - \beta\right) + b\left(\frac{1}{2a} - \alpha\right) + \frac{1}{4}\left(\frac{uv - av - bu + ab}{abuv}\right). \quad (3) \end{aligned}$$

The term $\frac{1}{4}\left(\frac{uv - av - bu + ab}{abuv}\right)$ is negligible since $|uv - av - bu + ab| < abuv$. Now, we need to show that

$$1 < a\left(\frac{1}{2b} - \beta\right) + b\left(\frac{1}{2a} - \alpha\right) < 1 + \delta. \quad (4)$$

to complete the proof. We can see that

$$\begin{aligned} a\left(\frac{1}{2b} - \beta\right) + b\left(\frac{1}{2a} - \alpha\right) &= \frac{a}{2b} - a\beta + \frac{b}{2a} - b\alpha \\ &= \frac{a^2 + b^2}{2ab} - (a\beta + b\alpha). \end{aligned}$$

where $(a\beta + b\alpha)$ is a negligible value because

$$\begin{aligned} (a\beta + b\alpha) &= \frac{a}{2u} + \frac{b}{2v} \\ &\approx \frac{2^{n+1}}{2^{2n}} \\ &< \frac{1}{2^{n-1}} \end{aligned}$$

Furthering we have

$$\begin{aligned} \frac{a^2 + b^2}{2ab} &= \frac{(b - a)^2 + 2ab}{2ab} \\ &= 1 + \frac{(b - a)^2}{2ab} \\ &< 2. \end{aligned}$$

From $a < b < 2a$, we have

$$\begin{aligned} 0 < b - a < 2a - a &\Rightarrow 0 < b - a < a \\ &\Rightarrow 0 < (b - a)^2 < a^2 < ab < 2ab. \end{aligned}$$

Thus $0 < \frac{(b-a)^2}{2ab} < 1$. Hence,

$$1 < a \left(\frac{1}{2b} + \beta \right) + b \left(\frac{1}{2a} + \alpha \right) < 1 + \delta \quad (5)$$

where $\frac{(b-a)^2}{2ab} < \delta < 1$. This completes the proof. \square

Corollary 3.1. *If a and b be two distinct integers such that $a < b < 2a$ then $\left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor - ab = 1$.*

Proof. Let

$$\sqrt{(a^2 + 1)(b^2 + 1)} = \left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor + \epsilon \quad (6)$$

From Theorem 3.1, we have $1 < \sqrt{(a^2 + 1)(b^2 + 1)} - ab < 1 + \delta$. If $0 < \delta < 1$ then (6) becomes

$$\begin{aligned} 1 < \left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor + \epsilon - ab < 1 + 1 \\ 1 < \left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor + \epsilon - ab < 2 \\ 0 < \left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor - ab \leq 1 \end{aligned}$$

where ϵ is a small positive number such that $\epsilon \leq 1$. As $\left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor$ is an integer,

$$\left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor - ab = 1.$$

\square

From the previous theorem, we conduct an attack against RSA in the next theorem.

Theorem 3.2. *Let $N = pq$ be a valid RSA modulus where $p = a^2 + 1$ and $q = b^2 + 1$ for any positive integer a, b such that $a < b < 2a$. Then N can be factored in polynomial time.*

Proof. We can see that

$$N = pq = (a^2 + 1)(b^2 + 1) \quad (7)$$

and

$$\begin{aligned} \phi(N) &= (p - 1)(q - 1) = a^2b^2. \\ \phi(N)^{0.5} &= ((p - 1)(q - 1))^{0.5} = ab. \end{aligned} \quad (8)$$

From Corollary 3.1 we know that

$$\left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor - ab = 1 \quad (9)$$

if a and b are two distinct positive integers such that $a < b < 2a$. Equation (9) can also be written as

$$\left\lfloor \sqrt{(a^2 + 1)(b^2 + 1)} \right\rfloor - 1 = ab$$

or

$$\left\lfloor \sqrt{N} \right\rfloor - 1 = \phi(N)^{0.5}$$

as shown in equations (7) and (8). Then

$$\left(\left\lfloor \sqrt{N} \right\rfloor - 1 \right)^2 = \phi(N).$$

That is we can obtain $\phi(N)$ only by knowing N which can be obtained publicly. We know that by knowing the value of $\phi(N)$, we can factor N . Hence, N can be factored in polynomial time. \square

The following proposition shows that we have conducted a successful attempt on primes of the form $p - 1 = a^2$ but a^2 is not a product of small primes.

Proposition 3.1. *If $a < b < 2a$ and $p = a^2 + 1 = (2s_1t_1)^2 + 1$ and $q = b^2 + 1 = (2s_2t_2)^2 + 1$ where s_1, s_2, t_1, t_2 are large primes with same bit size then $N = pq$ can only be factored using Theorem 3.2.*

Proof. To prove this, first we assume that if $a < b < 2a$ and $p = a^2 + 1$ and $q = b^2 + 1$ then $N = pq$ can be factored in polynomial time using current integer factorization method. This is a realistic assumption as algorithms in the previous section show that if $p - 1$ and $q - 1$ have small prime factors then N can be factored. We must notify that a^2 and b^2 must be even numbers for p and q to be prime. Let

$$p = a^2 + 1 = (2s_1t_1)^2 + 1 \quad (10)$$

and

$$q = b^2 + 1 = (2s_2t_2)^2 + 1 \quad (11)$$

where s_1, s_2, t_1, t_2 are large primes with same bit size. If $N = pq$ has $2n$ -bit in size, then

$$\begin{aligned} N &= pq \\ &= (a^2 + 1)(b^2 + 1) \\ &= ((2s_1t_1)^2 + 1)((2s_2t_2)^2 + 1) \end{aligned} \quad (12)$$

must be larger than 2^{2n-1} . If $a < b < 2a$ then $s_1t_1 < s_2t_2$ and (12) becomes

$$\begin{aligned} N &> ((2s_1t_1)^2 + 1)((2s_1t_1)^2 + 1) \\ &= ((2s_1t_1)^2 + 1)^2 \\ &> 2^{2n-1}. \end{aligned} \quad (13)$$

That is,

$$\begin{aligned}
 ((2s_1t_1)^2 + 1) &> 2^{n-\frac{1}{2}} \\
 ((2s_1t_1)^2) &> 2^{n-\frac{1}{2}} - 1 \\
 2s_1t_1 &> 2^{\frac{n}{2}-\frac{1}{4}} - 1 \\
 s_1t_1 &> 2^{\frac{n}{2}-\frac{5}{4}} - 1.
 \end{aligned} \tag{14}$$

As s_1 and t_1 has the same bit size then (14) can be approximated to

$$\begin{aligned}
 t_1^2 &> 2^{\frac{n}{2}-2} \\
 t_1 &> 2^{\frac{n}{4}-1}.
 \end{aligned}$$

Thus, each $s_1, t_1, s_2, t_2 > 2^{\frac{n}{4}-1}$. But if each $R_1 = s_1t_1$ and $R_2 = s_2t_2$ are large integers, current integer factorization method can not find s_1, t_1, s_2, t_2 given R_1 and R_2 . This is a contradiction with our assumption that if $a < b < 2a$ and $p = a^2 + 1$ and $q = b^2 + 1$ then $N = pq$ can be factored in polynomial time using current integer factorization method. Hence, only method shown in Theorem 3.2 can solve for $N = pq$ where $p = a^2 + 1 = (2s_1t_1)^2 + 1$ and $q = b^2 + 1 = (2s_2t_2)^2 + 1$ in all cases. \square

Now we present a numerical example for this attack.

Example 3.1. We use RSA-2048 modulus in this example. Specifically, we are given

$N =$ 277760248864169968873158968067901869922547278238355037590136727
 0240715147675160012833624528218201455199166813185429130368617419
 3971154144972985148142858883130417360160403706126708493567588828
 5980909262646789339002894853982864784073453668970595805015360357
 0349305396657254928545681160160413746193711515404385110255141997
 8532249156454524031480888840656697182797180989644228817301086228
 9252200300708568057626781500410230811407239230858237420913911912
 6939905482579660903725542803158791582967891614203750908579865707
 7703813533573519947179170160430599543149618207062701368130987691
 877537886244483937140946828720325250085369.

Then we calculate

$$\begin{aligned}
 L &= \left(\lfloor \sqrt{N} \rfloor - 1 \right)^2 \\
 &= 277760248864169968873158968067901869922547278238355037590136727 \\
 &024071514767516001283362452821820145519916681318542913036861741 \\
 &939711541449729851481428588831304173601604037061267084935675888 \\
 &285980909262646789339002894853982864784073453668970595805015360 \\
 &357034930539665725492854568116016041374619371151540438507686755 \\
 &931874570131235514879897396406589333629534460281975037863868702 \\
 &282117880551121820534487752559896204193123400136248129488310265 \\
 &889163447275523981760970896848274976635369964584038260481236430 \\
 &583171754578092224377606463756058793317856482872950763107180571 \\
 &33604199594424655882993586502629167421250332318096.
 \end{aligned}$$

Given N and L , we can calculate

$$p, q = \frac{(N - L + 1) \pm \sqrt{(N - L + 1)^2 - 4(1)(N)}}{(2)(1)}.$$

as $(X - p)(X - q) = X^2 - (N - L + 1)X + N$ as $N - L + 1 = p + q$. This is true because L is equal to ϕ . That is,

$$\begin{aligned} p = & 157333641302400043936149808430258159069378322856315566983292525 \\ & 651386768042551505902942334333741125818724724985312473023318232 \\ & 768860695131654895606976423743720919403690906598556269120123134 \\ & 487460216852133338533382436839315503082749221064232498772450702 \\ & 872681030062417577979661502327282505682061567696014789317 \end{aligned}$$

and

$$\begin{aligned} q = & 176542185488675192096753227611483089112345011373035585808847916 \\ & 779044089974585659014946469023246296341695878783488237274954642 \\ & 142204107143269775895451196768226651196843017780276951392192854 \\ & 897982523487768907470928698919985300481531877749662363550680049 \\ & 110629967321074705133568859163068132635599731378902977957 \end{aligned}$$

Remark 3.1. Both p and q in example 3.1 satisfy the conditions stated in FIPS PUB 186-4.

4 SECOND ATTACK

This attack considers both of the RSA primes, p and q to take form of $a^m + 1$ and $b^m + 1$ respectively where m is an even number. It is trivial to see that the attack is actually a generalized form of the first attack where $m = 2$. The attack is shown in the next theorem.

Lemma 4.1. Let m be a power of 2 and $a < b < 2a$. Then $\frac{a^m + b^m}{2(ab)^{m/2}} < m + \delta$ where $0 < \delta < 1$.

Proof. Using binomial expansion, we have

$$\begin{aligned} (b - a)^m &= \binom{m}{0}(-a)^m b^0 + \binom{m}{1}(-a)^{m-1} b^1 + \binom{m}{2}(-a)^{m-2} b^2 \dots + \\ &\quad \binom{m}{m-2}(-a)^2 b^{m-2} + \binom{m}{m-1}(-a)^1 b^{m-1} + \binom{m}{m}(-a)^0 b^m \\ &= (-a)^m + b^m + C \\ &= a^m + b^m + C \end{aligned} \tag{15}$$

as m will always be a positive even integer where

$$C = \binom{m}{1}(-a)^{m-1} b^1 + \binom{m}{2}(-a)^{m-2} b^2 + \dots + \binom{m}{m-2}(-a)^2 b^{m-2} + \binom{m}{m-1}(-a)^1 b^{m-1}. \tag{16}$$

From (15), we get

$$\frac{a^m + b^m}{2(ab)^{m/2}} = \frac{(b-a)^m - C}{2(ab)^{m/2}}. \quad (17)$$

We can rearrange equation (16) to get

$$\begin{aligned} C &= \binom{m}{2} a^{m-2} b^2 + \binom{m}{4} a^{m-4} b^4 + \dots + \binom{m}{m-4} a^4 b^{m-4} + \binom{m}{m-2} a^2 b^{m-2} - \\ &\quad \binom{m}{1} a^{m-1} b - \binom{m}{3} a^{m-3} b^3 - \dots - \binom{m}{m-3} a^3 b^{m-3} + \binom{m}{m-1} a b^{m-1} \\ &= \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i - \sum_{\substack{i \text{ is odd} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i. \end{aligned} \quad (18)$$

Then (17) will become

$$\begin{aligned} \frac{a^m + b^m}{2(ab)^{m/2}} &= \frac{1}{2(ab)^{m/2}} \left((b-a)^m - \left(\sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i - \sum_{\substack{i \text{ is odd} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i \right) \right) \\ &= \frac{1}{2(ab)^{m/2}} \left((b-a)^m + \sum_{\substack{i \text{ is odd} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i - \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} a^{m-i} b^i \right) \\ &= \frac{(b-a)^m}{2(ab)^{m/2}} + \sum_{\substack{i \text{ is odd} \\ 0 < i < m}} \frac{\binom{m}{i} a^{m-i} b^i}{2(ab)^{m/2}} - \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \frac{\binom{m}{i} a^{m-i} b^i}{2(ab)^{m/2}} \\ &= \frac{(b-a)^m}{2(ab)^{m/2}} + \sum_{\substack{i \text{ is odd} \\ 0 < i \leq m/2}} \frac{\binom{m}{i}}{2} \left(\frac{a}{b}\right)^{\frac{m}{2}-i} + \sum_{\substack{i \text{ is odd} \\ m/2 < i < m}} \frac{\binom{m}{i}}{2} \left(\frac{b}{a}\right)^{\frac{m}{2}-i} \\ &\quad - \sum_{\substack{i \text{ is even} \\ 0 < i \leq m/2}} \frac{\binom{m}{i}}{2} \left(\frac{a}{b}\right)^{\frac{m}{2}-i} - \sum_{\substack{i \text{ is even} \\ m/2 < i < m}} \frac{\binom{m}{i}}{2} \left(\frac{b}{a}\right)^{\frac{m}{2}-i} \end{aligned} \quad (19)$$

If $a < b < 2a$ then

$$\frac{a}{b} < 1 \quad \text{and} \quad \frac{b}{a} < 2.$$

Thus (19) will become

$$\begin{aligned}
 \frac{a^m + b^m}{2(ab)^{m/2}} &< \sum_{\substack{i \text{ is odd} \\ 0 < i \leq m/2}} \frac{\binom{m}{i}}{2} + \sum_{\substack{i \text{ is odd} \\ m/2 < i < m}} \frac{\binom{m}{i}}{2} 2^{\frac{m}{2}-i} - \sum_{\substack{i \text{ is even} \\ 0 < i \leq m/2}} \frac{\binom{m}{i}}{2} - \sum_{\substack{i \text{ is even} \\ m/2 < i < m}} \frac{\binom{m}{i}}{2} 2^{\frac{m}{2}-i} \\
 &+ \frac{(b-a)^m}{2(ab)^{m/2}} \\
 &= \frac{1}{2} \left(\sum_{\substack{i \text{ is odd} \\ 0 < i < m}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) - \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) \right) + \frac{(b-a)^m}{2(ab)^{m/2}} \\
 &= \frac{1}{2} \left(\binom{m}{m-1} (1 + 2^{m-\frac{m}{2}+1}) + \sum_{\substack{i \text{ is odd} \\ 0 < i < m-1}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) - \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) \right) \\
 &+ \frac{(b-a)^m}{2(ab)^{m/2}} \\
 &< \frac{1}{2} \left(\binom{m}{m-1} (1 + 2^{m-\frac{m}{2}+1}) + \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) - \sum_{\substack{i \text{ is even} \\ 0 < i < m}} \binom{m}{i} (1 + 2^{\frac{m}{2}-i}) \right) \\
 &+ \frac{(b-a)^m}{2(ab)^{m/2}} \\
 &= \frac{1}{2} m (1 + 2^{-\frac{m}{2}+1}) + \frac{(b-a)^m}{2(ab)^{m/2}} \\
 &= m \left(\frac{1}{2} + 2^{-\frac{m}{2}} \right) + \frac{(b-a)^m}{2(ab)^{m/2}} \tag{20}
 \end{aligned}$$

As $m \geq 2$, then (20) will become

$$\frac{a^m + b^m}{2(ab)^{m/2}} < m + \frac{(b-a)^m}{2(ab)^{m/2}}$$

If $a < b < 2a$ then

$$\begin{aligned}
 0 &< b - a < 2a - a \\
 0 &< (b-a)^m < a^m < a^{m/2} b^{m/2}. \tag{21}
 \end{aligned}$$

As $(b-a)^m < (ab)^{m/2}$ hence $0 < \frac{(b-a)^m}{2(ab)^{m/2}} < 1$. This completes the proof. \square

Theorem 4.1. *If $a < b < 2a$ where a, b are n -bit numbers. Let $\epsilon_1 = \frac{1}{2a^{m/2}} - \alpha$ and $\epsilon_2 = \frac{1}{2b^{m/2}} - \beta$ where $\alpha = \frac{1}{2u}$ and $\beta = \frac{1}{2v}$ for $u > a^m$ and $v > b^m$ and m is a power of 2. Then $1 < \sqrt{(a^m + 1)(b^m + 1)} - (ab)^{m/2} < m + \delta$ where $0 < \delta < 1$.*

Proof. We can see that

$$\begin{aligned}
 \sqrt{(a^m + 1)(b^m + 1)} &= \sqrt{a^m + 1} \sqrt{b^m + 1} \\
 &= (a^{m/2} + \epsilon_1)(b^{m/2} + \epsilon_2) \\
 &= (ab)^{m/2} + a^{m/2} \epsilon_2 + b^{m/2} \epsilon_1 + \epsilon_1 \epsilon_2. \tag{22}
 \end{aligned}$$

based on Theorem 3.1. If $\sqrt{(a^m + 1)(b^m + 1)} - (ab)^{m/2}$ then equation (22) will be

$$a^{m/2}\epsilon_2 + b^{m/2}\epsilon_1 + \epsilon_1\epsilon_2 = a^{m/2} \left(\frac{1}{2b^{m/2}} - \beta \right) + b^{m/2} \left(\frac{1}{2a^{m/2}} - \alpha \right) + \left(\frac{1}{2a^{m/2}} - \alpha \right) \left(\frac{1}{2b^{m/2}} - \beta \right)$$

where $\epsilon_1 = \frac{1}{2a^{m/2}} - \alpha$ and $\epsilon_2 = \frac{1}{2b^{m/2}} - \beta$. We can see that

$$\begin{aligned} \left(\frac{1}{2a^{m/2}} - \alpha \right) \left(\frac{1}{2b^{m/2}} - \beta \right) &= \left(\frac{1}{2a^{m/2}} - \frac{1}{2u} \right) \left(\frac{1}{2b^{m/2}} - \frac{1}{2v} \right) \\ &= \frac{1}{4} \left(\frac{1}{(ab)^{m/2}} - \frac{1}{ub^{m/2}} - \frac{1}{va^{m/2}} + \frac{1}{uv} \right) \\ &= \frac{1}{4} \left(\frac{uv - a^{m/2}v - b^{m/2}u + (ab)^{m/2}}{(ab)^{m/2}uv} \right) \end{aligned} \quad (23)$$

is a negligible value because $a, b > 2^n$ and $u, v > 2^{2n}$ thus $|uv - a^{m/2}v - b^{m/2}u + (ab)^{m/2}| < (ab)^{m/2}uv$ and $(4(ab)^{m/2}uv)^{-1} < 2^{-6n}$ where n is positive integer. We need to show that

$$1 < a^{m/2} \left(\frac{1}{2b^{m/2}} - \beta \right) + b^{m/2} \left(\frac{1}{2a^{m/2}} - \alpha \right) < 1 + \delta.$$

to complete the proof. We can see that

$$\begin{aligned} a^{m/2} \left(\frac{1}{2b^{m/2}} - \beta \right) + b^{m/2} \left(\frac{1}{2a^{m/2}} - \alpha \right) &= \frac{a^{m/2}}{2b^{m/2}} - a^{m/2}\beta + \frac{b^{m/2}}{2a^{m/2}} - b^{m/2}\alpha \\ &= \frac{a^m + b^m}{2(ab)^{m/2}} - (a^{m/2}\beta + b^{m/2}\alpha). \end{aligned}$$

where $(a^{m/2}\beta + b^{m/2}\alpha)$ is a negligible value because

$$\begin{aligned} (a^{m/2}\beta + b^{m/2}\alpha) &= \frac{a^{m/2}}{2u} + \frac{b^{m/2}}{2v} \\ &< \frac{2^n}{2^{2n}} + \frac{2^n}{2^{2n}} \\ &< 2^{-n+1} \end{aligned}$$

as $2^{n-1} < a, b < 2^n$ and $u, v > 2^{2n}$ where n is a positive integer. Based on Lemma 4.1,

$$\frac{a^m + b^m}{2(ab)^{m/2}} < m + \delta$$

where $0 < \delta < 1$. As m is a power of 2, $1 < \frac{a^m + b^m}{2(ab)^{m/2}} < m + \delta$. This terminates the proof. \square

Corollary 4.1. *If a and b be two distinct integers such that $a < b < 2a$ then $\left| \sqrt{(a^m + 1)(b^m + 1)} - (ab)^{m/2} \right| \leq m$.*

Proof. Let

$$\sqrt{(a^m + 1)(b^m + 1)} = \left\lfloor \sqrt{(a^m + 1)(b^m + 1)} \right\rfloor + \epsilon \quad (24)$$

From Theorem 4.1, we get

$$1 < \sqrt{(a^m + 1)(b^m + 1)} - (ab)^{m/2} < m + \delta.$$

If $0 < \delta < 1$ then (24) becomes

$$\begin{aligned} 1 &< \left\lfloor \sqrt{(a^m + 1)(b^m + 1)} \right\rfloor + \epsilon - (ab)^{m/2} < m + 1 \\ 0 &< \left\lfloor \sqrt{(a^m + 1)(b^m + 1)} \right\rfloor - (ab)^{m/2} \leq m \end{aligned}$$

where ϵ is a small positive number such that $\epsilon \leq 1$. This follows the result. \square

Theorem 4.2. Let $N = pq$ be a valid RSA modulus where $p = a^m + 1$ and $q = b^m + 1$ for any positive integer a, b such that $a < b < 2a$ and m is a power of 2. Then N can be factored in polynomial time.

Proof. We can see that

$$N = pq = (a^m + 1)(b^m + 1) \quad (25)$$

and

$$\begin{aligned} \phi(N) &= (p - 1)(q - 1) = a^m b^m \\ \phi(N)^{0.5} &= ((p - 1)(q - 1))^{1/2} = a^{m/2} b^{m/2} = (ab)^{m/2}. \end{aligned} \quad (26)$$

From Corollary 4.1 we know that

$$\left\lfloor \sqrt{(a^m + 1)(b^m + 1)} \right\rfloor - (ab)^{m/2} \leq m \quad (27)$$

where a and b are two distinct positive integers such that $a < b < 2a$. Equation (27) can also be written as

$$\left\lfloor \sqrt{(a^m + 1)(b^m + 1)} \right\rfloor - m \leq (ab)^{m/2}$$

or

$$\left\lfloor \sqrt{N} \right\rfloor - m \leq \phi(N)^{0.5}$$

as shown in equations (25) and (26). Then

$$\left(\left\lfloor \sqrt{N} \right\rfloor - m \right)^2 \leq \phi(N).$$

If N has the size of $2n$ -bit, then

$$\begin{aligned} N &< 2^{2n+1} \\ (a^m + 1)(b^m + 1) &< 2^{2n+1} \end{aligned}$$

If $a < b < 2a$ then

$$\begin{aligned} (a^m + 1) &< 2^{\frac{2n+1}{2}} \\ &= 2^{n+\frac{1}{2}}. \end{aligned}$$

If a is the smallest possible positive integer such that $a < b < 2a$ then

$$m < n + \frac{1}{2}$$

for some positive integer n (RSA-2048 has $n = 1024$). That is we can obtain $\phi(N)$ only by knowing N and m where N can be obtained publicly and m is small positive integer. We know that by knowing the value of $\phi(N)$, we can factor N . Hence, N can be factored in polynomial time. \square

Remark 4.1. For RSA-2048, if $m = 4$, the size of a and b is 128-bit in this attack which can be considered small as factoring 256-bit integer can be solved using current technology. However, if the size of N keeps increasing, the size of a and b will also increase. For example, using the same value of m , RSA-8192 will have the size of a and b in 512-bit. Current integer factorization algorithm still cannot factor integer with 1048-bit in size.

Remark 4.2. We only conduct our attack when m is a power of 2. The reason is $(a^m + 1)$ where m is a power of 2 can only be factored as complex numbers. Specifically,

$$a^m + 1 = (a + i)(a - i) \tag{28}$$

where i is a complex number. This means $(a^m + 1)$ will not produce more than one factor over integers which is the properties of a prime number that we need in p and q .

The attack can be organized using this next algorithm.

4.1 The Algorithm

Algorithm 1 : Factoring $N = (a^m + 1)(b^m + 1)$ where m is a power of 2.

Require: N, m

Ensure: p, q

- 1: Set $i = 1$.
 - 2: **while** $i < m$ **do**
 - 3: Calculate $\phi = \left(\left\lfloor \sqrt{N} \right\rfloor - i \right)^2$
 - 4: Calculate $p = \frac{(N-\phi) \pm \sqrt{(N-\phi)^2 - 4(1)(N)}}{(2)(1)}$.
 - 5: **if** p is an integer **then** Calculate $q = N/p$.
 - 6: **else** Set $i = i + 1$.
 - 7: **end if**
 - 8: **end while**
 - 9: **Output** p and q
-

Now we present a numerical example for this attack.

Example 4.1. We use RSA-2048 modulus in this example. Specifically, we are given

$$N = 250054066319159926797518951285221735286596577906960284744779850 \\ 631622518831735645790974955813086052707180793359005227400937930 \\ 425407658277522409375459892853673156176157765518910184774652653 \\ 552782261379953088518714892122925670341127961676618341033639458 \\ 002794253423097349566807799786088194512160928758222945802888639 \\ 247045462623082987151506519069930442208509086921938871994713627 \\ 429867923568320127047658054951872913557118685012555676804774834 \\ 017398569118755482784705813479389053567130070052475126005688056 \\ 414892505355759938150822689310863753686005033594845014761915705 \\ 03664483513744782158346574745071915844106715972129.$$

Then we calculate

$$L = \left(\lfloor \sqrt{N} \rfloor - 1 \right)^2 \\ = 250054066319159926797518951285221735286596577906960284744779850 \\ 631622518831735645790974955813086052707180793359005227400937930 \\ 425407658277522409375459892853673156176157765518910184774652653 \\ 552782261379953088518714892122925670341127961676618341033639458 \\ 002794253423097349566807799786088194512160928758222945799701502 \\ 159076989824402949825946919328656378992172057505562837062807846 \\ 668441880656253024356712251267237227449393584571483019425154690 \\ 693759687623392603692775824146402769024957150592209267891543012 \\ 343655164887319421724584741186707424894173996980728761108850688 \\ 76083583429901223520743304369828146195189508280576.$$

Given N and L , we can calculate

$$p, q = \frac{(N - L + 1) \pm \sqrt{(N - L + 1)^2 - 4(1)(N)}}{(2)(1)}.$$

as $(X - p)(X - q) = X^2 - (N - L + 1)X + N$ as $N - L + 1 = p + q$. This is true because L is equal to ϕ . That is,

$$p = 179085017309460215141171962633191852883468503124836370156331808 \\ 519129460459587664226369683817168273510584462292299574613012719 \\ 901602434173075037491623127181378425038750120917975952288612098 \\ 524105893517619084973880081958410499787449634499177132756359891 \\ 833428994145825320738412865591329095172535573957177292817$$

and

$$q = 139628691487387064726831769922768121243937818508866571481271684 \\ 671448615683016626980340585277412094952984148480210469494253018 \\ 060411898190813112044664782011620508259878333299315993737973712 \\ 890398513606114961870171560665384312628183244683926528655265473 \\ 473072633435074763105145772011941280071234074960030398737.$$

Remark 4.3. From example 4.1, notice that both p and q satisfy the conditions stated in FIPS PUB 186-4.

5 CONCLUSION

We present the new method to factor RSA modulus $N = pq$ where p and q have special structures that may hinder current factorization algorithm to solve it. While having this structure, we show that both primes can satisfy conditions required to generate key pair specified in latest NIST guidelines. The proposed method can run in polynomial time and increasing size of modulus N will not affect the efficiency of the new method.

REFERENCES

- Asbullah, M. A. and Ariffin, M. R. K. (2016). Design of Rabin-like Cryptosystem without Decryption Failure. *Malaysian Journal of Mathematical Sciences*, 10(S):1–18.
- FIPS, P. (2013). 186-4: Federal information processing standards publication. Digital Signature Standard (DSS). *Information Technology Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD*, pages 20899–8900.
- Gordon, J. (1984). Strong primes are easy to find. In *Workshop on the Theory and Application of Cryptographic Techniques*, pages 216–223. Springer.
- Lenstra Jr, H. W. (1987). Factoring integers with elliptic curves. *Annals of mathematics*, pages 649–673.
- Pollard, J. M. (1974). Theorems on factorization and primality testing. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 76, pages 521–528. Cambridge University Press.
- Rivest, R. L., Shamir, A., and Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126.
- Rivest, R. L. and Silverman, R. D. (1997). Are Strong primes needed for RSA? In *The 1997 RSA Laboratories Seminar Series, Seminars Proceedings*.
- Williams, H. C. and Schmid, B. (1979). Some remarks concerning the MIT public-key cryptosystem. *BIT Numerical Mathematics*, 19(4):525–538.

A New Simultaneous Diophantine Attack Upon RSA Moduli

$$N = pq$$

Saidu Isah Abubakar^{*1}, Muhammad Rezal Kamel Ariffin², and Muhammad Asyraf Asbullah³

^{1,2,3}Laboratory of Cryptography, Analysis and Structure, Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia.

²Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Malaysia.

E-mail: siabubakar82@gmail.com

*Corresponding author

ABSTRACT

This paper reports four new cryptanalytic attacks which show that the t instances of RSA moduli $N = pq$ can be simultaneously factored in polynomial time using simultaneous diophantine approximations and lattice basis reduction techniques. In our technique we utilize the relation given by $N - \left[\left(\frac{a^{\frac{i}{2}} + b^{\frac{i}{2}}}{(2ab)^{\frac{i}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N} \right] + 1$ as a good approximations of $\phi(N)$ for unknown positive integers d, d_i, k_i, k , and z_i . We construct four system of equations of the form $e_s d - k_s \phi(N_s) = 1$, $e_s d_s - k \phi(N_s) = 1$, $e_s d - k \phi(N_s) = z_s$ and $e_s d_s - k \phi(N_s) = z_s$ where $s = 1, 2, \dots, t$. In our attacks, we improve the short decryption exponent bounds of some reported attacks.

Keywords: RSA Moduli, Simultaneous, Diophantine, Approximations, Lattice, Basis, Reduction, LLL algorithm

1 INTRODUCTION

The increased application of shared telecommunications channels, particularly wireless and local area networks (LAN's), results to larger connectivity, but also to a much greater opportunity to intercept data and forge messages. The only practical way to maintain privacy and integrity of information is by using public-key cryptography (Yan, 2008).

The RSA public-key cryptosystem is reported to be the most widely used public-key cryptosystem invented in 1978 by Rivest, Shamir and Adleman. The RSA cryptosystem setup involves randomly selecting two large prime numbers p, q whose product $N = pq$ known as the RSA

modulus and a public key tuple (N, e) used in encrypting message where e is randomly generated and a private key tuple (N, d) which is used in decrypting the ciphertext. The two parameter e, d have a relation in the form of $ed \equiv 1 \pmod{\phi(N)}$ where $\phi(N) = (p-1)(q-1)$ is called the Euler totient function of N (Asbullah and Ariffin, 2016c). RSA has applications in many areas which include e-banking, secure telephone, smart cards, digital and communications in different types of networks (Dubey et al., 2014).

The security of RSA cryptosystem relies on the difficulty of factoring the RSA modulus $N = pq$ for prime numbers p and q , also known as integer factorization problem. The security of RSA can also be associated with the difficulty of solving the RSA key equation problem $ed \equiv 1 \pmod{\phi(N)}$ where the parameters $d, \phi(N)$ are unknown and (e, N) are public key pair. It is therefore recommended for RSA user to generate primes p and q in such a way that the problem of factoring $N = pq$ is computationally infeasible for an adversary. Choosing p and q as strong primes has been recommended as a way of maximizing the difficulty of factoring RSA modulus N .

The use of short decryption exponent is to reduce the decryption time or the signature generation time. In 1990, Wiener proved that RSA is insecure if the decryption exponent is $d < \frac{1}{3}N^{\frac{1}{4}}$ using continued fraction to show that d can be found from the convergent of $\frac{e}{N}$ Wiener (1990). Blömer and May reported an improved version of Wiener's attack using generalized key equation of the form $ex - y\phi(N) = z$ for unknown parameters $x < \frac{1}{3}N^{\frac{1}{4}}$ and $|z| < exN^{-\frac{3}{4}}$. Their techniques used combinations of continued fraction method and lattice basis reduction technique. We emphasize that the continued fraction technique is still widely used for current algebraic cryptanalysis, for instance, Asbullah and Ariffin (2016a) and Asbullah and Ariffin (2016b).

Also, Hinek (2007), proved that k RSA moduli N_i can be factored if $d < N^\gamma$ for $\gamma = \frac{k}{2(k+1)} - \varepsilon$ where ε is a small constant determine based on the size of $\max N_i$. Another instances were also presented by Nitaj et al. (2014). A Nitaj et al., 2014, presented two scenarios which showed that k RSA moduli $N_i = p_i q_i$ can be factored simultaneously in polynomial-time. In the first scenario, they proved that if the given equation $e_i x - y_i \phi(N_i)$ is satisfied where $x < N^\delta$, $y_i < N^\delta$, $|z_i| < \frac{p_i - q_i}{3(p_i + q_i)} y_i N^{\frac{1}{4}}$ for $\delta = \frac{k}{2(k+1)}$, $N = \min N_i$ then k RSA moduli N_i can be factored simultaneously and the second scenario showed that the k instances of RSA public key tuple (N_i, e_i) satisfying $e_i d_i - y \phi(N_i) = z_i$ for unknown integers x_i, y, z_i where $x < N^\delta$, $y_i < N^\delta$, $|z_i| < \frac{p_i - q_i}{3(p_i + q_i)} y_i N^{\frac{1}{4}}$ for $\delta = \frac{k(2\alpha-1)}{2(k+1)}$, $N = \min N_i$ and $\min_i e_i = N^\alpha$. They used simultaneous diophantine approximations and lattice basis reduction techniques and finally use the Coppersmith's method to compute prime factors p_i and q_i of RSA moduli N_i in polynomial time.

In Isah et al. (2018), we also presented some results where we established that if the short decryption exponent $d < \sqrt{\frac{a^j + b^j}{2}} \left(\frac{N}{e}\right)^{\frac{1}{2}} N^{0.375}$ then $\frac{k}{d}$ can be found from the convergents of the continued fraction expansion of $\frac{e}{N_1}$, where $N_1 = N - \left[\frac{a^{\frac{j}{2}} + b^{\frac{j}{2}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right] \sqrt{N} + 1$ where a, b, i, j are small positive integers less than $\log N$ which led to the factorization of N in polynomial time. This paper presents four attacks on t instances of RSA public key pair (N_s, e_s) for $s = 1, \dots, t$ satisfying the following equations $e_s d - k_s \phi(N_s) = 1$, $e_s d_s - k \phi(N_s) = 1$, $e_s d - k_s \phi(N_s) = z_s$ and $e_s d_s - k \phi(N_s) = z_s$ where d, d_s, k, k_s , and z_s are unknown positive integers. In the

first attack, we show that t RSA moduli $N = pq$ can be efficiently factored if there exists an integer d and t integers k_s , such that a given equation $e_s d - k_s \phi(N_s) = 1$ is satisfied. We show that the prime factors p_s and q_s of t moduli N_s for $s = 1, \dots, t$ can be found efficiently if $N = \max\{N_s\}$ and

$$d < N^\gamma, k_s < N^\gamma, \text{ for all } \gamma = \frac{t(1+\beta)}{3t+1}$$

for $\beta < \gamma \leq \frac{1}{2}$. In the second attack, we also show that the t instances of RSA moduli can be simultaneously factored if the equation $e_s d_s - k \phi(N_s) = 1$ is satisfied for integers $d_s < N^\gamma, k < N^\gamma, \text{ for } \gamma = \frac{t(\alpha+\beta)}{3t+1}, N = \max\{N_s\}$ and $e_s = \min e_s$. In the third attack, we also show that a generalized key equation $e_s d - k_s \phi(N_s) = z_s$ can be factored using simultaneous diophantine approximations and lattice basis reduction methods if $d < N^\gamma, k_s < N^\gamma, z_s < N^\gamma$ for all $\gamma = \frac{t(1+\beta)}{3t+1}$ and $N = \max N_s$. In the final attack, the paper presents an attack on t RSA moduli $N = pq$ satisfying an equation $e_s d_s - k \phi(N_s) = z_s$ in which we show that the attack can simultaneously factor t RSA moduli if $d_s < N^\gamma, k < N^\gamma, z_s < N^\gamma$ for all $\gamma = \frac{t(\alpha+\beta)}{3t+1}$ where $e_s = \min\{e_s\} = N^\alpha$ and $N = \max\{N_s\}$.

The rest of the paper is organized as follows. In Section 2, we present review of some preliminary results, some previous theorems on t instances of RSA public key pair (N, e) which simultaneously factored t RSA moduli $N = pq$ using simultaneous diophantine approximations and lattice basis reduction techniques. In Section 3, we present the proofs of our main results with lemmas and theorems and their respective numerical examples and finally in Section 4, we conclude the paper.

2 PRELIMINARIES

In this section, we state some definitions and theorems related to t instances of RSA public key pair (N, e) that simultaneously factored an RSA moduli $N = pq$ using simultaneous diophantine approximations and lattice basis reduction techniques.

Definition 2.1. Let $\vec{b}_1, \dots, \vec{b}_m \in \mathcal{R}^n$. The vectors \vec{b}_i 's are said to be linearly dependent if there exist $x_1, \dots, x_m \in \mathcal{R}$, which are not all zero such that

$$\sum_i^m (x_i \vec{b}_i = \mathbf{0}).$$

Otherwise, they are said to be linearly independent.

Definition 2.2 (Lenstra et al., 1982). Let n be a positive integer. A subset \mathcal{L} of an n -dimensional real vector space \mathcal{R}^n is called a lattice if there exists a basis b_1, \dots, b_n on \mathcal{R}^n such that $\mathcal{L} = \sum_{i=1}^n \mathcal{Z} b_i = \sum_{i=1}^n r_i b_i : r_i \in \mathcal{Z}, 1 \leq i \leq n$.

In this situation, we say that b_1, \dots, b_n are basis for \mathcal{L} or that they span \mathcal{L} .

Definition 2.3 (LLL Reduction, Nitaj (2012)). Let $\mathcal{B} = \langle b_1, \dots, b_n \rangle$ be a basis for a lattice \mathcal{L} and let $\mathcal{B}^* = \langle b_1^*, \dots, b_n^* \rangle$ be the associated Gram-Schmidt orthogonal basis. Let

$$\mu_{i,j} = \frac{\langle b_i, b_j^* \rangle}{\langle b_j^*, b_j^* \rangle} \text{ for } 1 \leq j < i$$

The basis \mathcal{B} is said to be LLL reduce if it satisfies the following two conditions:

1. $\mu_{i,j} \leq \frac{1}{2}$, for $1 \leq j < i \leq n$
2. $\frac{3}{4} \|b_{i-1}^*\|^2 \leq \|b_i^* + \mu_{i,i-1} b_{i-1}^*\|^2$ for $1 \leq i \leq n$. Equivalently, it can be written as

$$\|b_i^*\|^2 \geq \left(\frac{3}{4} - \mu_{i,i-1}^2\right) \|b_{i-1}^*\|^2$$

Theorem 2.1 (Blömer and May (2004)). Let (N, e) be RSA public key pair with modulus $N = pq$ and the prime difference $p - q \geq cN^{\frac{1}{2}}$. Suppose that the public exponent $e \in \mathcal{Z}_{\phi(N)}$ satisfies an equation $ex + y = k\phi(N)$ with

$$0 < x < \frac{1}{3}N^{\frac{1}{4}} \quad \text{and} \quad |y| \leq N^{\frac{-3}{4}}ex$$

for $c \leq 1$. Then N can be factored in polynomial time.

Theorem 2.2 (Lenstra et al. (1982)). Let \mathcal{L} be a lattice basis of dimension n having a basis v_1, \dots, v_n . The LLL algorithm produces a reduced basis b_1, \dots, b_n satisfying the following condition

$$\|b_1\| \leq \|b_2\| \leq \dots \leq \|b_j\| \leq 2^{\frac{n(n-1)}{4(n+1-j)}} \det(\mathcal{L})^{\frac{1}{n+1-j}}$$

for all $1 \leq j \leq n$.

Theorem 2.3 (Simultaneous Diophantine Approximations, Nitaj et al. (2014)). Given any rational numbers of the form $\alpha_1, \dots, \alpha_n$ and $0 < \varepsilon < 1$, there is a polynomial time algorithm to compute integers p_1, \dots, p_n and a positive integer q such that

$$\max_i |q\alpha_i - p_i| < \varepsilon \quad \text{and} \quad q \leq 2^{\frac{n(n-3)}{4}} \cdot 3^n \cdot \varepsilon^{-n}.$$

Theorem 2.4 (Nitaj et al. (2014)). Given $k \geq 2$ and let $N_i = p_i q_i$ be k RSA moduli. Let $N = \min_i N_i$. Let $e_i, i = 1, \dots, k$, be k public exponents. Define $\delta = \frac{k}{2(k+1)}$. If there exist an integer $x < N^\delta$ and k integers $y_i < N^\delta$ and $|z_i| < \frac{p_i - q_i}{3(p_i + q_i)} y_i N^{1/4}$ such that $e_i x - y_i \phi(N_i) = z_i$ for $i = 1, \dots, k$, then one can factor the k RSA moduli N_1, \dots, N_k in polynomial time.

Theorem 2.5 (Nitaj et al. (2014)). Given $k \geq 2$ and let $N_i = p_i q_i$ be k RSA moduli with the same size N . Let $e_i, i = 1, \dots, k$, be k public exponents with $\min_i e_i = N^\alpha$. Let $\delta = \frac{(2\alpha-1)k}{2(k+1)}$. If there exist an integer $y < N^\delta$ and k integers $x_i < N^\delta$ and $|z_i| < \frac{p_i - q_i}{3(p_i + q_i)} y N^{1/4}$ such that $e_i x_i - y \phi(N_i) = z_i$ for $i = 1, \dots, k$, then one can factor the k RSA moduli N_1, \dots, N_k in polynomial time.

3 RESULTS

In this section, we present some theorems and their proofs with numerical examples to show how the attacks are carried out to simultaneously factor t RSA moduli $N = pq$.

Lemma 3.1. If a and b are positive integers less than $\log N$ and p and q are prime numbers such that $a > b$ and $ap^j - bq^j \neq 0$ and $N = pq$, then $\phi(N) < N - \lceil \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(ab)^{\frac{1}{2j}}} \sqrt{N} \rceil + 1$.

Proof. Let $(ap^j - bq^j)(bp^j - aq^j) > 0$, then we get

$$\begin{aligned} abp^{2j} - a^2p^jq^j - b^2p^jq^j + abq^{2j} &> 0 \\ ab(p^{2j} + q^{2j}) &> (a^2 + b^2)p^jq^j \end{aligned}$$

Adding $2abp^jq^j$ to both sides we have:

$$\begin{aligned} ab(p^{2j} + 2p^jq^j + q^{2j}) &> (a^2 + 2ab + b^2)p^jq^j \\ (p^j + q^j)^2 &> \frac{(a+b)^2p^jq^j}{ab} \\ p^j + q^j &> \frac{(a+b)(p^jq^j)^{\frac{1}{2}}}{\sqrt{ab}} \end{aligned}$$

Since $(p+q)^j > p^j + q^j$, then

$$p+q > \frac{(a+b)^{\frac{1}{j}}}{(ab)^{\frac{1}{2j}}} \sqrt{N}$$

Finally, $\phi(N) < N - \lceil \frac{(a+b)^{\frac{1}{j}}}{(ab)^{\frac{1}{2j}}} \sqrt{N} \rceil + 1$ □

Lemma 3.2. If a and b are small positive integers and p and q are prime numbers such that $\frac{b^j}{a^j} < \frac{q^i}{p^i}$ for $a > b$ and $a^jp^i - b^jq^i \neq 0$ and $N = pq$, $e < \phi(N)$ then $\phi(N) < N - \lceil \frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(a^jb^j)^{\frac{1}{2i}}} \sqrt{N} \rceil + 1$, for $2 < j < i$.

Proof. Let $(a^jp^i - b^jq^i)(b^jp^i - a^jq^i) > 0$, then we get

$$\begin{aligned} a^jb^jp^{2i} - a^{2j}p^iq^i - b^{2j}p^iq^i + a^jb^jq^{2i} &> 0 \\ a^jb^j(p^{2i} + q^{2i}) &> (a^{2j} + b^{2j})p^iq^i \end{aligned}$$

Adding $2a^jb^jp^iq^i$ to both sides we have

$$\begin{aligned} a^jb^j(p^i + q^i)^2 &> (a^j + b^j)^2p^iq^i \\ (p^i + q^i)^2 &> \frac{(a^j + b^j)^2}{a^jb^j} N^i \\ p^i + q^i &> \frac{(a^j + b^j)}{(ab)^{\frac{j}{2}}} N^{\frac{i}{2}} \end{aligned}$$

Since $(p+q)^i > p^i + q^i$, then

$$p+q > \frac{(a+b)^{\frac{j}{i}}}{(ab)^{\frac{j}{2i}}} \sqrt{N}$$

Finally, $\phi(N) < N - \lceil \frac{(a+b)^{\frac{j}{i}}}{(a^jb^j)^{\frac{1}{2i}}} \sqrt{N} \rceil + 1$ □

Theorem 3.1. Let p and q be prime numbers and $N = pq$ be an RSA modulus such that (N, e) are public keys with $e < \phi(N)$. If $d < \sqrt{\frac{a^j+b^j}{2}} \left(\frac{N}{e}\right)^{\frac{1}{2}} N^{0.375}$ and $N_1 > N - \left\lceil \left[\frac{a^{\frac{j}{2}}+b^{\frac{j}{2}}}{(2ab)^{\frac{j}{2i}}} \sqrt{N} + \frac{a^{\frac{1}{j}}+b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \sqrt{N} \right] + 1 \right\rceil$, for $2 < j < i$ then one of the convergents $\frac{k}{d}$ can be found from the continued fraction expansion of $\frac{e}{N_1}$ which leads to the factorization of RSA modulus N in polynomial time.

Proof. See Isah et al. (2018) □

Simultaneous Attack Using $N - \left\lceil \left[\left(\frac{a^{\frac{j}{2}}+b^{\frac{j}{2}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}}+b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N} \right] + 1 \right\rceil$ as an Approximation of $\phi(N)$

In this section, we present four attacks on t RSA moduli $N = pq$ using system of equations of the form $e_s d - k_s \phi(N_s) = 1$, $e_s d_s - k \phi(N_s) = 1$, $e_s d - k_s \phi(N_s) = z_1$ and $e_s d_s - k \phi(N_s) = z_1$ for $s = 1, \dots, t$, for $3 \leq j < i$ in which we successfully factor t RSA moduli $N = pq$ in polynomial time.

3.1 The Attack on t RSA Moduli $N = pq$ Satisfying $e_s d - k_s \phi(N_s) = 1$

Taking $t \geq 2$, let $N_s = p_s q_s$, for $s = 1, \dots, t$. The attack works for t instances of public key pair (N_s, e_s) if there exists an integer d and t integers k_s , such that equation $e_s d - k_s \phi(N_s) = 1$ holds. We show that prime factors p_s and q_s of the t RSA moduli N_s for $s = 1, \dots, t$, can be found efficiently if $N = \max\{N_s\}$ and $d < N^\gamma$, $k_s < N^\gamma$ for $\gamma = \frac{t(1+\beta)}{3t+1}$ and $\beta < \gamma \leq \frac{1}{2}$.

Theorem 3.2. Let $N_s = p_s q_s$ be RSA moduli for $s = 1, \dots, t$ and $t \geq 2$ and let the pairs (e_s, N_s) be public keys and (d, N_s) be private keys with $e_s < \phi(N_s)$ and a relation $e_s d \equiv 1 \pmod{\phi(N_s)}$ is satisfied. Let $N = \max\{N_s\}$ if there exists positive integers $d < N^\gamma$, $k_s < N^\gamma$ for all $\gamma = \frac{t(1+\beta)}{3t+1}$ such that the equation $e_s d - k_s \phi(N_s) = 1$ holds, for $\beta < \gamma \leq \frac{1}{2}$, then the t RSA moduli N_s can be successfully recovered in polynomial time.

Proof. Suppose $N_s = p_s q_s$ be t RSA moduli for $s = 1, \dots, t$. Suppose that $N = \max\{N_s\}$ and $k_s < N^\gamma$. Then the equation $e_s d - k_s \phi(N_s) = 1$ can be rewritten as follows

$$e_s d - k_s (N_s - (p_s + q_s) + 1) = 1$$

$$e_s d - k_s (N_s - (N_s - \phi(N_s) + 1) + 1) = 1$$

Let $\Phi = \left\lceil \left[\left(\frac{a^{\frac{j}{2}}+b^{\frac{j}{2}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}}+b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_s} \right] \right\rceil$ for $3 \leq j < i$,

$$e_s d - k_s (N_s - \Phi + \Phi - (N_s - \phi(N_s) + 1) + 1) = 1$$

$$\left| \frac{e_s}{N_s - \Phi + 1} d - k_s \right| = \frac{|1 - k_s (\Phi - N_s + \phi(N_s) - 1)|}{N_s - \Phi + 1} \quad (1)$$

Setting $N = \max\{N_s\}$, $k_s < N^\gamma$, $d < N^\gamma$ be positive integers and suppose that

$$\begin{aligned} \frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(ab)^{\frac{j}{2i}}} \sqrt{N_s} + \phi(N_s) - N_s - 1 &< N^{2\gamma-\beta} \\ N_s - \frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(ab)^{\frac{j}{2i}}} \sqrt{N_s} + 1 &> \frac{a}{b^2} N \end{aligned}$$

Plugging the conditions into inequality (1) gives the following

$$\begin{aligned} \frac{|1 - k_s (\Phi - N_s + \phi(N_s) - 1)|}{N_s - \Phi + 1} &< \frac{\left| 1 + k_s \left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(ab)^{\frac{j}{2i}}} \sqrt{N_s} - N_s + \phi(N_s) - 1 \right) \right|}{N_s - \frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(ab)^{\frac{j}{2i}}} \sqrt{N_s} + 1} \\ &< \frac{1 + N^\gamma (N^{2\gamma-\beta})}{\frac{a}{b^2} N} \\ &= \frac{b^2 (1 + N^{3\gamma-\beta})}{aN} \\ &< \left(\frac{a}{b} \right)^{\frac{j}{i}} N^{3\gamma-\beta-1} \end{aligned}$$

Then, it follows that

$$\left| \frac{e_s}{N_s - \Phi + 1} d - k_s \right| < \left(\frac{a}{b} \right)^{\frac{j}{i}} N^{3\gamma-\beta-1}$$

We next proceed to show the existence of integer d and t integers k_s . We let $\varepsilon = \left(\frac{a}{b} \right)^{\frac{j}{i}} N^{3\gamma-\beta-1}$, with $\gamma = \frac{t(1+\beta)}{3t+1}$. Then we have

$$N^\gamma \varepsilon^t = N^\gamma \left(\left(\frac{a}{b} \right)^{\frac{j}{i}} N^{3\gamma-\beta-1} \right)^t = \left(\frac{a}{b} \right)^{\frac{jt}{i}} N^{\gamma+3\gamma t-\beta t-t} = \left(\frac{a}{b} \right)^{\frac{jt}{i}}$$

Since $\left(\frac{a^j}{b^i} \right)^{\frac{jt}{i}} < 2^{\frac{t(t-3)}{4}} \cdot 3^t$ for $t \geq 2$, then we get $N^\gamma \varepsilon^t < 2^{\frac{t(t-3)}{4}} \cdot 3^t$. It follows that if $d < N^\gamma$ then $d < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$, we have

$$\left| \frac{e_s}{N_s - \Phi + 1} d - k_s \right| < \varepsilon, \quad d < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$$

This satisfies the conditions of Theorem 2.3, and we proceed to find integer d and t integers k_s for $s = 1, \dots, t$. Next from the equation $e_s d - k_s \phi(N_s) = 1$ we compute the following:

$$\phi(N_s) = \frac{e_s d - 1}{k_s}$$

$$p_s + q_s = N_s - \phi(N_s) + 1$$

$$x^2 - (N_s - \phi(N_s) + 1)x + N_s = 0$$

Finally the prime factors p_s and q_s can be revealed which lead to the factorization of t RSA moduli N_s for $s = 1, \dots, t$. \square

Example 3.1. As an illustration to show how our attack of Theorem 3.2 works on t RSA Moduli, we consider the following three RSA moduli and their corresponding public exponents:

$$\begin{aligned} \text{Let } N_1 &= 315078539564032793014025183523942518374399 \\ N_2 &= 248643343706880670445537684832337617453461 \\ N_3 &= 313563765085645786587032668649968533858157 \\ e_1 &= 54785773022691739080967465778411770222115 \\ e_2 &= 233202275925527309532487719612540335362787 \\ e_3 &= 127773668166415955590004752976664794561699 \end{aligned}$$

Observe $N = \max\{N_1, N_2, N_3\}$

$$N = 315078539564032793014025183523942518374399$$

Taking $t = 3$, we have $\gamma = \frac{t(1+\beta)}{3t+1} = 0.360$ and

$\varepsilon = \left(\frac{a}{b}\right)^{\frac{j}{i}} N^{3\gamma-\beta-1} = 0.00001419901394$. Then, applying Theorem 2.3 for $n = t = 3$ we compute

$$C = [3^{t+1} \cdot 2^{\frac{(t+1)(t-4)}{4}} \cdot \varepsilon^{-t-1}] = 99637375850000000000$$

Taking

$$\begin{aligned} \Phi_1 &= N_1 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_1} \right] + 1 \\ \Phi_2 &= N_2 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_2} \right] + 1 \\ \Phi_3 &= N_3 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_3} \right] + 1 \end{aligned}$$

Consider the lattice \mathcal{L} spanned by the matrix

$$M = \begin{bmatrix} 1 & -[C \frac{e_1}{\Phi_1}] & -[C \frac{e_2}{\Phi_2}] & -[C \frac{e_3}{\Phi_3}] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}$$

Therefore, by applying the LLL algorithm to \mathcal{L} , we obtain the reduced basis with following matrix

$$K = \begin{bmatrix} -587415796688459 & -81248903596137 & -1084033091323991 & -545117608215071 \\ -2313427250843772 & -1987081530181396 & 4232019267889172 & -6511230567971468 \\ -7326215049658198 & -4449405116695314 & 3216416747729298 & 1289829390273538 \\ -2413352066722934 & 10101990792311438 & 10101990792311438 & -3155564744178846 \end{bmatrix}$$

Next we compute $J = K \cdot M^{-1}$

$$J = \begin{bmatrix} -587415796688459 & -102139703173206 & -550936528845291 & -239365256573193 \\ -2313427250843772 & -402258117752507 & -2169760477163234 & -942695294514397 \\ -7326215049658198 & -1273880332762023 & -6871247780170771 & -2985349312970351 \\ -2413352066722934 & -419633018276246 & -2263479834927063 & -983413520557661 \end{bmatrix}$$

From the first row of the matrix J we obtain d, k_1, k_2 and k_3 as follows:

$$d = 587415796688459, k_1 = 102139703173206, k_2 = 550936528845291, \\ k_3 = 239365256573193$$

We now compute $\phi(N_s) = \frac{e_s d - 1}{k_s}$ for $s = 1, 2, 3$. That is:

$$\phi(N_1) = 315078539564032793012647397356576267277664 \\ \phi(N_2) = 248643343706880670444208409556385023201152 \\ \phi(N_3) = 313563765085645786585874804790352134354880$$

Also, we proceed to compute $N_s - \phi(N_s) + 1$ for $s = 1, 2, 3$.

$$N_1 - \phi(N_1) + 1 = 1377786167366251096736 \\ N_2 - \phi(N_2) + 1 = 1329275275952594252310 \\ N_3 - \phi(N_3) + 1 = 1157863859616399503278$$

Finally solving the quadratic equation $x^2 - (N_s - \phi(N_s) + 1)x + N_s$ for $s = 1, 2, 3$ gives us p_1, p_2, p_3 and q_1, q_2, q_3 which lead to the factorization of t RSA moduli N_1, N_2, N_3 . That is:

$$p_1 = 1088261511556953635023, p_2 = 1104068913726932255413 \\ p_3 = 725895919498873764497, q_1 = 289524655809297461713 \\ q_2 = 225206362225661996897, q_3 = 431967940117525738781$$

We observe from our result, we get $d = N^{0.35589}$ which is larger than the Blömer-May's bound of $x < \frac{1}{3}N^{0.25}$. This shows that the Blömer-May's attack will not yield the factorization of the t RSA moduli in our case (see Blömer and May (2004)). Furthermore, our bound $d = N^{0.35589}$ is also greater than Nitaj bound of $x = N^{0.344}$ (see Nitaj et al. (2014)).

3.2 The Attack on t RSA Moduli $N = pq$ Satisfying $e_s d_s - k\phi(N_s) = 1$

In this section, we consider second case in which the t RSA moduli satisfy the t equation of the form $e_s d_s - k\phi(N_s) = 1$ for unknown positive integers d_s, k , for $s = 1, \dots, t$.

Theorem 3.3. Let $N_s = p_s q_s$ be RSA moduli for $s = 1, \dots, t$ and $t \geq 2$ and let the pairs (e_s, N_s) be public keys and (d_s, N_s) be private keys with $e_s < \phi(N_s)$ and the relation $e_s d \equiv 1 \pmod{\phi(N_s)}$ is satisfied. Let $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$ be t public exponents for $s = 1, \dots, t$, if there exist integers $d_s < N^\gamma$, $k < N^\gamma$ for all $\gamma = \frac{t(\alpha+\beta)}{3t+1}$ such that the equation $e_s d_s - k\phi(N_s) = 1$ holds, then the prime factors p_s and q_s of t RSA moduli N_s can be successfully recovered in polynomial time for all $s = 1, \dots, t$.

Proof. Let $N_s = p_s q_s$, be t RSA moduli and suppose $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$ be t public exponents for $s = 1, \dots, t$, and suppose that $d_s < N^\gamma$. Then the equation $e_s d_s - k\phi(N_s) = 1$ can be rewritten as

$$e_s d_s - k(N_s - (p_s + q_s) + 1) = 1$$

$$e_s d_s - k(N_s - (N_s - \phi(N_s) + 1)) = 1.$$

$$\text{Let } \Delta = \left\lceil \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_s} \right] \text{ for } 3 \leq j \leq i, \right.$$

$$\left. e_s d_s - k(N_s - \Delta + \Delta - (N_s - \phi(N_s) + 1) + 1) = 1 \right.$$

Then we can have:

$$\left| k \frac{(N_s - \Delta + 1)}{e_s} - d_s \right| = \frac{|1 - k(\Delta - N_s + \phi(N_s) - 1)|}{e_s}$$

Taking $N = \max\{N_s\}$ and suppose that $d_s < N^\gamma$, $k < N^\gamma$ be positive integers and

$$\Delta + \phi(N_s) - N_s - 1 < N^{2\gamma - \beta}.$$

Suppose also $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$, for $s = 1, \dots, t$ then we have

$$\begin{aligned} \frac{|1 - k(\Delta - N_s + \phi(N_s) - 1)|}{e_s} &\leq \frac{|1 + k(\Delta - N_s + \phi(N_s) - 1)|}{e_s} \\ &< \frac{1 + N^\gamma(N^{2\gamma - \beta})}{N^\alpha} \\ &= \frac{1 + N^{3\gamma - \beta}}{N^\alpha} \\ &< \left(\frac{a}{b}\right)^{\frac{j}{2i}} N^{3\gamma - \alpha - \beta} \end{aligned}$$

Hence we get:

$$\left| k \frac{(N_s - \Delta + 1)}{e_s} - d_s \right| < \left(\frac{a}{b}\right)^{\frac{j}{2i}} N^{3\gamma - \alpha - \beta}$$

We now proceed to show the existence of integer k and t integers d_s . Taking $\varepsilon = \left(\frac{a}{b}\right)^{\frac{j}{2i}} N^{3\gamma - \alpha - \beta}$ and $\gamma = \frac{t(\alpha+\beta)}{3t+1}$. Then we get

$$N^\gamma \varepsilon^t = N^\gamma \left(\left(\frac{a}{b}\right)^{\frac{j}{2i}} N^{3\gamma - \alpha - \beta} \right)^t = \left(\frac{a}{b}\right)^{\frac{jt}{2i}} N^{\gamma + 3\gamma t - \alpha t - \beta t} = \left(\frac{a}{b}\right)^{\frac{tj}{2i}}$$

Since $\left(\frac{a}{b}\right)^{\frac{tj}{2i}} < 2^{\frac{t(t-3)}{4}} \cdot 3^t$ for $t \geq 2$, then $N^\gamma \varepsilon^t < 2^{\frac{t(t-3)}{4}} \cdot 3^t$. It follows that if $k < N^\gamma$ then $k < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$ for $s = 1, \dots, t$, we have

$$\left| k \frac{(N_s - \Delta + 1)}{e_s} - d_s \right| < \varepsilon, \quad k < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}.$$

This also satisfies the conditions of Theorem 2.3 and we now proceed to reveal the private key d_s and k for $s = 1, \dots, t$. Next from the equation $e_s d_s - k \phi(N_s) = 1$ we can compute the following:

$$\begin{aligned} \phi(N_s) &= \frac{e_s d_s - 1}{k} \\ p_s + q_s &= N_s - \phi(N_s) + 1 \\ x^2 - (N_s - \phi(N_s) + 1)x + N_s &= 0 \end{aligned}$$

Finally the prime factors p_s and q_s can be found which lead to the factorization of t RSA moduli N_s for $s = 1, \dots, t$ □

Example 3.2. *In what follows, we give a numerical example to illustrate how our attack of Theorem 3.3 works on t RSA Moduli. We consider the following three RSA moduli and their corresponding public exponents*

$$\begin{aligned} N_1 &= 330887927826729358131406751905555113358427 \\ N_2 &= 909455241479718015703976451522306293699987 \\ N_3 &= 896255999831476423504365353752613393410129 \\ e_1 &= 260093505791357595269019761161559922357089 \\ e_2 &= 830211428275988442317142948578507842037903 \\ e_3 &= 260639236216424239075202140155225066663301 \end{aligned}$$

Observe

$$N = \max\{N_1, N_2, N_3\} = 909455241479718015703976451522306293699987$$

and

$$e_s = \min\{e_1, e_2, e_3\} = 260093505791357595269019761161559922357089$$

with $e_s = \min\{e_1, e_2, e_3\} = N^\alpha$ with $\alpha = 0.9870431932$. Taking $t = 3$ and $\beta = 0.25$ we have $\gamma = \frac{t(\alpha+\beta)}{3t+1} = 0.3711129579$ and $\varepsilon = 0.000007508475067$.

Applying Theorem 3.3, we compute

$$C = [3^{t+1} \cdot 2^{\frac{(t+1)(t-4)}{4}} \cdot \varepsilon^{-t-1}] = 12742306620000000000000$$

Taking

$$\begin{aligned} \Delta_1 &= N_1 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_1} \right] + 1 \\ \Delta_2 &= N_2 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_2} \right] + 1 \\ \Delta_3 &= N_3 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_3} \right] + 1 \end{aligned}$$

Consider the lattice \mathcal{L} spanned by the matrix

$$M = \begin{bmatrix} 1 & -[C \frac{\Delta_1}{e_1}] & -[C \frac{\Delta_2}{e_2}] & -[C \frac{\Delta_3}{e_3}] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}$$

where $a = 3, b = 2, i = 4, j = 3, t = 3$.

Therefore by applying the LLL algorithm to \mathcal{L} , we obtain the reduced basis with following matrix

$$K = \begin{bmatrix} -175146409612035 & -823228839795 & 174148519192170 & -114206584622820 \\ -84039951771888287 & 80666065160018481 & -87963455766549006 & -5966375445846324 \\ 76917823720099937 & 113434318528267569 & 264927030686706 & -118170963300717876 \\ 21604480682726699 & 152229348988955163 & 151359706383740262 & 196696397901374148 \end{bmatrix}$$

Next we compute $J = K \cdot M^{-1}$

$$J = \begin{bmatrix} -175146409612035 & -222819221750609 & -191864162336087 & -602273175529801 \\ -84039951771888287 & -106914647529743848 & -92061578568439781 & -288986846702386117 \\ 76917823720099937 & 97853959199204323 & 84259642258505725 & 264496098146466542 \\ 21604480682726699 & 27484968619764657 & 23666631808664282 & 74290984412871231 \end{bmatrix}$$

From the first row of matrix J we obtain k, d_1, d_2 , and d_3 as follows:

$$k = 175146409612035, d_1 = 222819221750609, \\ d_2 = 191864162336087, d_3 = 602273175529801$$

We now compute $\phi(N_i) = \frac{e_s d_s - 1}{k}$ for $i = 1, 2, 3$. That is:

$$\phi(N_1) = 330887927826729358130254895146939245547920 \\ \phi(N_2) = 909455241479718015702034073311041714951816 \\ \phi(N_3) = 896255999831476423502471935613753586474660$$

Also, we proceed to compute $N_s - \phi(N_s) + 1$ for $s = 1, 2, 3$.

$$N_1 - \phi(N_1) + 1 = 1151856758615867810508 \\ N_2 - \phi(N_2) + 1 = 1942378211264578748172 \\ N_3 - \phi(N_3) + 1 = 1893418138859806935470$$

Finally solving the quadratic equation $x^2 - (N_s - \phi(N_s) + 1)x + N_s = 0$ for $i = 1, 2, 3$ gives us p_1, p_2, p_3 and q_1, q_2, q_3 which lead to the factorization of t RSA moduli N_1, N_2, N_3 . That is:

$$p_1 = 604310949056531947721, p_2 = 1154909102962814371933, \\ p_3 = 948145143716756720671, q_1 = 547545809559335862787, \\ q_2 = 787469108301764376239, q_3 = 945272995143050214799$$

From our result, one can observe that we get $\min\{d_1, d_2, d_3\} = N^{0.3404}$ which is larger than the Blömer-May's bound of $x < \frac{1}{3}N^{0.25}$, (see Blömer and May (2004)). This shows that the Blömer-May's attack can not yield the factorization of the t RSA moduli in our case. Furthermore, our $\min\{d_1, d_2, d_3\} = N^{0.3404}$ is also greater than the $\min\{x_1, x_2, x_3\} = N^{0.337}$ of Nitaj et al. (see Nitaj et al. (2014)).

3.3 The Attack on t RSA Moduli $N = pq$ Satisfying $e_s d - k_s \phi(N_s) = z_s$

In this section, we consider another case in which the t RSA moduli satisfy the t equation of the form $e_s d_s - k_s \phi(N_s) = z_s$ for unknown parameters d, k_s, z_s for $s = 1, \dots, t$.

Taking $s \geq 2$, let $N_s = p_s q_s$, $s = 1, \dots, t$. The attack can work for t instances (N_s, e_s) when there exists an integer d and t integers k_s , satisfying the equation $e_s d - k_s \phi(N_s) = z_s$. We show that the prime factors $p_s q_s$ of the RSA t moduli N_s for $s = 1, \dots, t$ can be found efficiently if $N = \max\{N_1, \dots, N_t\}$ and

$$d < N^\gamma, \quad k_s < N^\gamma, \quad z_s < N^\gamma, \quad \text{for all } \gamma = \frac{t(1+\beta)}{3t+1}$$

Theorem 3.4. *Let $N_s = p_s q_s$ be RSA moduli for $s = 1 \dots, t$ and $t \geq 2$ and let pairs (e_s, N_s) be public keys and (d, N_s) be private keys with $e_s < \phi(N_s)$ and the relation $e_s d \equiv 1 \pmod{\phi(N_s)}$ is satisfied. Let $N = \max\{N_s\}$ if there exists positive integers $d < N^\gamma$, $k_s < N^\gamma$, $z_s < N^\gamma$, for all $\gamma = \frac{t(1+\beta)}{3t+1}$ such that the equation $e_s d - k_s \phi(N_s) = z_s$ holds, then the prime factors p_s and q_s of t RSA moduli N_s can be successfully recovered in polynomial time.*

Proof. Let $N_s = p_s q_s$, be t moduli. Also suppose $N = \max\{N_s\}$, and $k_s < N^\gamma$. Then the equation $e_s d - k_s \phi(N_s) = z_s$ can be rewritten as

$$e_s d - k_s (N_s - (p_s + q_s) + 1) = z_s$$

$$e_s d - k_s (N_s - (N_s - \phi(N_s) + 1)) = z_s$$

Let $\Psi = \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_s} \right]$ for $3 \leq j \leq i$, then we have

$$e_s d - k_s (N_s - \Psi + \Psi - (N_s - \phi(N_s) + 1) + 1) = z_s$$

$$\left| \frac{e_s}{N_s - \Psi + 1} d - k_s \right| = \frac{|z_s - k_s (\Psi - N_s + \phi(N_s) - 1)|}{N_s - \Psi + 1} \quad (2)$$

Let $N = \max\{N_s\}$ and $k_s < N^\gamma$, $z_s < N^\gamma$ be positive integers and also suppose

$$\begin{aligned} |\Psi + \phi(N_s) - N_s - 1| &< N^{2\gamma-\beta} \\ N_s - \Psi + 1 &> \frac{a}{b^2} N. \end{aligned} \quad (3)$$

Then plugging into the inequality (2) yields

$$\begin{aligned} \frac{|z_s - k_s (\Psi - N_s + \phi(N_s) - 1)|}{N_s - \Psi + 1} &< \frac{|z_s + k_s (\Psi - N_s + \phi(N_s) - 1)|}{N_s - \Psi + 1} \\ &< \frac{N^\gamma + N^\gamma (N^{2\gamma-\beta})}{\frac{a}{b^2} N} \\ &= \frac{b^2 (N^\gamma + N^{3\gamma-\beta})}{aN} \\ &< \left(\frac{a}{b} \right)^{\frac{i}{j}} N^{3\gamma-\beta-1} \\ \left| \frac{e_s}{N_s - \Psi + 1} d - k_s \right| &< \left(\frac{a}{b} \right)^{\frac{i}{j}} N^{3\gamma-\beta-1} \end{aligned}$$

We now proceed to show the existence of an integer d and t integers k_s . Taking $\varepsilon = \left(\frac{a}{b}\right)^{\frac{i}{j}} N^{3\gamma-\beta-1}$, with $\gamma = \frac{t(1+\beta)}{3t+1}$. Then we have

$$N^\gamma \varepsilon^t = N^\gamma \left(\left(\frac{a}{b}\right)^{\frac{it}{j}} N^{3\gamma-\beta-1} \right)^t = \left(\frac{a}{b}\right)^{\frac{it}{j}} N^{\gamma+3\gamma t-\beta t-t} = \left(\frac{a}{b}\right)^{\frac{it}{j}}$$

Since $\left(\frac{a}{b}\right)^{\frac{it}{j}} < 2^{\frac{t(t-3)}{4}} \cdot 3^t$ for $t \geq 3$, then, we get $N^\gamma \varepsilon^t < 2^{\frac{t(t-3)}{4}} \cdot 3^t$. It follows that if $d < N^\gamma$ then $d < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$ $s = 1, \dots, t$ we have

$$\left| \frac{e_s}{N_s - \Psi + 1} d - k_s \right| < \varepsilon, \quad d < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t} \quad (4)$$

This also satisfies the conditions of Theorem 2.3. We next proceed to reveal the integer d and t integers k_s for $s = 1, \dots, t$. Next from the equation $e_s d - k_s \phi(N_s) = z_s$ we can compute the following:

$$\begin{aligned} \phi(N_s) &= \frac{e_s d - z_s}{k_s} \\ p_s + q_s &= N_s - \phi(N_s) + 1 \\ x^2 - (N_s - \phi(N_s) + 1)x + N_s &= 0 \end{aligned}$$

Finally the prime factors p_s and q_s can be revealed which lead to the factorization of t RSA moduli N_s for $s = 1, \dots, t$. □

Example 3.3. *In what follows, we give a numerical example to illustrate how our attack of Theorem 3.4 works on t RSA Moduli. We consider the following three RSA moduli and their corresponding public exponents:*

$$\begin{aligned} \text{Let } N_1 &= 478202014692581203725714896104461606002087 \\ N_2 &= 407373410958533040905284126341361092602851 \\ N_3 &= 499175968661010756100202689444114150592723 \\ e_1 &= 182924502224696062844581056611029752935382 \\ e_2 &= 242054720451558422339567852547153515942673 \\ e_3 &= 499175968661010756100202689444114150592723 \end{aligned}$$

Observe $N = \max\{N_1, N_2, N_3\}$

$$N = 499175968661010756100202689444114150592723$$

Taking $t = 3, \beta = 0.2$, we have $\gamma = \frac{t(1+\beta)}{3t+1} = 0.36$ and

$$\varepsilon = \left(\frac{a}{b}\right)^{\frac{i}{j}} N^{3\gamma-\beta-1} = 0.00001702150298.$$

Applying Theorem 2.3, for $n = t = 3$ we compute

$$C = [3^{t+1} \cdot 2^{\frac{(t+1)(t-4)}{4}} \cdot \varepsilon^{-t-1}] = 482462321800000000000$$

Taking

$$\begin{aligned}\Psi_1 &= N_1 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_1} \right] + 1 \\ \Psi_2 &= N_2 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_2} \right] + 1 \\ \Psi_3 &= N_3 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_3} \right] + 1\end{aligned}$$

Consider the lattice \mathcal{L} spanned by the matrix

$$M = \begin{bmatrix} 1 & -[C \frac{e_1}{\Psi_1}] & -[C \frac{e_2}{\Psi_2}] & -[C \frac{e_3}{\Psi_3}] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}$$

Therefore, by applying the LLL algorithm to \mathcal{L} , we obtain the reduced basis with following matrix

$$K = \begin{bmatrix} -786769598471111 & 17551771343013 & 375471762735787 & 779076690941344 \\ -1240140677681243 & -2939260572286431 & -990535967249169 & 87848135184672 \\ -1811695573378447 & 1476087950072701 & 210913397314899 & -1945255928194912 \\ -1122173399445060 & 5105352219907980 & -8265461496459980 & 2160982162554240 \end{bmatrix}$$

Next we compute $J = K \cdot M^{-1}$

$$J = \begin{bmatrix} -786769598471111 & -300959495660788 & -467485825276640 & -784104294569916 \\ -1240140677681243 & -474385529929801 & -736871619457008 & -1235939509012960 \\ -1811695573378447 & -693019896948751 & -1076480334162222 & -1805558173955685 \\ -1122173399445060 & -429260028599514 & -666777362473709 & -1118371863207297 \end{bmatrix}$$

From the first row of matrix J we obtain $d, k_1, k_2,$ and k_3 as follows:

$$\begin{aligned}d &= 786769598471111, \quad k_1 = 300959495660788, \\ k_2 &= 467485825276640, \quad k_3 = 784104294569916\end{aligned}$$

We next compute $\phi(N_s) = \frac{e_s d - z_s}{k_s}$ for $s = 1, 2, 3$. That is: where z_1, z_2, z_3 are :

$$z_1 = 721100209071834, \quad z_2 = 363119586169143, \quad z_3 = 47354453320039$$

$$\begin{aligned}\phi(N_1) &= 478202014692581203724146376589903938235936 \\ \phi(N_2) &= 407373410958533040903863937079629053587104 \\ \phi(N_3) &= 499175968661010756098610235201732733930280\end{aligned}$$

Also, we proceed to compute $N_s - \phi(N_s) + 1$ for $s = 1, 2, 3$.

$$\begin{aligned} N_1 - \phi(N_1) + 1 &= 1568519514557667766152 \\ N_2 - \phi(N_2) + 1 &= 1420189261732039015748 \\ N_3 - \phi(N_3) + 1 &= 1592454242381416662444 \end{aligned}$$

Finally solving the quadratic equation $x^2 - (N_s - \phi(N_s) + 1)x + N_s = 0$ for $s = 1, 2, 3$ gives us p_1, p_2, p_3 and q_1, q_2, q_3 which lead to the factorization of t RSA moduli N_1, N_2, N_3 in polynomial time. That is:

$$\begin{aligned} p_1 &= 1154207526828096380209, \quad p_2 = 1021319587812302117329, \\ p_3 &= 1163380576653995747053 \quad q_1 = 414311987729571385943, \\ q_2 &= 398869673919736898419, \quad q_3 = 429073665727420915391 \end{aligned}$$

From our result, one can observe that we get $d = N^{0.3522}$ which is larger than the Blömer-May's bound of $x < \frac{1}{3}N^{0.25}$ (see Blömer and May (2004)). This shows that the Blömer-May's attack will not yield the factorization of t RSA moduli in our case. Furthermore, our bound $d = N^{0.3572}$ is greater than $x = N^{0.344}$ of Nitaj et al. (see Nitaj et al. (2014)).

3.4 The Attack on t RSA Moduli $N = pq$ Satisfying $e_s d_s - k\phi(N_s) = z_s$

In this section, we present another case in which the t RSA moduli satisfying an equation of the form $e_s d_s - k\phi(N_s) = z_s$ for unknown positive integers d_s, k, z_s for $s = 1, \dots, t$ can be simultaneously factored in polynomial time.

Theorem 3.5. *Let $N_s = p_s q_s$ be RSA moduli for $s = 1, \dots, t$ and $t \geq 2$ and let the pairs (e_s, N_s) be public keys and (d_s, N_s) be private keys with $e_s < \phi(N_s)$ and the relation $e_s d_s \equiv z_s \pmod{\phi(N_s)}$ is satisfied. Also, let $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$ be t public exponents for $s = 1, \dots, t$, if there exists positive integers $d_s < N^\gamma$, $k < N^\gamma$, $z_s < N^\gamma$, for all $\gamma = \frac{t(\alpha+\beta)}{3t+1}$ such that the equation $e_s d_s - k\phi(N_s) = z_s$ holds, then the prime factors p_s and q_s of t RSA moduli N_s can be successfully recovered in polynomial time for $s = 1, \dots, t$.*

Proof. Suppose $N_s = p_s q_s$, $1 \leq s \leq t$ be t RSA moduli. Setting $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$ be t public exponents for $s = 1, \dots, t$, and suppose that $d_s < N^\gamma$. Then equation $e_s d_s - k\phi(N_s) = z_s$ can be rewritten as

$$e_s d_s - k(N_s - (p_s + q_s) + 1) = z_s$$

$$e_s d_s - k(N_s - (N_s - \phi(N_s) + 1)) = z_s$$

Suppose $\Upsilon = \left[\left(\frac{a^{\frac{i}{2}} + b^{\frac{i}{2}}}{(2ab)^{\frac{i}{2i}}} + \frac{a^{\frac{j}{2}} + b^{\frac{j}{2}}}{(2ab)^{\frac{j}{2j}}} \right) \sqrt{N_s} \right]$ for $3 \leq j < i$, then we have

$$e_s d_s - k(N_s - \Upsilon + \Upsilon - (N_s - \phi(N_s) + 1) + 1) = z_s$$

$$\left| k \frac{(N_s - \Upsilon + 1)}{e_s} - d_s \right| = \frac{|z_s - k(\Upsilon - N_i + \phi(N_s) - 1)|}{e_s} \quad (5)$$

Suppose $N = \max\{N_s\}$, $d_s < N^\gamma$, $k < N^\gamma$, $z_s < N^\gamma$ are positive integers and

$$\Upsilon + \phi(N_s) - N_s - 1 < N^{2\gamma-\beta}$$

and taking $e_s = \min\{e_1, \dots, e_t\} = N^\alpha$. Plugging the above conditions into inequality (5), then we have:

$$\begin{aligned} \frac{|z_s - k(N_s - \Upsilon - N_s + \phi(N_s) - 1)|}{e_s} &\leq \frac{|z_s + k(N_s - \Upsilon - N_s + \phi(N_s) - 1)|}{e_s} \\ &< \frac{N^\gamma + N^\gamma(N^{2\gamma-\beta})}{N^\alpha} \\ &= \frac{N^\gamma + N^{3\gamma-\beta}}{N^\alpha} \\ &< \left(\frac{a}{b}\right)^{\frac{i}{2j}} N^{3\gamma-\alpha-\beta} \end{aligned}$$

Hence we get:

$$\left| k \frac{(N_s - \Upsilon + 1)}{e_s} - d_s \right| < \left(\frac{a}{b}\right)^{\frac{i}{2j}} N^{3\gamma-\alpha-\beta}.$$

We now proceed to show the existence of integer k and the t integers d_s . Let $\varepsilon = \left(\frac{a}{b}\right)^{\frac{i}{2j}} N^{3\gamma-\alpha-\beta}$ and $\gamma = \frac{t(\alpha+\beta)}{3t+1}$. Then we get

$$N^\gamma \varepsilon^t = N^\gamma \left(\left(\frac{a}{b}\right)^{\frac{i}{2j}} N^{3\gamma-\alpha-\beta} \right)^t = \left(\frac{a}{b}\right)^{\frac{it}{2j}} t N^{3\gamma t - t\alpha - \beta t} = \left(\frac{a}{b}\right)^{\frac{it}{2j}}$$

Since $\left(\frac{a}{b}\right)^{\frac{it}{2j}} < 2^{\frac{t(t-3)}{4}} \cdot 3^t$ for $t \geq 3$, then, it implies that $N^\gamma \varepsilon^t < 2^{\frac{t(t-3)}{4}} \cdot 3^t$. It follows that if $k < N^\gamma$ then $k < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$ for $s = 1, \dots, t$, we have

$$\left| k \frac{(N_s - \Upsilon + 1)}{e_s} - d_s \right| < \varepsilon, \quad k < 2^{\frac{t(t-3)}{4}} \cdot 3^t \cdot \varepsilon^{-t}$$

This fulfilled the conditions of Theorem 2.3. We next proceed to reveal the private key d_s and k for $s = 1, \dots, t$. Next from the equation $e_s d_s - k \phi(N_s) = z_s$ we can compute the following:

$$\phi(N_s) = \frac{e_s d_s - z_s}{k}$$

$$p_s + q_s = N_s - \phi(N_s) + 1$$

$$x^2 - (N_s - \phi(N_s) + 1)x + N_s = 0$$

Finally, the prime factors p_s and q_s can be found which lead to the factorization of t RSA moduli N_s for $s = 1, \dots, t$ in polynomial time. \square

Example 3.4. *In what follows, we give a numerical example to illustrate how our attack of Theorem 3.5 works on t RSA Moduli. We consider the following three RSA moduli and their corresponding public exponents:*

$$\begin{aligned} N_1 &= 329514818397907511194535067519744287 \\ N_2 &= 853577457696022637279536861717261139 \\ N_3 &= 689835688169708146675664504365049467 \\ e_1 &= 689835688169708146675664504365049467 \\ e_2 &= 737687793704945765120221919495997383 \\ e_3 &= 156091109112298242178765923428663298 \end{aligned}$$

Observe

$$N = \max\{N_1, N_2, N_3\} = 853577457696022637279536861717261139$$

and

$$e_s = \min\{e_1, e_2, e_3\} = 853577457696022637279536861717261139$$

with $e_s = \min\{e_1, e_2, e_3\} = N^\alpha$ for $\alpha = 0.9794645353$. Since $t = 3$, we have $\gamma = \frac{t(\alpha+\beta)}{3t+1} = 0.3688, \varepsilon = 0.00006564013470$.

Applying Theorem 2.3, we compute

$$C = [3^{t+1} \cdot 2^{\frac{(t+1)(t-4)}{4}} \cdot \varepsilon^{-t-1}] = 2181610267000000000$$

Taking

$$\begin{aligned} \Upsilon_1 &= N_1 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_1} \right] + 1 \\ \Upsilon_2 &= N_2 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_2} \right] + 1 \\ \Upsilon_3 &= N_3 - \left[\left(\frac{a^{\frac{j}{i}} + b^{\frac{j}{i}}}{(2ab)^{\frac{j}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N_3} \right] + 1 \end{aligned}$$

Consider the lattice \mathcal{L} spanned by the matrix

$$M = \begin{bmatrix} 1 & -[C \frac{\Upsilon_1}{e_1}] & -[C \frac{\Upsilon_2}{e_2}] & -[C \frac{\Upsilon_3}{e_3}] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}$$

where $a = 3, b = 2, i = 4, j = 3, t = 3$

Therefore, by applying the LLL algorithm to \mathcal{L} , we obtain the reduced basis with following matrix

$$K = \begin{bmatrix} -1424579461243 & -20025631349712 & -7755984254469 & -31187830209288 \\ -27436166288609 & 32412192492944 & -13886788876447 & -6055604609944 \\ -35676502412466 & 12602114484256 & 46739915404722 & -22467451342256 \\ 69688967857759 & 43943589280656 & 30572521380897 & -56265908213656 \end{bmatrix}$$

Next we compute $J = K \cdot M^{-1}$

$$J = \begin{bmatrix} -1424579461243 & -2804695406341 & -1648378792750 & -6295847076671 \\ -27436166288609 & -54016003775688 & -31746347532657 & -121252560508265 \\ -35676502412466 & -70239481301555 & -41281228303608 & -157670252541326 \\ 69688967857759 & 137202826055599 & 80637001887655 & 307986389317072 \end{bmatrix}$$

From the first row of J we obtain $k, d_1, d_2,$ and d_3 as follows:

$$k = 1424579461243, d_1 = 2804695406341, \\ d_2 = 1648378792750, d_3 = 121252560508265$$

We now compute $\phi(N_s) = \frac{e_s d_s - z_s}{k}$ for $s = 1, 2, 3$. That is where z_1, z_2, z_3 are :

$$z_1 = 579057474385, z_2 = 1556015073242, z_3 = 38593801470$$

$$\phi(N_1) = 329514818397907510033962670013247816 \\ \phi(N_2) = 853577457696022635407743651209932856 \\ \phi(N_3) = 689835688169708144943019327714137216$$

Also, we proceed to compute $N_s - \phi(N_s) + 1$ for $s = 1, 2, 3$.

$$N_1 - \phi(N_1) + 1 = 1160572397506496472 \\ N_2 - \phi(N_2) + 1 = 1871793210507328284 \\ N_3 - \phi(N_3) + 1 = 1732645176650912252$$

Finally, solving the quadratic equation $x^2 - (N_i - \phi(N_i) + 1)x + N_i = 0$ for $i = 1, 2, 3$ gives us p_1, p_2, p_3 and q_1, q_2, q_3 which lead to the factorization of k RSA moduli N_1, N_2, N_3 . That is:

$$p_1 = 665240622214224083, p_2 = 1085312126633841397, \\ p_3 = 1112653948231598779, q_1 = 495331775292272389, \\ q_2 = 786481083873486887, q_3 = 619991228419313473$$

From our result, one can observe that we get $\min\{d_1, d_2, d_3\} = N^{0.3400}$ which is larger than the Blömer-May's bound of $x < \frac{1}{3}N^{0.25}$, (see Blömer and May (2004)). This shows that the Blömer-May's attack can not yield the factorization of the t RSA moduli in our case. Furthermore, our $\min\{d_1, d_2, d_3\} = N^{0.340}$ is greater than the $\min\{x_1, x_2, x_3\} = N^{0.337}$ of Nitaj et al. (see Nitaj et al. (2014)).

4 CONCLUSION

It has been shown in the presented attacks, this paper reports some improvement of bounds over some former attacks on t instances of RSA moduli $N_s = pq$. It has been proven that t instances of RSA moduli $N = pq$ satisfying equations of the form $e_s d - k_s \phi(N_s) = 1, e_s d_s - k \phi(N_s) = 1,$

$e_s d - k_s \phi(N_s) = z_1$ and $e_s d_s - k \phi(N_s) = z_1$ for $s = 1, \dots, t$ for unknown positive integers d, d_s, k, k_s and z_s , and by using $N - \left[\left(\frac{a^{\frac{1}{i}} + b^{\frac{1}{i}}}{(2ab)^{\frac{1}{2i}}} + \frac{a^{\frac{1}{j}} + b^{\frac{1}{j}}}{(2ab)^{\frac{1}{2j}}} \right) \sqrt{N} \right] + 1$ as a good approximation of $\phi(N)$, one can simultaneously factor in polynomial time the t instances of RSA moduli using simultaneous diophantine approximations and lattice basis reductions methods.

REFERENCES

- Asbullah, M. A. and Ariffin, M. R. K. (2016a). Analysis on the AA_β Cryptosystem. In *The 5th International Cryptology and Information Security Conference 2016 (Cryptology2016)*, pages 41–48.
- Asbullah, M. A. and Ariffin, M. R. K. (2016b). Analysis on the Rabin- p cryptosystem. In *The 4th International Conference on Fundamental and Applied Sciences (ICFAS2016)*, pages 080012–1–080012–8. AIP Conf. Proc. 1787.
- Asbullah, M. A. and Ariffin, M. R. K. (2016c). Design of Rabin-like Cryptosystem without Decryption Failure. *Malaysian Journal of Mathematical Sciences*, 10(S):1–18.
- Blömer, J. and May, A. (2004). A Generalized Wiener Attack on RSA. In *Public Key Cryptography (PKC 2004)*, volume 2947 of *LNCS*, pages 1–13. Springer.
- Dubey, M. K., Ratan, R., Verma, N., and Saxena, P. K. (2014). Cryptanalytic Attacks and Countermeasures on RSA. In *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, pages 805–819. Springer, New Delhi.
- Hinek, J. (2007). *On the Security of Some Variants of RSA*. PhD thesis, University of Waterloo, Ontario, Canada.
- Isah, S. A., Asbullah, M. A., and Ariffin, M. R. K. (2018). A New Improved Bound for Short Decryption Exponent on RSA Modulus $N = pq$ Using Wiener’s Method. In *3rd International Conference on Mathematical Sciences and Statistics (ICMSS2018)*, UPM, Malaysia.
- Lenstra, A., Lenstra, H., and Lovsz, L. (1982). Factoring Polynomials with Rational Coefficients. *Mathematische Annalen.*, 261(1):513–534.
- Nitaj, A. (2012). Diophantine and Lattice Cryptanalysis of RSA Cryptosystem. *Artificial Intelligence Evolutionary Computation and Metaheuristics (AIECM)*, 2(11):139–168.
- Nitaj, A., Ariffin, M., D. Naasr, D., and Bahig, H. (2014). New Attacks on the RSA Cryptosystem. In *Lecture Notes in Computer Science, Progress in Cryptology (AFRICACRYPT’2014)*, pages 178–198, Springer, Berlin, Heidelberg.
- Wiener, M. (1990). Cryptanalysis of Short RSA Secret Exponents. *IEEE Trans. Inform. Theory*, 36(3):553–558.
- Yan, S. Y. Y. (2008). *Cryptanalytic Attacks on RSA*. Springer, 1st edition.

New Vulnerability on System of $N_i = p_i^2 q_i$ Using Good Approximation of $\phi(N)$

Normahirah Nek Abd Rahman^{*1}, Muhammad Rezal Kamel Ariffin^{2,3},
Muhammad Asyraf Asbullah², and Faridah Yunos^{2,3}

¹*Pusat PERMATApintar Negara, Universiti Kebangsaan Malaysia , 43600 UKM Bangi, Selangor, Malaysia.*

²*Al-Kindi Cryptography Research Laboratory, Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*

³*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia , 43400 UPM Serdang, Selangor, Malaysia.*

E-mail: normahirah@ukm.edu.my, rezal@upm.edu.my, ma_asyraf@upm.edu.my, faridahy@upm.edu.my
**Corresponding author*

ABSTRACT

This paper proposes new vulnerability in m moduli of the form $N_i = p_i^2 q_i$ for $m \geq 2$ and $i = 1, \dots, m$. The attack works when m public keys (N_i, e_i) there exist m relations of the form $e_i d - k \phi(N_i) = 1$ or $e_i d_i - k_i \phi(N_i) = 1$ with the parameters d, d_i, k and k_i are suitably small. By using the term $N - 2N^{2/3} - N^{1/3}$ as a good approximation of $\phi(N)$ applying the LLL algorithm enables one to factor m moduli of the form $N_i = p_i^2 q_i$ simultaneously.

Keywords: Factorization, LLL algorithm, Simultaneous diophantine approximations

1 INTRODUCTION

RSA cryptosystem become a well-known public key cryptosystem for transmission of data since it was invented by Rivest et al. (1978). The RSA cryptosystem has played a very great role in the development of modern cryptography. The simple structure of the RSA cryptosystem has also attracted many cryptanalysts.

Basically, by multiplying two large primes are computationally easy to compute but factoring the resulting product is very hard. The mathematical operations in RSA depend on three parameters, the modulus $N = pq$, the public exponent e and the private exponent d , related by

the congruence relation $ed \equiv 1 \pmod{\phi(N)}$ where $\phi(N) = (p-1)(q-1)$. Hence, the difficulty of breaking the RSA cryptosystem is based on three hard mathematical problems which is the integers factorization problem of $N = pq$, the e -th root problem from the congruence relation given by $C \equiv M^e \pmod{N}$ and to solve the Diophantine key equation $ed + 1 = \phi(N)k$. That is, to solve the key equation which contain three variables namely $(d, \phi(N), k)$ (Menezes et al. (1997)).

In 1990, Wiener (1990) proved that if the secret exponent $d < \frac{1}{3}N^{1/4}$ based on the convergent of the continued fraction expansion of $\frac{e}{N}$ making RSA totally insecure. Wiener was able to obtain the integer solution of the Diophantine key equation and eventually factoring N through the continued fraction of $\frac{e}{N}$. Furthering this, Boneh and Durfee (1999) presented an attack on RSA by improved the bound to $d < N^{0.292}$ based on Coppersmiths lattice reduction based method. Later, de Weger (2002) proposed an extension of these attacks to an RSA modulus with small difference between its prime factors. As studied by de Weger, if the difference between p and q is small, then $N - 2\sqrt{N} + 1$ is better approximation to $\phi(N)$ instead of N . Hence, $\frac{k}{d}$ is one of the convergent of the continued fractions expansion of $\frac{e}{N-2\sqrt{N}+1}$.

Later, Maitra and Sarkar (2008) consider the case that p and $2q$ are too close. That means the difference between $2q$ and p is small. They used $N - \frac{3}{2}\sqrt{N} + 1$ as a good approximation to $\phi(N)$ instead of N . Hence, in their studied revealed that $\frac{k}{d}$ is one of the convergent of the continued fractions expansion of $\frac{e}{N-\frac{3}{2}\sqrt{N}+1}$. Over the past few years, we have witnessed steady progress toward cryptanalysis using the continued fraction expansion as a tool. For instance, Asbullah and Ariffin (2016a,b) and Bunder and Tonien (2017).

In general, the use of short secret exponent encounters serious security problem in various instances of RSA. As the work that has been proposed by Howgrave-Graham and Seifert (1999), showed an extension of Wiener's attack that allows the RSA system to be insecure in the presence of two decryption exponents (d_1, d_2) with $d_1, d_2 < N^{\frac{5}{14}}$. In the presence of three decryption exponents, they improved the bound to $N^{\frac{2}{5}}$. As studied by Hinek (2007) showed that it is possible to factor the k modulus N_i using equations $e_i d - k_i \phi(N_i) = 1$ if $d < N^\delta$ with $\delta = \frac{k}{2(k+1)} - \epsilon$ where ϵ is a small constant depending on the size of $\max N_i$. Later on, Nitaj et al. (2014) presented new method to factor all the RSA moduli N_1, \dots, N_k in the scenario that RSA instances satisfy k equations of the shape $e_i x - y_i \phi(N_i) = z_i$ or of the shape $e_i x_i - y \phi(N_i) = z_i$ with suitable parameters x, x_i, y, y_i and $z_i, \phi(N_i) = (p_i - 1)(q_i - 1)$ based on the LLL algorithm which has been introduced by Lenstra et al. (1982) for lattice basis reduction.

As described in May (2004) the modulus of the form $N = p^2q$ is frequently used in cryptography and therefore they represent one of the most important cases. For example Takagi (1998) showed that the decryption process is about three times faster than RSA cryptosystem when using the RSA modulus of the form $N = p^2q$. On the other hand, the HIME(R) cryptosystem (Nishioka et al., 2002) became a standard in Japan because it was able to encapsulate and sends more data securely than the original RSA cryptosystem. Additionally, the AA_β cryptosystem (Asbullah and Ariffin, 2014) incorporating the hardness of factoring integer $N = p^2q$ coupled with the square root problem as its cryptographic primitive which gives advantage for encryption without "expensive" mathematical operation. Recently, by incorporating the modulus $N = p^2q$, a variant of Rabin cryptosystem successfully eliminate the decryption failure which was due to a 4-to-1 mapping scenario.

In 2015, motivated from de Weger's generalization attack and Maitra and Sarkars attack, Asbullah and Ariffin (2015) proposed new attack on RSA-type modulus $N = p^2 q$ using the term $N - (2N^{2/3} - N^{1/3})$ as a good approximation to $\phi(N)$ satisfying the equation $ed - k\phi(N) = 1$. Hence, they showed that $\frac{k}{d}$ is one of the convergent of the continued fractions expansion of $\frac{e}{N - (2N^{2/3} - N^{1/3})}$ and later led to the factorization of $N = p^2 q$ in polynomial time.

Our contribution. In this work, we will look at a variant of the RSA modulus of the form $N = p^2 q$. We introduce new attack to factor m moduli of the form $N = p^2 q$ using the term $N - 2N^{2/3} - N^{1/3}$ as a good approximation of $\phi(N)$ satisfying the equation $ed - k\phi(N) = 1$. The first and second attack are upon m -instances (N_i, e_i) .

The first attack works when there exist an integer d and m integers k_i satisfy $e_i d - k_i \phi(N_i) = 1$. We show that m moduli N_i can be factored in polynomial time satisfying $d < N^\delta, k_i < N^\delta$ where $\delta = \frac{m(1-\beta)}{1+m}, \beta < 2/3$ and $N = \min N_i$. The second attack works when there exist an integer k and m integers d_i satisfy $e_i d_i - k\phi(N_i) = 1$. Similarly, we show that m moduli $N_i = p_i^2 q_i$ can be factored in polynomial time if $d_i < N^\delta, k < N^\delta$ where $\delta = \frac{m(\alpha-\beta)}{m+1}, \beta < 2/3, N = \max N_i$ and $\min e_i = N^\alpha$. For both attacks, we transform the equations into a simultaneous Diophantine problem and apply lattice basis reduction techniques to find parameters (d, k_i) or (k, d_i) in order to compute the prime factor p_i and q_i of each $N_i = p_i^2 q_i$ simultaneously.

The layout of the paper is as follows. In Section 2, we begin with a brief review on lattice basic reduction, simultaneous Diophantine approximation and also some useful results that will be used throughout the paper. In Sections 3 and 4 we present our first and second attack consecutively together with examples. Then, we conclude the paper in Section 5.

2 PRELIMINARIES

2.1 Lattice Basis Reductions

Let u_1, \dots, u_d be d linearly independent vectors of \mathbb{R}^n with $d \leq n$. The set of all integer linear combinations of the vectors u_1, \dots, u_d is called a lattice and is in the form

$$\mathcal{L} = \left\{ \sum_{i=1}^d x_i u_i \mid x_i \in \mathbb{Z} \right\}.$$

The set (u_1, \dots, u_d) is called a basis of \mathcal{L} and d is its dimension. The determinant of \mathcal{L} is defined as $\det(\mathcal{L}) = \sqrt{\det(U^T U)}$ where U is the matrix of the u_i 's in the canonical basis of \mathbb{R}^n . Define $\|v\|$ to be the Euclidean norm of a vector $v \in \mathcal{L}$. A central problem in lattice reduction is to find a short non-zero vector in \mathcal{L} . The LLL algorithm of Lenstra et al. (1982) produces a reduced basis and the following result fixes the sizes of the reduced basis vector (see May (2003)).

Theorem 2.1. *Let \mathcal{L} be a lattice of dimension ω with a basis $\{v_1, \dots, v_\omega\}$. The LLL algorithm produces a reduced basis $\{b_1, \dots, b_\omega\}$ satisfying*

$$\|b_1\| \leq \|b_2\| \leq \dots \leq \|b_i\| \leq 2^{\frac{\omega(\omega-1)}{4(\omega+1-i)}} \det(L)^{\frac{1}{\omega+1-i}},$$

for all $1 \leq i \leq \omega$.

One of the important application that the LLL algorithm provided is a way to solve the simultaneous Diophantine approximations problem. Lenstra, Lenstra and Lovász proposed a way to compute simultaneous Diophantine approximations to rational numbers. They considered a lattice with real entries as shown in the following proposition (Lenstra et al., 1982).

Proposition 2.1. (Lenstra et al., 1982). *There exists a polynomial time algorithm that, given a positive integer n and rational numbers $\alpha_1, \alpha_2, \dots, \alpha_n, \varepsilon$ satisfying $0 < \varepsilon < 1$, finds integers p_1, p_2, \dots, p_n, q for which*

$$|q\alpha_i - p_i| < \varepsilon, \quad 1 \leq q \leq 2^{n(n+1)/4} \varepsilon^{-n} \quad \text{for } 1 \leq i \leq n.$$

The above proposition follows immediately from the following classical theorem of Dirichlet (Cassels (1971), Section V.10).

Theorem 2.1. (Dirichlet Theorem). *Let $\vartheta_1, \dots, \vartheta_n$ be n real numbers and Q a real number such that $0 < Q < 1$. There exist integers s_1, \dots, s_n and a positive integer $r \leq Q^{-n}$ such that*

$$|r\vartheta_i - s_i| < Q \quad \text{for } 1 \leq i \leq n.$$

As studied by Nitaj et al. (2014) revealed a similar result for a lattice with integer entries as shown in the following theorem.

Theorem 2.2. (Simultaneous Diophantine Approximations). (Nitaj et al., 2014). *There is a polynomial time algorithm, for given rational numbers $\alpha_1, \dots, \alpha_n$ and $0 < \varepsilon < 1$, to compute integers p_1, \dots, p_n and a positive integer q such that*

$$\max |q\alpha_i - p_i| < \varepsilon \quad \text{and} \quad q \leq 2^{n(n-3)/4} \cdot 3^n \cdot \varepsilon^{-n}.$$

Next, we introduce some useful lemmas and theorem proposed by Asbullah and Ariffin (2015) that have been used throughout this paper.

Lemma 2.1. (Asbullah and Ariffin, 2015). *Let $N = p^2q$ with $q < p < 2q$. Then*

$$2^{-2/3} N^{1/3} < q < N^{1/3} < p < 2^{1/3} N^{1/3}.$$

Let $N = p^2q$ with $q < p < 2q$. Then $\phi(N) = N - (p^2 + pq - p)$. The following result gives an interval for $N - \phi(N) = p^2 + pq - p$ in terms of N . It shows that if $p \approx q$, then $N - 2N^{2/3} - N^{1/3}$ is a good approximation to $\phi(N)$ while if $p \approx 2q$, then $N - \left((2^{2/3} + 2^{-1/3})N^{2/3} + 2^{1/3}N^{1/3} \right)$ is a good approximation to $\phi(N)$.

Lemma 2.2. (Asbullah and Ariffin, 2015) *Let $N = p^2q$ and $\phi(N) = N - (p^2 + pq - p)$ with $q < p < 2q$. Then*

$$2N^{2/3} - N^{1/3} < N - \phi(N) < (2^{2/3} + 2^{-1/3})N^{2/3} - 2^{1/3}N^{1/3}$$

and

$$|N - (2N^{2/3} - N^{1/3}) - \phi(N)| < 2p^{5/3}|p^{1/3} - q^{1/3}|$$

Theorem 2.3. (Asbullah and Ariffin, 2015). Let $N = p^2 q$ be an RSA modulus with $q < p < 2q$. Let $1 < e < \phi(N) < N - (2N^{2/3} - N^{1/3})$ satisfying $ed - k\phi(N) = 1$ for some unknown integers $\phi(N)$, d and k . Assume $\phi(N) > \frac{2}{3}N$ and $N > 6d$. Let $2p^{5/3}|p^{1/3} - q^{1/3}| < \frac{1}{3}N^\beta$ and $d < N^\delta$. If $\delta < N^{\frac{1-\beta}{2}}$, then $\left| \frac{e}{N - (2N^{2/3} - N^{1/3})} - \frac{k}{d} \right| < \frac{1}{2d^2}$.

3 THE FIRST ATTACK ON m MODULI $N_i = p_i^2 q_i$

In this section, we propose our first attack. We extend the result proposed by Asbullah (2015) as in Theorem 2.3 which is the basis of our analysis. The following theorem proved that we are given m moduli of the form $N_i = p_i^2 q_i$. We consider in this scenario given $e_i d - k_i \phi(N_i) = 1$ with fixed d and the term $N - (2N^{2/3} - N^{1/3})$ as a good approximation of $\phi(N_i)$ satisfying the key equation. We transform the equation into simultaneous Diophantine approximation problem, then we apply lattice basis reduction algorithm in order to obtain the parameter (d, k_i) and later compute the prime factor p_i and q_i of each $N_i = p_i^2 q_i$ simultaneously in polynomial time.

Theorem 3.1. Suppose that $m \geq 2$, $N_i = p_i^2 q_i$, $1 \leq i \leq m$ be m moduli each with the same size N where $N = \min N_i$. Assume e_i , $i = 1, \dots, m$ be m public exponents. Let $\Phi = 2N^{2/3} - N^{1/3}$, $1 < e < \phi(N_i) < N_i - \Phi$. Define $\delta = \frac{m(1-\beta)}{m+1}$. If there exist an integer $d < N^\delta$ and m integers $k_i < N^\delta$ such that $e_i d - \phi(N_i) k_i = 1$, then it is possible to factor m moduli $N_i = p_i^2 q_i$ in polynomial time.

Proof.

Suppose $m \geq 2$ and $i = 1, \dots, m$, the equation $e_i d - \phi(N_i) k_i = 1$ can be written as $e_i d - k_i(N_i - \Phi) = 1 - k_i(N_i - \phi(N_i) - \Phi)$. Hence,

$$\left| \frac{e_i}{N_i - \Phi} d - k_i \right| = \frac{|1 - k_i(N_i - \phi(N_i) - \Phi)|}{N_i - \Phi} \quad (1)$$

Let $N = \min N_i$ and suppose that $k_i < N^\delta$ and $(2N^{2/3} - N^{1/3}) < p^2 + \frac{N}{p} - p < ((2^{2/3} + 2^{-1/3})N^{2/3} - 2^{1/3}N^{1/3})$, we will get,

$$\begin{aligned} \frac{|1 - k_i(N_i - \phi(N_i) - \Phi)|}{N_i - \Phi} &\leq \frac{|1 + k_i(N_i - \phi(N_i) - \Phi)|}{N - \Phi} \\ &< \frac{1 + N^\delta (N - (2N^{2/3} - N^{1/3}) - \phi(N))}{\phi(N)} \\ &< \frac{1 + N^\delta (2p^{5/3}|p^{1/3} - q^{1/3}|)}{\phi(N)} \\ &< \frac{N^\delta (\frac{1}{6}N^\beta)}{\frac{2}{3}N} \\ &= \frac{1}{4}N^{\beta-1+\delta} \end{aligned} \quad (2)$$

By applying Theorem 2.2, we substitute (2) in (1). Then, we obtain

$$\left| \frac{e_i}{N_i - \Phi} d - k_i \right| < \frac{1}{4} N^{\beta-1+\delta}.$$

We can see the relation between $\left| \frac{e_i}{N_i - \Phi} d - k_i \right| < \frac{1}{4} N^{\beta-1+\delta}$ and $|q\alpha_i - p_i| < \varepsilon$ which is the condition of Theorem 2.2.

Now, we proceed to show the existence of integer d and the integers k_i . We assume $\varepsilon = \frac{1}{4} N^{\beta-1+\delta}$ and $\delta = \frac{m(1-\beta)}{m+1}$. We have

$$N^\delta \cdot \varepsilon^m = \left(\frac{1}{4}\right)^m \cdot N^{m\beta-m+\delta m+\delta} = \left(\frac{1}{4}\right)^m.$$

Since $\left(\frac{1}{4}\right)^m < 2^{\frac{m(m-3)}{4}} \cdot 3^m$ for $m \geq 2$, we get $N^\delta \cdot \varepsilon^m < 2^{\frac{m(m-3)}{4}} \cdot 3^m$ by applying Theorem 2.2. It follows that if $d < N^\delta$, then $d < 2^{\frac{m(m-3)}{4}} \cdot 3^m \cdot \varepsilon^{-m}$.

Summarizing for $i = 1, \dots, m$, we have

$$\left| \frac{e_i}{N_i - \Phi} d - k_i \right| < \varepsilon \text{ and } d < 2^{\frac{m(m-3)}{4}} \cdot 3^m \cdot \varepsilon^{-m}.$$

If the conditions of Theorem 2.2 are fulfilled, then this lead us to find d and k_i for $i = 1, \dots, m$ using the LLL algorithm.

Next, we look at the relation $\frac{e_i d - 1}{k_i} = \phi(N_i) = p_i(p_i - 1)(q_i - 1)$ of the equation $e_i d - k_i \phi(N_i) = 1$. We can compute $\gcd\left(\frac{e_i d - 1}{k_i}, N_i\right)$ which later give the prime factor p_i and q_i . This leads to the factorization of m moduli of the form $N_i = p_i^2 q_i$. This terminates the proof. ■

Example 3.1. To illustrate our proposed attack, we consider the three moduli and three public exponents as follows:

$$\begin{aligned} N_1 &= 263516113503680842149862744216952225120329, \\ N_2 &= 128365563717380133604870768381638907086713, \\ N_3 &= 187981680633838672011014961959439098246519, \\ e_1 &= 80792911443410905692392637406380886780007, \\ e_2 &= 123578136992200918273607094829994961422647, \\ e_3 &= 20961451900454998790761026472974297795827. \end{aligned}$$

Then, $N = \min(N_1, N_2, N_3) = 128365563717380133604870768381638907086713$. If $m = 3$ and $\beta < 2/3$, then we have $\delta = \frac{m(1-\beta)}{m+1} = \frac{3}{8}$ and $\varepsilon = \frac{1}{4} N^{\beta-1+\delta} \approx 0.000001817121$. Suppose that we consider the parameter C as defined in [Nitaj et al. (2014), Appendix A, page 196], $n = m = 3$, leads to

$$C = \left[3^{n+1} \cdot 2^{\frac{(n+1)(n-4)}{4}} \cdot \varepsilon^{-n-1} \right] = 3714661586044869472880425.$$

Now, we consider the lattice \mathcal{L} spanned by the rows of the matrix

$$M = \begin{bmatrix} 1 & -\left[\frac{Ce_1}{N_1-\Phi}\right] & -\left[\frac{Ce_2}{N_2-\Phi}\right] & -\left[\frac{Ce_3}{N_3-\Phi}\right] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}.$$

Then, the LLL algorithm is applied to lattice \mathcal{L} leads to the reduced basis together with the matrix as follows.

$$K = \begin{bmatrix} 531150414983 & 301257228587 & 499463626192 & 178001942286 \\ -168389513847154765291 & -51627503530491328427 & -162109383604390765943 & -18776769540238268234 \\ 124074656146938841057 & 38040758013500360884 & 119447259923583684692 & 13835310590480351043 \\ 100588675112833001015 & 30840056847179847024 & 96837194594698950848 & 11216436984704075968 \end{bmatrix}.$$

Now, we obtain

$$K \cdot M^{-1} = \begin{bmatrix} 531150414983 & 162848441677 & 511341023600 & 59227494073 \\ -294151413649686019969 & -90185563220300129309 & -283180960997331683543 & -32800221212401676852 \\ 52224763545169658695 & 16011888761402698006 & 50277027552185434292 & 5823476337553763421 \\ 135845721544043483999 & 41649716234605959720 & 130779320408203043648 & 15147877965697829272 \end{bmatrix}.$$

According to the first row of the above matrix, we obtain

$$d = 531150414983, \quad k_1 = 162848441677, \quad k_2 = 511341023600, \quad k_3 = 59227494073.$$

By applying d and k_i for $i = 1, 2, 3$, we look at the relation $\frac{e_i d - 1}{k_i} = \phi(N_i) = p_i(p_i - 1)(q_i - 1)$ of the equation $e_i d - k_i \phi(N_i) = 1$, we get

$$\frac{e_1 d - 1}{k_1} = 263516113503672185830235590020192839191440,$$

$$\frac{e_2 d - 1}{k_2} = 128365563717374975692327918192177017411200,$$

$$\frac{e_3 d - 1}{k_3} = 187981680633832049824018263550877466963780.$$

Then, for each $i = 1, 2, 3$, we find $p_i = \gcd\left(\frac{e_i d - 1}{k_i}, N_i\right)$ and we obtain

$$p_1 = 69904752761407, \quad p_2 = 51738406927601, \quad p_3 = 58282229331211.$$

It is possible to factor three moduli N_1, N_2 and N_3 since

$$q_1 = 53925448837321, \quad q_2 = 47953733769113, \quad q_3 = 55340517648239.$$

4 THE SECOND ATTACK ON m MODULI $N_i = p_i^2 q_i$

In this section, we propose our second attack. Suppose that we are given m moduli $N_i = p_i^2 q_i$. We consider in this scenario that the following system of equation given by $e_i d_i - \phi(N_i)k = 1$

with fixed k will lead us the factor of each moduli which are all of the same size. We show that it is possible to factor such moduli of the form $N_i = p_i^2 q_i$. This is achievable when the unknown parameters d_i and k are suitably small. We couple this information together with the execution of the LLL algorithm to achieve our objective since we use the term $N - 2N^{2/3} - N^{1/3}$ as a good approximation of $\phi(N)$ satisfying the key equation.

We prove the following theorem based on Theorem 2.3. We transform the key equation into the simultaneous Diophantine approximation and then we apply lattice basis reduction algorithm to find parameters (d_i, k) which later lead to the factorization of m moduli of the form $N_i = p_i^2 q_i$ simultaneously.

Theorem 4.1. *Suppose that $m \geq 2$ and $N_i = p_i^2 q_i$, $1 \leq i \leq m$ be m moduli each with the same size N where $N = \max N_i$. Assume e_i , $i = 1, \dots, m$ be m public exponents with $\min e_i = N^\alpha$ and $\Phi = 2N^{2/3} - N^{1/3}$, $1 < e < \phi(N_i) < N_i - \Phi$. Define $\delta = \frac{m(\alpha-\beta)}{m+1}$. If exist an integer $k < N^\delta$ and m integers $d_i < N^\delta$ such that $e_i d_i - \phi(N_i)k = 1$ for $i = 1, \dots, m$, then it is possible to factor m moduli of the form $N_i = p_i^2 q_i$ in polynomial time simultaneously.*

Proof. For $i = 1, \dots, m$, starting with the equation $e_i d_i - k\phi(N_i) = 1$, we get

$$\left| \frac{N_i - \Phi}{e_i} k - d_i \right| = \frac{|1 - k(N_i - \phi(N_i) - \Phi)|}{e_i} \tag{3}$$

Let $N = \max N_i$ and suppose that $k_i < N^\delta$ and $(2N^{2/3} - N^{1/3}) < p^2 + \frac{N}{p} - p < ((2^{2/3} + 2^{-1/3})N^{2/3} + 2^{1/3}N^{1/3})$, we will get,

$$\begin{aligned} \frac{|1 - k(N_i - \phi(N_i) - \Phi)|}{e_i} &\leq \frac{|1 + k(N_i - \phi(N_i) - \Phi)|}{N^\alpha} \\ &< \frac{1 + N^\delta (N - (2N^{2/3} - N^{1/3}) - \phi(N))}{N^\alpha} \\ &< \frac{1 + N^\delta (2p^{5/3} |p^{1/3} - q^{1/3}|)}{N^\alpha} \\ &< \frac{N^\delta (\frac{1}{6} N^\beta)}{N^\alpha} \\ &= \frac{1}{6} N^{\beta+\delta-\alpha} \end{aligned} \tag{4}$$

By applying Theorem 2.2, we substitute (4) in (3). Then, we obtain

$$\left| \frac{N_i - \Phi}{e_i} k - d_i \right| < \frac{1}{6} N^{\beta+\delta-\alpha}.$$

We can see the relation between $\left| \frac{N_i - \Phi}{e_i} k - d_i \right| < \frac{1}{6} N^{\beta+\delta-\alpha}$ and $|q\alpha_i - p_i| < \varepsilon$ which is the condition of Theorem 2.2.

Now, we proceed to show the existence of integer k and the integers d_i . Let $\varepsilon = \frac{1}{6} N^{\beta+\delta-\alpha}$, $\delta = \frac{m(\alpha-\beta)}{m+1}$. We have

$$N^\delta \cdot \varepsilon^m = \left(\frac{1}{6}\right)^m \cdot N^{m\beta+m\delta-m\alpha+\delta} = \left(\frac{1}{6}\right)^m$$

Since $\left(\frac{1}{6}\right)^m < 2^{\frac{m(m-3)}{4}} \cdot 3^m$ for $m \geq 2$, we get $N^\delta \cdot \varepsilon^m < 2^{\frac{m(m-3)}{4}} \cdot 3^m$ by applying Theorem 2.2. It follows that if $k < N^\delta$, then $k < 2^{\frac{m(m-3)}{4}} \cdot 3^m \cdot \varepsilon^{-m}$.

Summarizing for $i = 1, \dots, m$, we have

$$\left| \frac{N_i - \Phi}{e_i} k - d_i \right| < \varepsilon \text{ and } k < 2^{\frac{m(m-3)}{4}} \cdot 3^m \cdot \varepsilon^{-m}$$

If the conditions of Theorem 2.2 are fulfilled, then this lead us to find k and d_i for $i = 1, \dots, m$ using the LLL algorithm.

Next, we look at the relation $\frac{e_i d_i - 1}{k} = \phi(N_i) = p_i(p_i - 1)(q_i - 1)$ of the equation $e_i d_i - k\phi(N_i) = 1$. We can compute $\gcd\left(\frac{e_i d_i - 1}{k}, N_i\right)$ which later give the prime factor p_i and q_i . This leads to the factorization of m moduli of the form $N_i = p_i^2 q_i$. This terminates the proof. ■

Example 4.1. As an illustration of this proposed attack, we look at three moduli and three public exponents as follows

$$\begin{aligned} N_1 &= 75008312957047469348732538527519866244681, \\ N_2 &= 123382641656631392246631643235917883162209, \\ N_3 &= 236640302081303358584350414610670308043781, \\ e_1 &= 31754220269884338669277120219619839718053, \\ e_2 &= 52233138249590873962310663900788852848217, \\ e_3 &= 185287571937475026853160676658447544281045. \end{aligned}$$

Then, $N = \max(N_1, N_2, N_3) = 236640302081303358584350414610670308043781$. We also get $\min(e_1, e_2, e_3) = N^\alpha$ with $\alpha \approx 0.9789170642$. Since $m = 3$ and $\beta < 2/3$, we get $\delta = \frac{m(\alpha-\beta)}{m+1} = 0.3591877982$ and $\varepsilon = \frac{1}{6}N^{\beta+\delta-\alpha} \approx 0.000001854212$.

Suppose that we consider the parameter C as defined in [Nitaj et al. (2014), Appendix A, page 196], $n = m = 3$, leads to

$$C = \left[3^{n+1} \cdot 2^{\frac{(n+1)(n-4)}{4}} \cdot \varepsilon^{-n-1} \right] = 3426234753047132066310644.$$

Next, we look at the lattice \mathcal{L} spanned by the rows of the matrix

$$M = \begin{bmatrix} 1 & -\left[\frac{C(N_1-\Phi)}{e_1}\right] & -\left[\frac{C(N_2-\Phi)}{e_2}\right] & -\left[\frac{C(N_3-\Phi)}{e_3}\right] \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix}.$$

Then, the LLL algorithm is applied to lattice \mathcal{L} leads to the reduced basis together with the matrix as follows

K

$$= \begin{bmatrix} -182524372654 & -146456473850 & -116350956196 & -161112380644 \\ -16683193846076224169 & -96006951081027194305 & -12803413578135971700 & 115420170547926858788 \\ -677461226345054183660 & 267316396489637852564 & 399791487973525630452 & 235778569220139902640 \\ 91011227434211881899 & -507503522055241331933 & 831338501094560594144 & -242139184242751706576 \end{bmatrix}.$$

Now, we obtain

$K \cdot M^{-1}$

$$= \begin{bmatrix} -182524372654 & -431150415597 & -431150415607 & -233111277947 \\ -16683193846076224169 & -39408249186843288687 & -39408249187757314228 & -21306966193871351804 \\ -677461226345054183660 & -1600266775567642689324 & -1600266775604758895587 & -865220627451237659280 \\ 91011227434211881899 & 214982404605424684025 & 214982404610410934689 & 116235126444902519447 \end{bmatrix}.$$

According to the first row of the above matrix, we obtain

$$k = 182524372654, d_1 = 431150415597, d_2 = 431150415607, d_3 = 233111277947.$$

By using k and d_i for $i = 1, 2, 3$, we look at the relation $\frac{e_i d_i - 1}{k} = \phi(N_i) = p_i(p_i - 1)(q_i - 1)$ of the equation $e_i d_i - k\phi(N_i) = 1$, we get

$$\begin{aligned} \frac{e_1 d_1 - 1}{k} &= 75008312957043771921448942071280309312160, \\ \frac{e_2 d_2 - 1}{k} &= 123382641656626251332238684396564345529200, \\ \frac{e_3 d_3 - 1}{k} &= 236640302081295426831936822683007388436496. \end{aligned}$$

Then, for each $i = 1, 2, 3$, we find $p_i = \gcd\left(\frac{e_i d_i - 1}{k}, N_i\right)$ and we obtain

$$p_1 = 45104908616699, \quad p_2 = 53063352809947, \quad p_3 = 65874665082797.$$

It is possible to factor three moduli N_1, N_2 and N_3 since

$$q_1 = 36869036060081, \quad q_2 = 43819224726601, \quad q_3 = 54532055826109.$$

5 CONCLUSION

In conclusion, this paper presents two new vulnerabilities on m moduli $N_i = p_i^2 q_i$. We focus on the system of key equation of the form $e_i d - \phi(N_i)k_i = 1$ for the first attack and the form $e_i d_i - \phi(N_i)k = 1$ for the second attack. We show that both of the attacks are successful when the parameters d, d_i, k and k_i are suitably small. It shows that these attacks are not dangerous. Additionally, we prove that both our attacks enable us to factor m moduli of the form $N_i = p_i^2 q_i$ simultaneously based on the LLL algorithm.

REFERENCES

- Asbullah, M. A. and Ariffin, M. R. K. (2014). Comparative Analysis of Three Asymmetric Encryption Schemes Based Upon the Intractability of Square Roots Modulo $N = p^2q$. In *The 4th International Cryptology and Information Security Conference 2014 (Cryptology2014)*, pages 86–99.
- Asbullah, M. A. and Ariffin, M. R. K. (2015). New Attack on RSA With Modulus $N = p^2q$ Using Continued Fractions. *Journal of Physics*, 622:191–199.
- Asbullah, M. A. and Ariffin, M. R. K. (2016a). Analysis on the AA_β Cryptosystem. In *The 5th International Cryptology and Information Security Conference 2016 (Cryptology2016)*, pages 41–48.
- Asbullah, M. A. and Ariffin, M. R. K. (2016b). Analysis on the Rabin- p cryptosystem. In *The 4th International Conference on Fundamental and Applied Sciences (ICFAS2016)*, pages 080012–1–080012–8. AIP Conf. Proc. 1787.
- Boneh, D. and Durfee, G. (1999). Cryptanalysis of RSA with private key d less than $N^{0.292}$. *Advance in Cryptology-Eurocrypt'99, Lecture Notes in Computer Science*, 1592:1–11.
- Bunder, M. and Tonien, J. (2017). A New Attack on the RSA Cryptosystem Based on Continued Fractions. *Malaysian Journal of Mathematical Sciences*, 11(S):45–57.
- Cassels, J. (1971). *Introduction to the Geometry of Numbers*. Springer-Verlag Berlin Heidelberg.
- de Weger, B. (2002). Cryptanalysis of RSA With Small Prime Difference. *Applicable Algebra in Engineering, Communication and Computing*, 13(1):17–28.
- Hinek, J. (2007). *On the security of some variants of RSA*. PhD thesis, Waterloo, Ontario, Canada.
- Howgrave-Graham, N. and Seifert, J. (1999). Extending Wiener attack in the presence of many decrypting exponents. In *Secure Networking-CQRE (Secure)'99 LNCS 1740 Springer-Verlag*, 1740:153–166.
- Lenstra, A. K., Lenstra, H. W., and Lovász, L. (1982). Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261:513–534.
- Maitra, S. and Sarkar, S. (2008). Revisiting Wieners Attack-New Weak Keys In RSA. In *Information Security Springer-Verlag*, pages 228–243.
- May, A. (2003). *New RSA vulnerabilities using lattice reduction methods*. PhD thesis, University of Paderborn.
- May, A. (2004). Secret exponent attacks on RSA-type scheme with moduli $N = p^r q$. In *PKC 2004 LNCS Springer-Verlag*, 2947:218–230.
- Menezes, A., Oorschot, P., and Vanstone, S. (1997). *Handbook Of Applied Cryptography*. CRC Press.

- Nishioka, M., Satoh, H., and Sakurai, K. (2002). Design and Analysis of Fast Provably Secure Public-Key Cryptosystems Based on a Modular Squaring. *In Information Security And Cryptology - ICISC 2001*, pages 81–102.
- Nitaj, A., Ariffin, M. R. K., Nassr, D. I., and Bahig, H. M. (2014). *New attacks on the RSA cryptosystem*, volume 8469 of *Lecture Notes in Computer Science*, pages 178–198. Springer-Verlag.
- Rivest, R., Shamir, A., and Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communication of the ACM* 21(2), 21(2):17–28.
- Takagi, T. (1998). Fast RSA-Type Cryptosystem Modulo p^kq . *Advances in Cryptology-CRYPTO'98*, 1462:318–326.
- Wiener, M. (1990). Cryptanalysis of short RSA secret exponents. *IEEE Transaction on Information Theory* IT-36, 36:553–558.

Accelerating DGHV's Fully Homomorphic Encryption With GPU

Jia-Zheng Goey^{*1}, Bok-Min Goi¹, Wai-Kong Lee², and Raphael C.-W. Phan³

¹*Lee Kong Chian Faculty of Science and Engineering, Universiti Tunku Abdul Rahman, Sungai Long, Malaysia.*

²*Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar, Malaysia.*

³*Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia.*

E-mail: goey.jz93@lutar.my, goibm@utar.edu.my, wklee@utar.edu.my, raphael@mmu.edu.my

**Corresponding author*

ABSTRACT

Fully Homomorphic Encryption schemes allow arithmetic operations to be performed on the encrypted data, which allows secure computation on sensitive data hosted in the cloud server. However, the speed performances of these schemes are very slow as it involves very large integer modular multiplication. In this paper, we present implementation of DGHV's encryption scheme (a variant of Fully Homomorphic Encryption) in GPU platform to accelerate the encryption operation. Speed up is achieved by parallelizing the modular multiplication using GPU with many computational cores. Schönhage-Strassen Multiplication Algorithm is used to realize very large integer multiplication, while Barrett's reduction is used to speed up the modular reduction. The timing performance of GPU acceleration is presented at the end of this paper.

Keywords: Homomorphic encryption, Security, Cloud Computing, GPU, Cryptography.

1 INTRODUCTION

1.1 Fully Homomorphic Encryption (FHE)

Fully Homomorphic Encryption scheme allows addition or multiplication to be performed on ciphertext. The mathematical manipulation of plaintext can be done on ciphertext without having to decrypt or acquire the secret key. The property is useful in cloud computing environment whereby it allows the processing of data in encrypted form. The first FHE scheme using ideal lattice was introduced by Gentry (2009) while the first integer based FHE is proposed later by

Van Dijk et al. (2010). The later scheme is relatively simpler and mathematically easier compared to lattice based FHE by Gentry (2009). Generally, FHE scheme consists of five stages as described below:

- **KeyGen** : Generate a large integer public and private key.
- **Encryption**: Encrypt one bit of data using the public key.
- **Evaluation**: Perform addition or multiplication operation on ciphertext.
- **Recreation**: Perform ciphertext refresh after each evaluation to reduce noise.
- **Decryption**: Decrypt ciphertext using private key.

1.2 General Purpose Computing on Graphic Processing Unit (GPU)

Graphic Processing Unit is utilized to handle computation on various complex algorithms. GPU's highly parallel architecture makes GPU more efficient in handling multiple tasks simultaneously. It enables the acceleration of an algorithm by offloading parallel parts to GPU. The non-parallelizable code runs in the CPU while parallelizable code runs in GPU. With proper implementation, application can run significantly faster compared to CPU-only implementation.

2 BACKGROUND

2.1 Fully homomorphic encryption over the integers

Van Dijk et al. (2010) describe the construction of their FHE scheme in their work. The authors start with construction of somewhat homomorphic encryption scheme, whereby the scheme only supports limited number of homomorphic operations on ciphertext.

- **DGHV.KeyGen**: Generate an odd number η -bit private key. Generate public key using the formula:

$$x_i = q_i \cdot p + r_i \quad (1)$$

Where q and r are random numbers, p is the private key. Public key set is rearranged so that x_0 is the largest public key element. $pk = \langle x_0, x_1, \dots, x_\tau \rangle$.

- **DGHV.Encrypt**: Encrypt one bit of message $\{0, 1\}$.

$$c = (m + 2R + 2 \sum_{i \in S} x_i) \text{ mod } x_0 \quad (2)$$

Where R is a random number, x_0 : largest integer in public key, S : random subset of $\{1, 2, \dots, \tau\}$.

However, there is a problem with somewhat homomorphic encryption scheme. Whenever homomorphic operation is executed, noise level increases in ciphertext. Van Dijk et al. (2010)'s

scheme can only support up to certain level of homomorphism due to the increasing noise level. Any ciphertext with noise higher than the permitted level will result in wrong decryption. For the ciphertext to decrypt correctly, noise parameter (ρ) must be smaller than the private key (p). In order to overcome this bottleneck, Van Dijk et al. (2010) adapted Gentry (2009)'s transformation decryption circuit squashing technique.

$$m \leftarrow (c \bmod p) \bmod 2 \quad (3)$$

The old decryption function cannot be expressed as a low depth Boolean circuit, thus they introduced a new decryption function.

- DGHV.Decrypt: Decrypt ciphertext to show the original message.

$$m \leftarrow [c - (\sum S_i \cdot Z_i)] \bmod 2 \quad (4)$$

Where m is the message, c is the ciphertext, S is the random hamming weight vector, and Z is the vector of re-encryption of ciphertext.

Then, ciphertext refresh technique is introduced to reduce the noise so that it is able to undergo homomorphic operation again. To refresh a ciphertext, homomorphically evaluate encryption of ciphertexts bits and encryption of secret-key bits with the new decryption circuit. The circuit will produce an encryption of plaintext bits, which is the refreshed ciphertext. With ciphertext refresh technique, unlimited homomorphic operations are now possible because refreshed ciphertext contains lesser noise than original ciphertext.

2.2 Public Key Compression and Modulus Switching for Fully Homomorphic Encryption over the Integers

The original DGHV's scheme by Van Dijk et al. (2010) is not practical due to its large public key size for security purpose, thus resulting in very slow encryption and decryption speed. Coron et al. (2012) described a compression technique that reduces public key size of DGHV's FHE from $O(\lambda^7)$ to $O(\lambda^5)$. Their improved variant remains semantically secure, but in random oracle model. They obtain an implementation of full scheme with 10.1MB public key.

Instead of storing a set of large bit size public key integers, authors propose to only store the small correction δ_i and random seed se and stated that it is sufficient to recover the public key x_i . The steps to compress the public key is shown in Algorithm 1. Coron et al. (2012) introduce a new encryption equation (Algorithm 2) to pair with their compressed public key. Decryption equation and ciphertext refresh equation remain the same as original DGHV's scheme.

2.3 Schönhage-Strassen Multiplication Algorithm (SSMA)

SSMA is a fast multiplication algorithm with low computational complexity of $O(n \log n \log \log n)$. It is often used for large integer of several thousand bits as analysed by Wang et al. (2015). This multiplication algorithm is being adopted in our implementation to speed up encryption of DGHV's scheme. Algorithm 3 shows the pseudocode for SSMA.

Algorithm 1 Public Key Compression

Generate random prime integer p of size ρ bits.
 Pick random odd integer $q_0 \in [0, 2^\gamma/p]$.
 Let $x_0 = q_0 \cdot p$.
 Initialize pseudo-random number generator f with random seed se .
 Generate a set of integers $X_i \in [0, 2^\gamma/p]$ for $1 \leq i \leq \tau$.
for $1 \leq i \leq \tau$ **do**
 $\delta_i = \langle x_i \rangle + \xi \cdot p - r_i$
endfor
for $1 \leq i \leq \tau$ **do**
 $x_i = X_i - \delta_i$
endfor
 $pk = (se, x_0, \delta_1, \dots, \delta_\tau)$
 Compute small corrections δ_i where $x_i = X_i - \delta_i$.
 Store δ_i in public key.

Algorithm 2 Compressed Public Key Encryption

Recover public key X_i using $f(se)$.
for $1 \leq i \leq \tau$ **do**
 $x_i = X_i - \delta_i$
 Choose random integer vector b .
for $1 \leq i \leq \tau$ **do**
 $b_i \in [0, 2^a]^\tau$
endfor
 Choose random integer $r \in (-2^p, 2^p)$.
 Encrypt 1-bit m message $c = [m + 2r + 2 \sum_{i=1}^\tau (b_i \cdot x_i)] \bmod x_0$

Algorithm 3 SSMA

Input: x_i and y_i , the series of input data of multiplier and multiplicand in time domain.
Output: $z_i = x_i \times y_i$
 $X \leftarrow FFT(x), Y \leftarrow FFT(y)$
for $0 \leq i \leq N - 1$ **do**
 $Z[i] \leftarrow X[i] \times Y[i]$ ▷ Convolution
endfor
 $z \leftarrow IFFT(Z)$
 $z \leftarrow Evaluation(z)$ ▷ Resolve carries
 Return Z

In the pseudocode, x and y represent two series of input data of the operands (multiplicand and multiplier). First, both x and y are forward transformed into their respective frequency domain, X and Y . Secondly, X and Y undergo convolution (pair-wise multiplication) and resulted Z , the product of the multiplication in frequency domain. Then, Z is converted back to its time domain, z using invert FFT. Lastly, in the evaluation step, the carries of each of the element in z is resolved to produce the final answer.

2.4 Accelerating fully homomorphic encryption using GPU

Wang et al. (2012) believes that FHE plays an important role in secure cloud computing. Despite the promising claim, FHE is far from being practical due to its large public key size and long encryption time. They proposed GPU realization of lattice based FHE by Gentry and Halevi (2011).

Their implementation shows significant speed up over the existing CPU implementation. They combined Schönhage-Strassen's FFT based integer multiplication algorithm with Barrett's modular reduction algorithm to implement an efficient modular multiplier that supports operands with million bits in FHE. They achieved speed up with factor of 8, 8 and 7.6 for decryption, encryption and reryption respectively using NVIDIA C2050 over CPU.

3 IMPLEMENTATION TECHNIQUES

In this paper, we aim to improve the encryption time of DGHV's encryption scheme by Van Dijk et al. (2010) because encryption is one of the frequently used function. Other parts of the DGHV's scheme by Van Dijk et al. (2010) will be parallelize and optimize in the future.

The FHE scheme is design by Deevashwer using GNU Multi Precision Arithmetic library function (GMP), which supports large integer operations. The most frequently used operations are multiplication and modulus in encryption function. The pseudocode of DGHV's encryption by Van Dijk et al. (2010) is as shown in Algorithm 4.

3.1 Schönhage-Strassen Multiplication Algorithm with Cooley-Tukey Fast Fourier Transform

The next stage of porting encryption operation into GPU is by replacing GMP multiplication with Schönhage-Strassen Multiplication Algorithm (SSMA) with Cooley-Tukey Fast Fourier Transform (CT-FFT). Total of three kernels are designed for forward CT-FFT.

The twiddle factors are constants for each of the FFT size, hence can be precomputed before load into GPU. There will be N number of twiddle factors for N -points FFT. Twiddle factors of small FFTs(8-points, 16-points, 32-points) are stored in the GPU registers for fast access. Only 64 twiddle factors are precomputed and stored for the 4K-points FFT while the remaining will

Algorithm 4 DGHV's Encryption

```

for  $0 \leq i \leq 12$  do
  for  $0 \leq j \leq 12$  do
    Generate random  $tmp$ 
     $tmp = public\_key[2 \times j] \times public\_key[2 \times (i - 1)]$ 
     $tmp$  modulo with  $public\_key[0]$ 
     $tmp = tmp \times tmp$ 
     $tmp$  modulo with  $public\_key[0]$ 
     $tmp$  multiply by 2
     $tmp$  modulo with  $public\_key[0]$ 
     $ct = ct + tmp$ 
     $ct$  modulo with  $public\_key[0]$ 
  endfor
endfor

```

be computed on the fly. Total of three kernels are designed for forward CT-FFT.

First kernel: Column-FFT implemented with $2 * N_2$ number of blocks with 64 threads each. Each thread will compute 64-points FFT simultaneously. Therefore, each block is capable of computing 4K-points FFT (64 threads * 64-points FFT). $2 * N_2$ number of blocks are launched in parallel to compute the column-FFTs for both input operands (x and y).

Second kernel: Twiddle Factors multiplication implemented to multiply the output from first kernel with respective twiddle factors. Every twiddle factors are only used once, therefore shared memory is not required.

Third kernel: Row-FFT implemented with 8 blocks, 1024 threads each. First 4096 threads will compute row FFTs for X operand while the remaining 4096 threads will compute for Y operand.

After forward CT-FFT is completed for both operands, another kernel will take over for the convolution process. Inverse CT-FFT is implemented the same way as forward CT-FFT with an additional kernel to multiply the outcome inverse FFT with FFT's multiplicative inverse N^{-1} .

3.2 Barrett's Reduction using SSMA with CT-FFT

Referring to encryption Algorithm 1 above, modulus operation follows after each multiplication to reduce the size of intermediate values. In order to be more efficient in multiplication and reduction, we implemented Barrett's reduction to replace GMP modulus function. The basic Barrett's reduction algorithm by Katz et al. (1996) is shown in Algorithm 5.

Barrett's reduction is more efficient compare to normal modulus because all division in the algorithm can be implemented using right shifting. We replaced multiplication in Barrett's reduction using SSMA with CT-FFT for faster computation in GPU. Our variant of Barrett's Reduction requires precomputation for q and μ using GMP library. We created a few extra functions to simplify the arithmetic operations of multi limbs integer, they are `poly_sub`-subtraction for multi limb integer, `poly_rshift`-right shifting for multi limb integer, `poly_cmp`-comparison for

Algorithm 5 Barrett's Reduction Algorithm

```

procedure BARRETT( $t, M$ )
   $q \leftarrow 2 \lceil \log_2(M) \rceil$ 
   $\mu \leftarrow \lfloor 2^q/M \rfloor$ 
   $r \leftarrow t - M \lfloor t\mu/2^q \rfloor$ 
  while  $r \geq M$  do
     $r \leftarrow r - M$ 
  endwhile
  return  $r$ 
endprocedure

```

▷ Output: $r = t \bmod M$
 ▷ Precomputation
 ▷ Precomputation
 ▷ $r = t \bmod M$

multi limb integer. The following is our implementation of Barrett's reduction with SSMA in Algorithm 6 and the timing for operations within Barrett's Reduction with SSMA is in Table 1.

Algorithm 6 Barrett's Reduction Algorithm with SSMA

```

procedure REDUCTION WITH SSMA(* $output$ , * $input$ , * $M, \mu, q$ )
   $q \leftarrow 2 \lceil \log_2(M) \rceil$ 
   $\mu \leftarrow \lfloor 2^q/M \rfloor$ 
   $temp = \mu \times input$ 
   $temp = temp \gg q$ 
   $temp = temp \times M$ 
   $temp = input - temp$ 
  while  $temp \geq M$  do
     $output = temp - M$ 
  endwhile
  return  $output$ 
endprocedure

```

▷ Precomputation
 ▷ Precomputation
 ▷ Multiply using SSMA
 ▷ Shifting using poly_rshift
 ▷ Multiply using SSMA
 ▷ Subtract using poly_sub
 ▷ Compare using poly_cmp
 ▷ Subtract using poly_sub

Operation	Time taken per iteration(s)	Minimum iteration	Time taken (s)
SSMA	0.195	2	0.390
poly_rshift	0.001	1	0.001
poly_sub + poly_cmp	0.001	1	0.001
Total time taken:			0.392

Table 1: Detail Operations within Barrett's Reduction with SSMA.

4 IMPLEMENTATION RESULT

The implementation is executed for security level $\lambda = 42$, the minimum security for cryptography scheme. The largest multiplication are two operands of 160k bits multiply together.

With all the modification being applied into encryption algorithm and precomputation of public key multiplication, our variant of the encryption algorithm is as describe in Algorithm

7. Precomputation of public key multiplication is possible because for any encryption using the same public key, the value of `pk_mul` will remain the same. By not computing the same value for each encryption, we are able to accelerate the encryption significantly. Timing for accelerated encryption is tabulated in Table 2.

Algorithm 7 DGHV’s Encryption with SSMA and Barrett’s Reduction

```

procedure ENCRYPT(ct, m)
  for  $1 \leq i \leq 12$  do
    for  $1 \leq j \leq 12$  do
       $pk\_mul = pk[2 \times j] \times pk[2 \times (i - 1)]$  ▷ Precomputation
    endfor
  endfor
  for  $1 \leq i \leq 144$  do
     $m = pk\_mul \bmod pk[0]$  ▷ Barrett’s Reduction
     $m = m \times gen\_tmp$  ▷ SSMA
     $m = m \bmod pk[0]$  ▷ Barrett’s Reduction
     $m = m \ll 2$  ▷ Poly_lshift
     $m = m \bmod pk[0]$  ▷ Barrett’s Reduction
     $ct = ct + m$  ▷ Poly_add
     $ct = ct \bmod pk[0]$  ▷ Barrett’s Reduction
  endfor
endprocedure

```

Operation	Time taken per iteration(s)	Iterations required	Time taken (s)
Barrett’s Reduction	0.390	576	224.60
SSMA	0.195	144	28.08
poly_lshift	0.046	144	6.62
poly_add	0.024	144	3.46
Total time taken:			262.76

Table 2: Accelerated encryption timing.

5 CONCLUSION

The implementation of DGHV’s encryption in GPU aims to speed up DGHV’s encryption for fully homomorphic cryptography to be practical for common use. Future work involves accelerating DGHV’s scheme by part, decreasing encryption, decryption, evaluation, reencryption time.

ACKNOWLEDGMENTS

We would like to express our gratitude to Universiti Tunku Abdul Rahman Research Fund (UTARRF) for supporting this research under the grant number IPSR/RMC/UTARRF/2016-C1/G1.

REFERENCES

- Coron, J.-S., Naccache, D., and Tibouchi, M. (2012). Public key compression and modulus switching for fully homomorphic encryption over the integers. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 446–464. Springer.
- Gentry, C. (2009). *A fully homomorphic encryption scheme*. Stanford University.
- Gentry, C. and Halevi, S. (2011). Implementing gentry's fully-homomorphic encryption scheme. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 129–148. Springer.
- Katz, J., Menezes, A. J., Van Oorschot, P. C., and Vanstone, S. A. (1996). *Handbook of applied cryptography*. CRC press.
- Van Dijk, M., Gentry, C., Halevi, S., and Vaikuntanathan, V. (2010). Fully homomorphic encryption over the integers. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 24–43. Springer.
- Wang, W., Hu, Y., Chen, L., Huang, X., and Sunar, B. (2012). Accelerating fully homomorphic encryption using gpu. In *High Performance Extreme Computing (HPEC), 2012 IEEE Conference on*, pages 1–5. IEEE.
- Wang, W., Hu, Y., Chen, L., Huang, X., and Sunar, B. (2015). Exploring the feasibility of fully homomorphic encryption. *IEEE Transactions on Computers*, 64(3):698–706.

Evaluation Criteria on Random Ambience for Cryptographic Keys

Nur Azman Abu^{*1}, Shekh Faisal Abdul Latip¹, and Shahrin Sahib¹

¹Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM),, Hang Tuah Jaya, 76100 Durian Tunggal, , Melaka, Malaysia

E-mail: nura@utem.edu.my

**Corresponding author*

ABSTRACT

There is a need nowadays to ensure information security which is independent of the security mechanism of physical medium. Cryptography still remains an important science in daily life, be it for sovereignty use or the privacy of individuals. Ultimately, the security of the interoperating open cryptosystem must reside only in the key being used. It must be true random key. It is a challenging task to generate a true random key live on demand suitable for cryptographic applications. This paper shall formally propose an evaluation criteria on a true random number generator(TRNG). An evaluation has been done on various TRNG for cryptographic keys. This paper will also classify recent TRNGs into several groups. A new true random generator has been developed based on the air ambience for cryptographic application and evaluated against other TRNGs under the proposed evaluation criteria.

Keywords: True Random Numbers, Random Cryptographic Keys, Evaluation Criteria of Random Numbers

1 INTRODUCTION

In a modern cryptographic design, a chosen algorithm must provide a high level of security, completely specified in a clear easy manner for cryptographic community to understand, analyze and validate. The security of the cryptographic system should not depend on the secrecy of the algorithms nor apparatus being used. They shall be made public and available to all users. For the purposes of interoperability, they must be efficient, exportable, economically implementable in electronic devices and adaptable for use in diverse future applications. The security of modern cryptographic system should be no longer based on the secrecy of the system design but rather on the randomness of the key being used. Most of the cryptographic operations nowadays mandate a fresh random key as an input. Apparently, these operations are also mostly designed and taken care of by the developers of the cryptosystem.

2 TRUE RANDOM NUMBER GENERATORS

This section shall give an overview on several various types of true random number generators. Entropy is often used as a measure of the unpredictability of a cryptographic key. Most TRNGs make use of a natural phenomenon as an entropy source to produce random bits. They come in many types, ranging from Johnson noise in the case of Petrie and Connelly (2000) and ring oscillators by Sunar et. al. (2007) to a key generation scheme exploiting randomness of hyperchaos (Teh and et. al., 2016) and road surface plus driving behavior (Uz-Zaman et. al., 2017). Each TRNG captures and measures unpredictable natural processes by using a dedicated hardware device. The measurement usually follows certain principle of physics.

2.1 TRNGs around the Globe

There are various cryptographic research groups concentrating on TRNG around the globe. They have recently managed to produce their own random number generators. The selected research groups under review are listed in Table 1.

As the realm of cryptography has embarked on an open system, a cryptosystem will require a freshly minted key for each cryptographic operation. An entropy of a random variable is a mathematical measure (Barker and Roginsky, 2012). Recent research on cryptographic key generation based on physical biometrics is a popular topic (Dinca and Hancke, 2017). However, a key generator with a low entropy, such as using a fuzzy logic (Koeberl and et. al., 2014) and deterministic pilot signals from random fading gains (Fritschek and Wunder, 2017), is out of scope of this reseach project. Based on the literature survey on recent TRNGs above, every developer and country should have their own TRNG as the source of a secure random cryptographic key.

2.2 Special Devices of TRNG

Unlike the PRNG, a physical hardware random number generator has a greater advantage, since it can produce completely unpredictable and un-reproducible random sequences. Most of the true random number generators recently are designed based on special hardware devices such as a quantum detector device, an electronic flip-flop circuit and a chaos oscillator digital circuit. For instances, a quantum detector devices has been used by Stipcevi and Rogina (2007), Dynes et. al. (2008), Kwon et. al. (2009) and Frst et. al. (2010). An electronic flip-flop has been used by Sunar et. al. (2007), Thamrin et. al. (2008) and Fechner and Osterloh (2010). A chaos oscillator has been used by Ergn and zoguz (2008), Blaszczyk and Guinee (2008) and Danger et. al. (2009). A literature survey has been done on the special hardware requirement in TRNGs. They are based on certain theoretical principles and special hardware to capture their source of randomness. In this research project, a systematic comparison on TRNGs has been hardly found. Some authors also have pointed out that to the best of their knowledge, no effective comparison of several TRNGs appears in the literature (Santoro et. al., 2009). Most research groups have proposed their techniques and reported their performances in terms of passing a random statistical test suite and output bitrates only. In order to achieve a viable TRNG, it

ID	Recent TRNG	Institution	City/State/Country
1	Sunar <i>et. al.</i> (2007)	Worcester Polytechnic Institute	Worcester, Massachusetts
2	Stipcević and Rogina (2007)	Rudjer Bosković Institute	Zagreb, Croatia
3	Drutarovský and Galajda (2007)	Technical University of Košice	Slovak Republic
4	Tokunaga <i>et. al.</i> (2008)	University of Michigan	Ann Arbor, Michigan
5	Dynes <i>et. al.</i> (2008)	Cambridge Research Laboratory	Cambridge, United Kingdom
6	Ahmed and Naganathan (2008)	VLB Engineering College	Coimbatore, India
7	Ergün and Özoguz (2008)	Istanbul Technical University	Istanbul, Turkey
8	Thamrin <i>et. al.</i> (2008)	MIMOS	Malaysia
9	Błaszczuk and Guinee (2008)	Cork Institute of Technology	Bishopstown, Ireland
10	Wayne <i>et. al.</i> (2009)	University of Illinois	Urbana-Champaign, Illinois
11	Holcomb <i>et. al.</i> (2009)	University of Massachusetts	Amherst, Massachusetts
12	Kwon <i>et. al.</i> (2009)	Pohang University of Science and Technology	Pohang, South Korea
13	Danger <i>et. al.</i> (2009)	Telecom ParisTech	Paris, France
14	Wei and Guo (2009)	Peking University	Beijing, China
15	Bardis <i>et. al.</i> (2009)	University of Military Education	Vari, Greece
16	Hars (2009)	Seagate Technology	Longmont, Colorado
17	Hirano <i>et. al.</i> (2010)	Takushoku University	Tokyo, Japan
18	Pironio <i>et. al.</i> (2010)	Université Libre de Bruxelles	Bruxelles, Belgium,
19	Fürst <i>et. al.</i> (2010)	Ludwig-Maximilians Universität	München, Germany
20	Fechner and Osterloh (2010)	University of Hagen	Hagen, Germany
21	Argyris (2010)	University of Athens,	Panepistimiopolis, Ilisia, Greece
22	Abu and Sahib (2010)	Universiti Teknikal Malaysia (UTeM)	Melaka, Malaysia

Table 1: The institution, city, state or country of origin from TRNG's research group.

should satisfy certain modern criteria as a competitive random cryptographic generator.

3 MODERN CRITERIA ON RANDOM CRYPTOGRAPHIC KEY GENERATION AND APPARATUS

In this paper, the authors shall propose a set of evaluation criteria to rate a TRNG according to the Table 2 below. In general, the TRNG should only use minimum hardware and utilize only few basic computer algorithms. The users and owners of a cryptosystem are mostly nontechnical. It is important to have a generation process and apparatus which are physical, economics, convenient, efficient to use and secure. The measuring device should work automatically on demand from user’s physical environment.

Index	Criterion
1	Minimum Hardware
2	Minimum Formula
3	Basic Algorithm
4	Efficiency
5	Economics
6	Mobility

Table 2: Modern Evaluation Criteria on TRNG for cryptographic keys.








The last criterion shall be referred to as mobility of the TRNG. This is the most critical criterion in this research project. It will determine whether the TRNG can and shall be used in generating the cryptographic key in each crypto operation independent of the developer preset RNG. It is common for the developer to use pseudo RNG instead of true RNG. In summary, this research project shall consider the six criteria listed in Table 2 above to evaluate each TRNG. These criteria are critical factors in practical usage of embedding the TRNG in current modern cryptosystem.

Symbol	1. 	2. 	3. 	4. 	5. 
Quality	Poor	Satisfactory	Good	Very Good	Outstanding

Table 3: Legend symbols for proxy qualitative variable for TRNG evaluation.



The symbolic scores shall be given according to the legend symbols which get the quantitative values from 1 to 5 as shown in Table 3 . In the end, a proxy random variable will take an average qualitative score on each TRNG according to the modern criteria above. Every criterion shall be utilised to evaluate the chosen TRNG for cryptographic keys based on basic evaluation rules as prescribed in the following subsections. The objective of these criteria is partly to distinguish a practical TRNG from pseudo RNG for cryptographic keys.


3.1 Evaluation Rule on Minimum Hardware

The use of additional hardware in a TRNG shall be evaluated in this subsection according to the complexity and technology in use. The laser source quantum device in general shall be given qualitative score 1 or symbolic score  since it takes more complex hardware to control the environment during the random number generating process. Any TRNG using semi-conductor laser shall be given qualitative score 2 or symbolic score  on minimum hardware requirement. The LED source QTRNG has been given qualitative score 2 or symbolic score . Whereas since an oscillator are suitable for high-frequency and high-performance integrated circuit, such TRNG shall be given qualitative score 3 or symbolic score  on minimum hardware requirement. Similarly, chaos system can be embedded as microcontroller, this TRNG system should be given a relatively balanced qualitative score 3 or symbolic score . Any system which uses common computer component with certain requirement of specific brand or model shall be given almost full qualitative score 4 or symbolic score  on minimum hardware requirement. Lastly, a TRNG utilising a common standard peripheral or computer component shall be given full qualitative score 5 or symbolic score  on this criterion.

3.2 Evaluation Rule on Minimum Formula

A practical TRNG should be distinct from pseudo RNG for cryptographic keys which uses heavy mathematical formula to generate the random output. In general the minimum formula refers to the mathematical formula being used in capturing the candidate bits from the random source. The more complex mathematical formula being used shall be given the lower qualitative score.

For instance, Stipcević and Rogina (2007) basic idea of the method for extracting random bits is to consider a pair of non overlapping random time intervals. However, the time interval between subsequent random pulses is measured by counting periodic pulses from a continuous non-restartable quartz-controlled clock. Nevertheless, the statistics of time intervals between emission of subsequent photoelectrons is governed by convolution of exponential and differential binomial distributions. The probability density function (pdf) of measured time intervals between subsequent pulses has to be measured by the fast digital oscilloscope. Since the whole bit extraction process has been embedded in logic circuitry, this TRNG shall be given qualitative score 2 or symbolic score  instead of 1. Similarly, Tokunaga et. al. (2008) makes use of statistical analysis in order to tune the system into metastability. A simpler mathematical formulas being utilized such as mapping equations by Drutarovský and Galajda (2007) shall be given the medium qualitative score 3 or symbolic score .

In another instance, Sunar *et. al.* (2007) design uses jitter (or random vibration) in clock signals present in all digital clocked circuits as the random source. They harvest the jitter by sampling an output of coupled oscillators. The output of the XOR combines periodic transition zones contributed by each oscillator ring. Since the method uses more than a simple XOR and digital sampling, this TRNG has been given almost full qualitative score 4 or symbolic score . In an ideal instance, Dynes *et. al.* (2008) captures the output bit by recording the time tagged count events. The time tagged events acquired were converted into random bits by assigning detection events in even clock cycles as a “1” and “0” for detection events in odd clock cycles.

A direct capture of random output bit such as this TRNG shall be given full qualitative score 5 or symbolic score ● on this criterion.

3.3 Evaluation Rule on Basic Algorithm

This evaluation rule on basic algorithm specifically pays attention to the computer algorithm in the generating process of random key. Special attention is given to the post-processing of the random bits being captured. Even though hashing algorithm is well known to secure the random number and make it irreversible, in this research project, the evaluation rule shall give it the lowest qualitative score 1 or symbolic score ●. A TRNG should stand on its own without relying on such a hashing algorithm. However, few TRNG uses some complex electronic algorithm without hashing function has been given a qualitative score 2 or symbolic score ● such as Tokunaga *et. al.* (2008). Similarly, Bardis *et. al.* (2009) utilises normalization transformation to create the packet of the uniformly distributed random variables.



A TRNG which uses the von Neumann method, in general shall be given the qualitative score 3 or symbolic score ●. A TRNG which uses the XOR corrector, in general, shall be given the qualitative score 4 or symbolic score ●. An unbiased TRNG shall be given full qualitative score 5 or symbolic score ● on this criterion such as Dynes *et. al.* (2008) without a need of any post-processing algorithm.



3.4 Evaluation Rule on Efficiency


Efficiency here specifically refers to the speed or potential capacity of TRNG. A qualitative score shall be based on the random output bit-rate such as few hundred bits per second, kilobits per second, few hundred kilobits per second, several megabits per second and in terms of gigabits per second and beyond. A TRNG which produces 500 random bits per second or less shall be given the lowest qualitative score 1 or symbolic score ●. Next, a TRNG which is capable of producing few thousand bits per second less than 50 Kbps shall be given the qualitative score 2 or symbolic score ●. Third, a TRNG which has random output bitrate between 50 Kbps and 500 Kbps shall be given the medium qualitative score 3 or symbolic score ●. Fourth, a TRNG which has random output bitrate few megabits per second between 500 Kbps and 500 Mbps shall be given the almost full qualitative score 4 or symbolic score ●.

Lastly, a TRNG which has random output bitrate in terms few gigabits per second specifically higher than 500 Mbps shall be given the full qualitative score 5 or symbolic score ● on this criterion. Even though (Wei and Guo, 2009) has reported the process and apparatus to generate random keys at the rate of 500 Kbps, they have also shown a clear technique and modus operandi on how to generate the random key in terms of gigabits per second. Thus, they are given the full qualitative score 5 or symbolic score ● on this criterion.




3.5 Evaluation Rule on Economics




The economic issue here refers to the cost or relative price of certain TRNG. The cost includes the cost deployment of a TRNG as a cryptographic key generator for nontechnical user applications. For instance, in Dynes *et. al.* (2008), the avalanche photodiode (APD) is held at a temperature of -30°C . This requirement is certainly very costly for deployment in general cryptographic environment. In general, the quantum TRNG shall be given qualitative score 1 or symbolic score . Simpler setup on the quantum TRNG shall be given a qualitative score 2 or symbolic score .

A quantum photonic TRNG comprises an optical system and extremely compact circuitry which can be implemented on any advance crypto hardware devices. For instance, Argyris (2010) make use of a photonic integrated circuit as its bit extraction controller. A photonic TRNG are mostly given a qualitative score 3 or symbolic score  on economic criterion. Even better, an oscillator TRNG is suitable for high-frequency and high-performance integrated circuit. An oscillator TRNG without any requirement of specific hardware brand or model shall be given almost full qualitative score 4 or symbolic score .

Lastly, a simple TRNG without any special circuit which makes use of only common computer component shall be given full qualitative score 5 or symbolic score  on this criterion. It is a popular idea to generate the random bit based on nature collected using the common computer peripherals. These types of TRNGs shall be discussed in the second half of the next section.

3.6 Evaluation Rule on Mobility

A score on mobility of any TRNG shall be evaluated here according to its size and portability. The laser source quantum device in general shall been given qualitative score 1 or symbolic score  since it takes more complex hardware to control the environment during the random number generating process. Any TRNG using semi-conductor laser shall be given qualitative score 2 or symbolic score  on minimum hardware requirement. The LED source QTRNG has been given qualitative score 2 or symbolic score .

Chaos system, however, can be embedded as microcontroller. So, this TRNG system should be given a relatively balanced qualitative score 3 or symbolic score . Since an oscillator are suitable for high-frequency and high-performance integrated circuit, they TRNGs are mostly given qualitative score 3 or 4 on mobility criterion. Similarly, any system which uses common computer component with certain requirement of specific brand or model shall be given almost full qualitative score 4 or symbolic score  on mobility criterion since it is not ever ready to generate a random cryptographic key generation live on demand. Lastly, a TRNG utilising a common standard peripheral or computer component shall be given full qualitative score 5 or symbolic score  on this criterion.



4 CLASSIFICATION OF RECENT TRNGS

A TRNG shall always be an important primitive in a cryptosystem. This literature review shall be miniturized by classifying recent TRNGs into several groups. The security of every cryptographic operation primarily relies on the unpredictability of the random key being used. The source of randomness in every TRNG is the real world phenomenon. Some kind physical hardware device is needed to detect and record a continuous event. This section shall pay a particular attention to the minimum hardware and mobility criteria on classifying recent TRNGs under review.

4.1 Quantum TRNG

A quantum TRNG relies upon a physical process, extracting randomness from the inherent uncertainty in quantum mechanics. Notably, it relies on a photon detector as its special apparatus. It also requires another hardware as a source of light. There are 2 main sources of lights. First, it is using a light emitting diode (LED) such as in Stipcević and Rogina (2007), Wayne *et. al.* (2009) and Fürst *et. al.* (2010). Second, it is using a laser diode such as in Dynes *et. al.* (2008), Kwon *et. al.* (2009), Wei and Guo (2009) and Pironio *et. al.* (2010).


At the same time, it requires another set of tools to control the light source and the environment of the random bit generating process. Stipcević and Rogina (2007) makes use of a photo-multiplier with photocathode. Wayne *et. al.* (2009) utilises an optical laser diode attenuator. Fürst *et. al.* (2010) uses a photomultiplier tube. Whereas Dynes *et. al.* (2008) uses an avalanche photodiode (APD). Kwon *et. al.* (2009) requires a pair of interference filters, a fiber beam splitter and polarization controllers. Wei and Guo (2009) uses a flexible attenuator and an avalanche photodiode. Pironio *et. al.* (2010) requires a pair of independent vacuum chambers placed in the magnetic field and photomultiplier tubes.

The LED source QTRNG has been given qualitative score 2 or symbolic score . Whereas the laser source QTRNG has been given qualitative score 1 or symbolic score  since it takes more hardware to control the environment during the random number generating process.

4.2 Photonic TRNG



A quantum photonic TRNG relies on a photo detector/receiver as its special apparatus. The design comprises an optical system and extremely compact and digitally random bit extraction circuitry which can be implemented on any advance crypto hardware devices.

Thamrin *et. al.* (2008) use T-shape optical system and a bit extraction controller. The optical setup consists of an optical source, attenuators, a half-wave plate, a polarize beam splitter and two detectors. The bit extraction microcontroller consists of D flip-flop circuit. Argyris (2010) uses a photonic integrated circuit, a photo receiver/detector and a real-time oscilloscope as part of bit extraction controller.


Hirano *et. al.* (2010) have constructed by far a complex setup to reach high random output. They uses two distributed-feedback (DFB) semi-conductor lasers, a temperature controller, a fiber coupler, a variable fiber reflector, polarization maintaining fibers and AC photo detectors. The readings go through electronic amplifiers and a digital oscilloscope a radio-frequency (RF) spectrum analyzer. For this reason Photonic RNG has been typically given qualitative score 2 or symbolic score  on minimum hardware requirement.

4.3 Oscillator TRNG

Oscillators provide a simple and effective method to build TRNGs. They are suitable for high-frequency and high-performance integrated circuit. Typically, an oscillator TRNG is build upon chaotic system. Sunar *et. al.* (2007) have made use of oscillator rings, XOR gates and a D flip-flop sampler. This particular Oscillator TRNG is based on sampling phase jitter in oscillator rings or random vibration in clock signals present in a typical digital clock circuit. Ergün and Özoguz (2008) use a continuous-time chaotic oscillators, a voltage-controlled oscillator (VCO) and CMOS transistor arrays. Blaszczyk and Guinee (2008) use a double-scroll attractor from a chaotic oscillator based on Chua's circuit for a nonlinear operation leading to its chaotic behaviour. The circuit has been modified to obtain TRNG's performance using a simple temperature dependent control resistor in the oscillator circuit and optimal voltage threshold settings. To achieve optimal voltage, three current feedback operational amplifiers are used along with a voltage comparator. Danger *et. al.* (2009) have presented a new method to build a very high speed TRNG based on an open loop structure. This principle can be implemented in an FPGA. They also need to use an external adjustable potentiometer RC delay and a Peltier sensor.

Since Oscillator TRNGs are suitable for high-frequency and high-performance integrated circuit, they are mostly given qualitative score 3 or symbolic score  on minimum hardware requirement. For the same reason, they are mostly given a qualitative score 4 or symbolic score  on mobility.

4.4 Chaos TRNG

Drutarovský and Galajda (2007) has proposed a new robust chaos-based TRNG embedded in a true PSoC integrating configurable analog and digital peripheral functions, memory and a microcontroller on a single chip. The system requires a powerful Harvard CPU architecture. At the same time, the system uses advanced peripherals, namely, four rail-to-rail continuous analog PSoC blocks, eight Switched Capacitor (SC) analog blocks, eight digital PSoC blocks and a mixed-signal PSoC hardware which includes also an embedded microcontroller. Even though chaos system is popular within the cryptosystem, this system is given a relatively balanced qualitative score 3 or symbolic score  on minimum hardware requirement and mobility.

4.5 User Interaction TRNG

Ahmed and Naganathan (2008) have proposed a user interaction model consisting of the time-stamp of mouse movements, character input pressed on the keyboard, hard disk reading time and operating system states. This interaction system is certainly an ideal TRNG for this research project in term of cryptographic key generation live on demand at the user physical location. This system has been given full qualitative score 5 or symbolic score ● on minimum hardware requirement, economically and mobility. Unfortunately, it is not efficient and relies strongly on the hashing algorithm being used.

4.6 CMOS TRNG


Tokunaga *et. al.* (2008) have proposed a meta-stable system to generate individual bits that result from the effect of thermal noise. The system also comes with meta-stable latch, completion detector and a time-to-digital converter (TDC) and fabricated chips on 8-metal-layer bulk CMOS. The dynamic control module that tunes the latch into the meta-stable region responds to both process and temperature variations, as well as external noise sources. Holcomb *et. al.* (2009) use a 64-bit SRAM logical device consists of cross-coupled CMOS inverters and access transistors. The system generates random numbers from the power-up of SRAM and existing volatile CMOS memory without requiring any dedicated circuitry. The system relies on the large number of cells to ensure that some cells will be influenced by noise when the chip is powered-up. The primary limitation of this TRNG is that entropy is only generated during power-up which makes it unpractical for random cryptographic key generation to be done live on demand.

Since both systems use common computer component, they are given almost full qualitative score 4 or symbolic score ● on minimum hardware requirement. Holcomb *et. al.* (2009) should get a slightly lower score on mobility criterion since it is not ever ready to generate random cryptographic key generation live on demand.

4.7 Disk Drive TRNG


Hars (2009) specifically uses a Seagate Momentus FDE disk drive and a diagnostic interface between the main control ASIC and the channel signal processor to access the coefficients of an adaptive channel-filter. Since a disk drive is considered as a common computer component, this system is given almost full qualitative score 4 or symbolic score ● on minimum hardware requirement. A full score may be given if the system may utilise any standard disk drive without referring to any specific brand or model. Nevertheless, the system is given full qualitative score 5 or symbolic score ● on mobility.

4.8 SRAM TRNG



Although SRAM has been used in earlier TRNGs such as in Holcomb *et. al.* (2009), it has not been used as the main hardware or the principle source of random extraction. Fechner and Osterloh (2010) use a six-transistor SRAM in 26 nm process technology and meta-stable flip-flops. Since SRAM is considered as common computer component, this system is given almost full qualitative score 4 or symbolic score  on minimum hardware requirement.



4.9 Ambience TRNG

Bardis *et. al.* (2009) proposal use a basic peripheral of a personal computer, namely, a microphone. This device is a standard part of modern personal computers, laptop computers, PDAs and of course mobile phones. Bardis *et. al.* (2009) have investigated several environmental sounds captured under four different scenarios *id est* single person speaking in an office environment, almost silent office noise, cocktail party noise and mixed noise (office noise, multimedia sounds and human conversations).

Since microphone is just a common computing device, this system is given full qualitative score 5 or symbolic score  on minimum hardware requirement and mobility criteria. On the practical application of the idea, this research project shall follow and broaden this strategy in moving forward.

5 TRUE ENVIRONMENTAL RANDOM AMBIENCE NUMBER GENERATOR

In this research project, a true environmental random ambience number generator (TERANG) has been proposed. TERANG utilises only a common standard peripheral or computer component such as web camera (Abu and Sahib, 2011) and regular microphone (Abu and Sahib, 2010a). On the case of one-megabit random number generation, it uses high fidelity digital camera which can be purchased over the shelves (Abu and Sahib, 2010b). TERANG shall be given full qualitative score 5 or symbolic score  on the first criterion, Minimum Hardware requirement. The random output bit has been captured directly from an image pixel or an audio signal. TERANG uses only basic minimum formula to do the direct capture of random output bit. TERANG shall be given full qualitative score 5 or symbolic score  on the second criterion, Minimum Formula being used prior to post-processing.

For the post-processing stage, however, TERANG uses the XOR operation among output bit being captured in the colour digital image. It shall be given the qualitative score 4 or symbolic score  on the third criterion, Basic Algorithm. In the case of the fourth criterion, Efficiency here specifically refers to the speed or potential capacity of TRNG. TERANG's the random output bit-rate in the current form is categorised and fall within several megabits per second. It shall be given the qualitative score 4 or symbolic score .

Random ambience follows the basic principle to generate the random bit from natural phenomena collected using the common computer peripherals. They carry minimum cost and widely available without any special circuit to operate on. TERANG shall be given full qualitative score 5 or symbolic score ● on the fifth criterion, Economics. At the same time, utilising a common standard peripheral or computer component, they are ready to be deployed anywhere and certainly very mobile. TERANG shall be given full qualitative score 5 or symbolic score ● on this last criterion, Mobility.

Following the proxy qualitative variable on the evaluation criteria, every TRNG has been evaluated and ranked in ascending order by their average scores as shown in the Table 4. TERANG has achieved the highest score compared to the rest.

0	Min Hardware	Min Formula	Basic Algorithm	Efficiency	Economic	Mobility	Average
18	1. ●	3. ●	4. ●	2. ●	1. ●	1. ●	2.00
12	1. ●	3. ●	3. ●	3. ●	1. ●	2. ●	2.17
10	2. ●	3. ●	1. ●	4. ●	2. ●	2. ●	2.33
2	2. ●	2. ●	5. ●	4. ●	2. ●	2. ●	2.83
5	1. ●	5. ●	5. ●	4. ●	1. ●	1. ●	2.83
14	1. ●	3. ●	3. ●	5. ●	1. ●	4. ●	2.83
9	3. ●	3. ●	3. ●	2. ●	4. ●	3. ●	3.00
4	3. ●	2. ●	2. ●	4. ●	5. ●	3. ●	3.17
8	2. ●	3. ●	4. ●	5. ●	3. ●	2. ●	3.17
3	3. ●	3. ●	4. ●	3. ●	4. ●	3. ●	3.33
16	4. ●	3. ●	1. ●	2. ●	5. ●	5. ●	3.33
17	2. ●	4. ●	4. ●	5. ●	3. ●	2. ●	3.33
19	2. ●	3. ●	5. ●	4. ●	2. ●	4. ●	3.33
20	4. ●	3. ●	3. ●	2. ●	5. ●	3. ●	3.33
13	3. ●	3. ●	3. ●	4. ●	4. ●	4. ●	3.50
1	3. ●	4. ●	3. ●	4. ●	4. ●	4. ●	3.67
6	5. ●	5. ●	1. ●	1. ●	5. ●	5. ●	3.67
7	3. ●	3. ●	3. ●	5. ●	4. ●	4. ●	3.67
11	4. ●	4. ●	4. ●	2. ●	5. ●	3. ●	3.67
15	5. ●	3. ●	2. ●	2. ●	5. ●	5. ●	3.67
21	2. ●	4. ●	5. ●	5. ●	3. ●	4. ●	3.83

22	5. 	5. 	4. 	4. 	5. 	5. 	4.67
----	--	--	--	--	--	--	------

Table 4: Evaluation scores of the random ambience as a source of randomness among the recent TRNGs for cryptographic keys.

This concept of random ambience is certainly by far the most practical, convenient and robust TRNG suitable for cryptographic applications. The proposed TRNG method in this project is not only capable of generating random cryptographic keys efficiently but also tested in real time live on demand.

6 DISCUSSION

A secure communication is meant to be used by many users. For a large number of applications, especially for those intended for use on mobile devices, the use of dedicated, specialized hardware for RNG is not feasible, due to the cost, volume and power consumption limitations. A large number of applications may significantly benefit by the availability of independent random number inputs. Therefore, development of an innovative and efficient TRNG is an urgent prerequisite for current information security system.

In this research project, however, the true random number generator does not depend on a special device. This research project shall make use of already available devices on the shelf or common peripherals. It is the environment which becomes the source of randomness. Even the vacuum was once thought to be just an empty dark silent space. In fact, vacuum is an extent of space that has virtual sub-atomic particles spontaneously appearing and disappearing. It is the presence of these virtual particles that give rise to random noise. This 'vacuum noise' is omnipresent and may be exploited and used to generate random numbers (Symul *et. al.*, 2011).

7 CONCLUSION

True random key generator is the most crucial components of modern cryptosystem. For cryptographic use, however, it is important that the numbers used to generate any cryptographic keys are not just seemingly random; they must be truly unpredictable. In fact, each operation in cryptography requires a new fresh random key. A set of evaluation criteria of modern TRNG for cryptographic keys has been proposed in this paper. In this research project, a true environmental random ambience number generator has been developed upon which the recent TRNGs have been evaluated against. Air ambience is the natural choice here. A TRNG based on air ambience has a strong potential to perform well according to the quantitative evaluation criteria proposed in this paper.

8 ACKNOWLEDGMENT

The authors would like to express sincere appreciation to Universiti Teknikal Malaysia Melaka and Fundamental Research Grant Scheme FRGS/1/2015/ICT05/FTMK/02/F00293 funded by the Ministry of Higher Education, Malaysia for giving full technical and financial supports in this research.

REFERENCES

- Abu, N.A. and Sahib, S. (2010a). *Random Ambience Key Generation Live on Demand*, 2nd IEEE International Conference on Signal Processing Systems, Vol. 1, pp. 110-114. 5-7 July 2010, Dalian.
- Abu, N.A. and Sahib, S. (2010b). *One Megabit Random Ambience*, International Journal of Cryptology Research, Vol. 2, No. 1, pp. 073-087.
- Abu, N.A. and Sahib, S. (2011). *Random Ambience Using High Fidelity Images*, 3rd International Conference on Digital Image Processing, 15-17 April 2011, Proc. of SPIE (International Society for Optical Engineering), Volume 8009, 80092H, Chengdu.
- Ahmed, M. S. I. and Naganathan, E. R. (2008). *A Secured Key Generation Scheme Using Enhanced Entropy*, International Journal of Computer Science and Network Security, Vol. 8, No. 2, pp. 236-240, February 2008.
- Argyris, A., Deligiannidis, S., Pikasis, E., Bogris, A. and Syvridis, D. (2010). *Implementation of 140 Gb/S True Random Bit Generator based on A Chaotic Photonic Integrated Circuit*, Optics Express, Vol. 18, Issue 18, pp. 18763-18768, 18 August 2010.
- Bardis, N. G., Markovskiy, A. P., Doukas, N. and Karadimas, N. V. (2009). *True Random Number Generation based on Environmental Noise Measurements for Military Applications*, Proceedings of the 8th WSEAS International Conference On Signal Processing, Robotics and Automation, 21-23 February 2009, pp. 68-73, Cambridge.
- Barker, E. and Roginsky A. (2012). *Recommendation for Cryptographic Key Generation*, NIST Special Publication 800-133, December 2012.
- Blaszczyk, M. and Guinee, R. A. (2008). *A True Random Binary Sequence Generator based on Chaotic Circuit*, Signals and Systems Conference, 18-19 June 2008, pp. 294 – 299, Galway.
- Danger, J. L., Guilley, S. and Hoogvorst, P. (2009). *High Speed True Random Number Generator based on Open Loop Structures in FPGAs*, Microelectronics Journal, Vol. 40, No. 11, November 2009, pp. 1650-1656.
- Dinca, L. M. and Hancke, G. (2017). *User-Centric Key Entropy: Study of Biometric Key Derivation Subject to Spoofing Attacks*, Special Issue on Entropy-Based Applied Cryptography and Enhanced Security for Ubiquitous Computing, Vol. 19, pp. 70-91, 21 February 2017.

- Drutarovský, M. and Galajda, P. (2007). *A Robust Chaos-Based True Random Number Generator Embedded in Reconfigurable Switched-Capacitor Hardware*, Radio Engineering, Vol. 16, No. 3, pp. 120-127, September 2007.
- Dynes, J. F., Yuan, Z. L., Sharpe, A. W. and Shields, A. J. (2008). *A High Speed, Post-processing Free, Quantum Random Number Generator*, Applied Physics Letters, Vol. 93, No. 3, Article ID. 031109(3 pages), 25 July 2008.
- Ergün, S. and Özoguz, S. (2008). *Truly Random Number Generators Based On Non-Autonomous Continuous-Time Chaos*, International Journal of Circuit Theory and Applications, Vol. 38, No. 1, 31 July 2008, pp. 1–24.
- Fechner, B. and Osterloh, A. (2010). *A Meta-Level True Random Number Generator*, International Journal of Critical Computer-Based Systems, Vol. 1, No. 1-3, 2010, pp. 267-279.
- Fritschek R. and Wunder, G.(2017). *On-the-Fly Secure Key Generation with Deterministic Models*, IEEE International Conference on Communications (ICC), 21-25 May 2017, Paris, pp. 1-6.
- Fürst, M., Weier, H., Nauerth, S., Marangon, D. G., Kurtsiefer, C. and Weinfurter, H. (2010). *High Speed Optical Quantum Random Number Generation*, Optics Express, Vol. 18, No. 12, pp. 13029-13037, 2 June 2010.
- Hars, L. (2009). *Random Number Generators in Secure Disk Drives*, EURASIP Journal on Embedded Systems, Vol. 2009, Article ID 598246(10 pages), 9 June 2009.
- Hirano, K., Yamazaki, T., Morikatsu, S., Okumura, H., Aida, H., Uchida, A., Yoshimori, S., Yoshimura, K., Harayama, T. and Davis, P. (2010). *Fast Random Bit Generation with Bandwidth-Enhanced Chaos in Semiconductor Lasers*, Optics Express, Vol. 18, No. 6, pp. 5512-5524, 15 March 2010.
- Holcomb, D. E., Bursleson, W. P. and Fu, K. (2009). *Power-up SRAM State as an Identifying Fingerprint and Source of True Random Numbers*, IEEE Transactions on Computers, Vol. 58, No. 9, September 2009, pp. 1198-1210.
- Koeberl, P., Li, J., Rajan A. and Wu W.(2014). *Entropy loss in PUF-based key generation schemes: The repetition code pitfall*, IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), 6-7 May 2014, Arlington, pp. 44-49.
- Kwon, O., Cho, Y. W. and Kim, Y. H. (2009). *Quantum Random Number Generator Using Photon-Number Path Entanglement*, Applied Optics, Vol. 48, No. 9, pp. 1774-1778, 19 March 2009.
- Petrie, C. and Connelly, J. (2000). *A Noise-Based IC Random Number Generator for Applications in Cryptography*, IEEE Transaction on Circuits Syst. I: Fundamental Theory and Applications, Vol. 47, pp. 615–621.
- Pironio, S., Acín, A., Massar, S., de la Giroday, A. B., Matsukevich, D. N., Maunz, P., Olmschenk, S., Hayes, D., Luo, L., Manning, T. A. and Monroe, C. (2010). *Random Numbers Certified by Bell's Theorem*, Nature, Vol. 464, pp. 1021-1024, 15 April 2010.

- Santoro, R., Sentieys, O. and Roy, S. (2009). *On-the-Fly Evaluation of FPGA-Based True Random Number Generator*, Proceedings IEEE Computer Society Annual Symposium on VLSI ISVLS 2009, pp. 055-060.
- Stipcević, M. and Rogina, B. M. (2007). *Quantum Random Number Generator based on Photonic Emission in Semiconductors*, Review of Scientific Instruments, 9 April 2007, Vol. 78 No. 4, 045104, pp. 001-007.
- Sunar, B., Martin, W. J. and Stinson, D. R. (2007). *A Provable Secure True Random Number Generator with Build-In Tolerance to Active Attacks*, IEEE Transactions on Computers, Vol. 56, No. 1, January 2007, pp. 109-119.
- Symul, T., Assad, S. M. and Lam P. K. (2011). *Real Time Demonstration of High Bitrate Quantum Random Number Generation with Coherent Laser Light*, Applied Physics Letters: Lasers, Optics and Optoelectronics, Vol. 98, No. 23, 17 May 2011.
- Teh, J. S., Teng, W. J. and Azman Samsudin, A. (2016) A True Random Number Generator based on Hyperchaos and Digital Sound, 3rd International Conference on Computer and Information Sciences, pp. 246-269, 15-17 August 2016, Kuala Lumpur.
- Thamrin, N. M., Witjaksono, G., Nuruddin, A. and Abdullah, M.S., (2008). *A Photonic-based Random Number Generator for Cryptographic Application*, 9th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 6-8 August 2008, pp. 356-361.
- Tokunaga, C., Blaauw, D. and Mudge, T. (2008). *True Random Number Generator with a Metastability-Based Quality Control*, IEEE Journal of Solid-State Circuits, Vol. 43, No. 1, January 2008, pp. 078-085.
- Uz-Zaman, I., Lopez, A. B., Al Faruque, M. A. and Boyraz, O.(2017). *A Physical Layer Security Key Generation Technique for Inter-Vehicular Visible Light Communication*, Signal Processing in Photonic Communications, Article SpTu1F.3, New Orleans, 24–27 July 2017
- Wayne, M. A., Jeffrey, E. R., Akselrod, G. M. and Kwiat, P. G. (2009). *Photon Arrival Time Quantum Random Number Generation*, Journal of Modern Optics, Vol. 56, No. 4, pp. 516-522, February 2009.
- Wei, W. and Guo, H. (2009). *Bias-Free True Random-Number Generator*, Optics Letters, Vol. 34, No. 12, 11 June 2009, pp. 1876-1878.

A Practical SCADA Testbed in Electrical Power System Environment for Cyber-security Exercises

**Norziana Jamil^{*1,2}, Qais Qassim¹, Maslina Daud⁴, Izham Zainal Abidin³,
Norhamadi Ja'afar⁴, and Wan Azlan Wan Kamarulzaman⁵**

¹*Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Malaysia*

²*College of Computer Science and Information Technology, Universiti Tenaga Nasional, Malaysia*

³*College of Engineering, Universiti Tenaga Nasional, Malaysia*

⁴*CyberSecurity Malaysia, Malaysia*

⁵*Tenaga Nasional Berhad, Malaysia*

E-mail: Norziana@uniten.edu.my

**Corresponding author*

ABSTRACT

The impact from Stuxnet worm to SCADA systems in 2010 has been one of the most significant signals of a well-coordinated cyber-attack is now towards disrupting national critical infrastructures such as power grid governed by SCADA system. The discovery of this worm has put a lot of attention on the strength and security level of security countermeasures of existing critical infrastructure systems such as SCADA that has been long used as a legacy system. One way to assess the strength and security level of a system is through penetration testing and vulnerability assessment that would help in determining weaknesses, loopholes and potential breaches for exploitation in system defences. However, performing a real penetration test and vulnerability assessment in a real critical infrastructure system is infeasible and unlikely to happen because an unintended consequence that might occur can propagate its effect to a wider scale. On the other hand, a replicated system is also infeasible due to the high cost and huge effort required. Therefore, developing a realistic SCADA testbed is the best available option for the cyber-security exercise to take place. This paper describes in-detail a scalable and reconfigurable SCADA testbed for cyber-security analysis.

Keywords: SCADA security, Scada testbed, Vulnerability assessment, Electric power grid

1 INTRODUCTION

Supervisory Control and Data Acquisition (SCADA) system controls and monitors public infrastructures such as electrical power, oil and gas, manufacturing and transportation networks as

well as industrial processes (Stouffer et al., 2011). A SCADA system gathers and analyses data from industrial field instruments for real-time control and management to improve the performance of the industrial critical system operations as well as to provide a better protection to the utilised equipment (Stefanov et al., 2015). Due to its wide range of applications, many attacks have been targeting SCADA systems because the failure-to-operate could cause great financial loss and have serious impacts on public safety and the environment (He and Yan, 2016, Lin et al., 2017).

In the early history of SCADA systems it was widely believed that such systems were secure because they were physically and electronically isolated from other networks. However, Stuxnet attack incident demonstrated that the security by obscurity concept is no longer a valid approach for such systems (Karnouskos, 2011). Stuxnet attack crossed both the cyber and physical world by manipulating the control system of the critical infrastructure. During the last few years, a number of security incidents targeted highly sensitive facilities such as the cyber-attack that induced power outage in Ukraine's power grid in 2015 by compromising the corporate networks using spear-phishing emails with BlackEnergy malware (Liang et al., 2017), and the attack on German steel plant in late 2014 where hackers successfully took control of the production software and caused significant material damage to the site (Lee et al., 2014). These incidents have raised concern among cyber-security researchers. Therefore, many cyber-security agencies, providers and researchers have taken substantial initiatives to address SCADA systems vulnerabilities and their security loopholes as an effort to protect these systems and their control from security threats, attacks and malware (Stoian et al., 2014). However, activities that help in identifying security vulnerabilities and potential breaches, and investigating the effect of attack on an actual system is neither recommended due to the unintended consequences, nor feasible on a replicated system due to the cost and effort involved (Hahn et al., 2010, Singh et al., 2015). Therefore, SCADA cyber-security researchers mostly rely on developments of SCADA-specific cyber-security testbed which imitate a realistic SCADA system through emulating, virtualising or simulating SCADA devices, software and communication infrastructure to analyse the security and defence countermeasures of these systems. Such testbed would help researchers to build a better and a more resilient SCADA network architecture, to find methods to do penetration testing on live systems, to build tools to check intrusion, malware, and other vulnerabilities etc. The testbed is also aimed at training engineers with hands-on exercises on how to secure their SCADA systems.

SCADA communication protocols are seen as the weakest link in the cyber security analysis of these systems (Pidikiti et al., 2013), where protocol security is crucial for the functioning of the SCADA systems Maynard et al. (2014). Security researchers urge on that, protecting data in-transit should be essential part of system protection strategy since data will be moved back and forth from many locations (Al Baalbaki et al., 2013). Whereas, any disruption or modification occur to the communication link may result in loss of availability and/or integrity of the entire system. That suggests a vulnerability in the protocol implementation in the SCADA system may compromise the entire system. Therefore, cyber-security analysis for SCADA-specific protocols should be considered. In the work, the proposed SCADA testbed was designed to utilise one of the commonly used communication protocols namely: IEC 60870-5-104. This protocol was intentionally chosen because it is crucial for the communication between the control stations and distribution stations in many electrical power facilities around the world. For example, a simple search for IEC 60870-5-104 devices using Shodan service reveals that about 584 SCADA

servers are connected to the Internet and are remotely accessible.

The rest of this paper is organised into the following sections. Section 2 provides an overview of the electrical power grid SCADA system. Section 3 presents research works related to SCADA testbed implementation and design. Section 4 describes the SCADA testbed requirements. In Section 5 the proposed SCADA testbed for cyber-security analysis is presented. Finally, Section 6 concludes the work.

2 ELECTRICAL POWER GRID SCADA SYSTEM

Electrical power systems have been early adopters of SCADA systems for their operations, making them one of the earliest cases of cyber-physical systems (Shaw, 2006). They were designed to provide voltage and current levels, circuit breaker status information and other field related indicators in a real-time to identify problems as they occur and to take corrective actions when assistance is needed as an attempt to prevent significant system failures. In electrical power systems, SCADA is used to remote control and monitor the flow of electricity from generators to customer premises and factories through transmission and distribution subsystems (Chikuni and Dondo, 2007). SCADA systems enable an efficient power production and distribution of electric systems through the use of automated data collection and equipment control (Spellman, 2016). It improves the overall efficiency of the system for optimising, supervising and controlling the generation and transmission systems. SCADA function in the power system network provides greater system reliability and stability for integrated grid operation.

The architecture of electrical power SCADA system can be envisaged as three main areas as shown in Figure 1 (Stouffer et al., 2008). At the field devices area; sensors, relays and actuators provide an interface for control and monitoring of the physical process. Remote Terminal Unit (RTU) and Programmable Logic Controllers (PLC) are also reside in this area where they aggregate control (acting as master) for many field devices by passing commands and responses through the communications network to the SCADA server. On the other side, a dedicated control centre to govern the entire control and monitor process. The control centre commonly consists of SCADA application servers for process monitoring and control, database servers for historical record storage and in some cases interoperability servers for interconnecting SCADA software and hardware from different vendors. Moreover, the system's operator monitors the process state through Human-machine Interface (HMI) and controls the process by activating commands as required.

Generally, SCADA system could have multiple supervisory systems, PLCs, RTUs, HMIs, process and control instrumentation, sensors and actuator devices over a large geographical area, interconnected through a communications network. The communication network is intended to provide the means by which data can be transferred between the main control centre and field sites. Historically, SCADA networks have been isolated from other networks through the use of dedicated communication links and proprietary communication protocols (Stouffer et al., 2008). However, with the increased deployment of geographically distributed substations and economical consideration, SCADA systems are becoming increasingly interconnected, rapidly adapt internet enabled devices as well as open communication standards to significantly reduce infras-

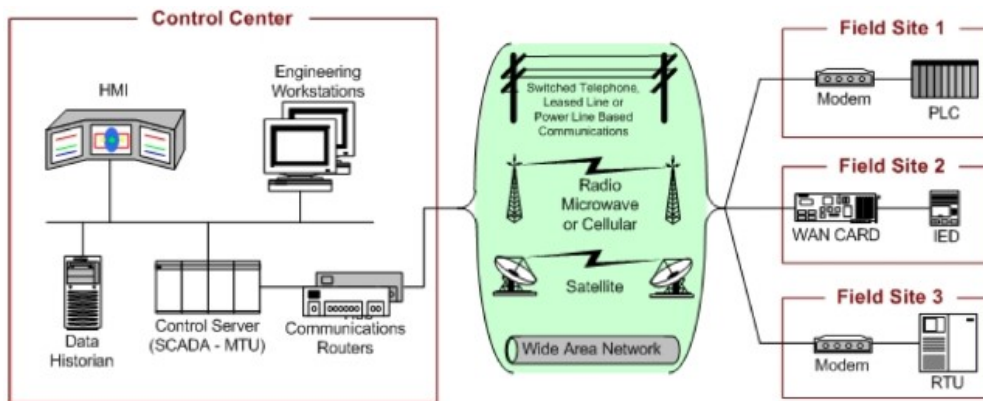


Figure 1: General architecture of a SCADA system (Stouffer et al., 2008)

structure costs and increase ease of maintenance and integration (Chalamasetty et al., 2016, Urias et al., 2012). Moreover, for better decision making and provide real-time updates utility companies have integrating their SCADA network with their enterprise networks (such as business and corporate networks) to streamline operations (Sridhar et al., 2012). Consequently, SCADA systems have to face different types of vulnerabilities and threats that are associated with cyber and physical devices, software as well as communication and control protocols (Singh, 2013). Therefore, identifying the vulnerabilities and threats and protecting SCADA systems is of vital importance.

3 RELATED WORKS

To ensure the highest level of protection of critical infrastructure systems driven and controlled by SCADA systems, the SCADA system has to go through multiple security tests including vulnerability assessment and penetration testing as well as security countermeasure evaluation. However, conducting such tests on a production system is technically difficult to audit without compromising and/or impact its reliability and performance (Poudel et al., 2017). Therefore, security researchers attempt to clone the SCADA systems in isolated environments where experiments can safely be performed. SCADA testbed allows cyber-security researchers and system engineers to at first investigate cyber-security vulnerabilities on functional control systems, then implement exploits to attack the systems in the testbed to understand the implications of the vulnerability. In this section, relevant published literature in terms of SCADA simulation and testbed implementation for SCADA cyber-security research is surveyed.

Several SCADA testbeds have been developed at various entities such as national labs, universities and research centres for purpose of studying the consequences associated with these cyber-physical threats and mitigate those consequences. At the national laboratories level, the Department of Energy, US, in 2003 has established a SCADA testbed at Idaho National Laboratory called National SCADA testbed (Laboratory, 2011, Ralston et al., 2007) it aims to provide testing, research and training facilities to help improve the security of SCADA control systems. Another effort developed by a national lab is by Sandia National Laboratory (Parks, 2007) which uses hybrid modelling and simulation architecture in order to understand the possible impact

of particular cyber threats, cyber defence training and exploring power system vulnerabilities. Comparable efforts have been made by Universities and research laboratories in the design and deployment of a realistic and reliable SCADA testbed for various applications especially in electrical power control system. For instance, a SCADA testbed has been developed at University of Arizona using PowerWorld simulation tool to simulate the operations of large scale power distribution systems to implement and test an anomaly-based intrusion detection system (Mallouhi et al., 2011). Another SCADA testbed example is implemented at power and energy research laboratory of Mississippi State University (Adhikari et al., 2012). The testbed was used for simulation of common power system contingencies (generator loss, transmission loss and sudden load loss), and event detection using data mining of phasor measurement unit data. Similarly, at Royal Melbourne Institute of Technology University, a SCADA testbed has been developed for building SCADA simulations which support combination of network simulation and real device connectivity (Queiroz et al., 2011).

Furthermore, University College Dublin have implemented a SCADA testbed for cyber-security practices using computer networks and power grids simulators (Stefanov and Liu, 2014). The testbed provides a tool for analysing cyber-physical vulnerabilities, allows monitoring of the dynamic behaviour of power system as response to cyber-attacks (impact analysis), and mitigation of cyber-attacks. The simulations of cyberattack were performed with IEEE 39-bus system and three attack scenarios including unauthorised access to control assets such as RTUs and Intelligent Electronic Devices (IEDs), denial of service by flooding the SCADA network, man-in-the-middle by modifying packets carrying measurements and control commands, configuration change of protective relays. One more SCADA testbed for cyber-security analysis was developed at Washington State University (Liu et al., 2015). The developed testbed was used to study the impact of three types of real-life cyber-attacks on IEEE14-bus test system controlled by SCADA. The testbed was designed using an integrated cyber-power modelling and simulation tools such that a real-time modelling of end-to-end cyber-power systems have been developed with hardware-in-the-loop capabilities. Real-time digital simulator, synchro-phasor devices, DeterLab, and network simulator-3 (NS-3) were utilised. The testbed was used to simulate Man-in-the-middle and denial-of-service attacks which were modelled in DeterLab.

Recently, a real-time hardware-in-the-loop based SCADA testbed were developed at South Dakota State University, US (Poudel et al., 2017). The testbed was developed to study the cyber-security and voltage stability control of power grid SCADA systems. The testbed utilises SEL 351S protection system with OPAL-RT including control functions and communications to build a cyber-physical environment. The study presented two different mitigation strategies using optimal power flow for system reconfiguration in order to restore normal or next steady state operating condition after failures. Additionally, different reconfiguration plans were presented for avoiding the cascading failures following any kind of cyber-attack.

4 SCADA TESTBED REQUIREMENTS

A trustworthy and accurate simulation results require a testbed that is able to reproduce the real system as accurately as possible. The fidelity concept defined by Siaterlis and Genge (2008) ensures the implementation of an adjustable level of realistic SCADA testbed through the use

of real hardware devices when its really required rather than emulators, simulators or other abstraction methods. Another key requirement for a reliable testbed design is repeatability (Gao et al., 2013). This requirement reflects the need to repeat the experiment and to obtain the same or statistically consistent results. Prior to conducting the experiment, the researcher has to define clearly and in-detail the experiments initial and final states as well as all events in between the two states. According to Hahn et al. (2010), a reliable testbed requires accurate measurements; an accurate conduct of experiment should be maintained. In other words, researchers should not interfere with the experiment in such a way that they might alter the experiments outcome. Safe execution of tests is another requirement defined in Siaterlis and Genge (2014) and has been a focus area for most of the testbeds intended for cyber-security exercise (Holm et al., 2015). In most of the cases, the experiments for cyber-security exercise assume the presence of an adversary who utilises malicious software and injects malicious code and command or traffic into the testbed to reach the desired goal. The effect of these activities can be unpredictable as it is difficult to predict the outcome of these activities beforehand which may have disruptive effects on physical systems. Such cases need to be carried out in an isolated and controlled environment to ensure the safety of physical devices as well as to protect security researchers from potential danger resulting from simulating an attack or executing a malicious activity.

In the previous work by Qassim et al. (2017), several SCADA testbed implementation approaches were evaluated based on the SCADA testbed requirements. The study showed that, to address the given requirements and to come up with a realistic testbed for cyber-security exercise, SCADA testbed should utilise a combination of both virtualisation and physical replication in a hybrid environment that ensures a high degree of fidelity. Therefore, virtual-physical replication is used to overcome the limitations of both virtualisation and physical replication approaches.

5 IMPLEMENTATION OF THE PROPOSED SCADA TESTBED

This section describes the architecture of the proposed electrical power SCADA testbed. The experimental setup is as shown in Figure 2.

It includes several SCADA key components such as real-time digital simulator to emulate the power system, master and local HMI for monitoring and control, an RTU to control the emulated physical system and several engineering workstations for testing and analysis purposes. As illustrated in Figure 2, four functional levels (or zones) have been considered in the design of our SCADA testbed, which are: process, bay, communication and stations levels. The functional levels denote for the hierarchical levels of power system management based on IEC-62264 reference model for industrial communication networks. The following subsections highlight the testbed components at the different functional levels.

5.1 Implementation of process level

In this testbed, OPAL-RT model OP5600 is used to simulate the electrical power system. OPAL-RT is a versatile and high-performance real-time digital simulator; it is a platform equipped with

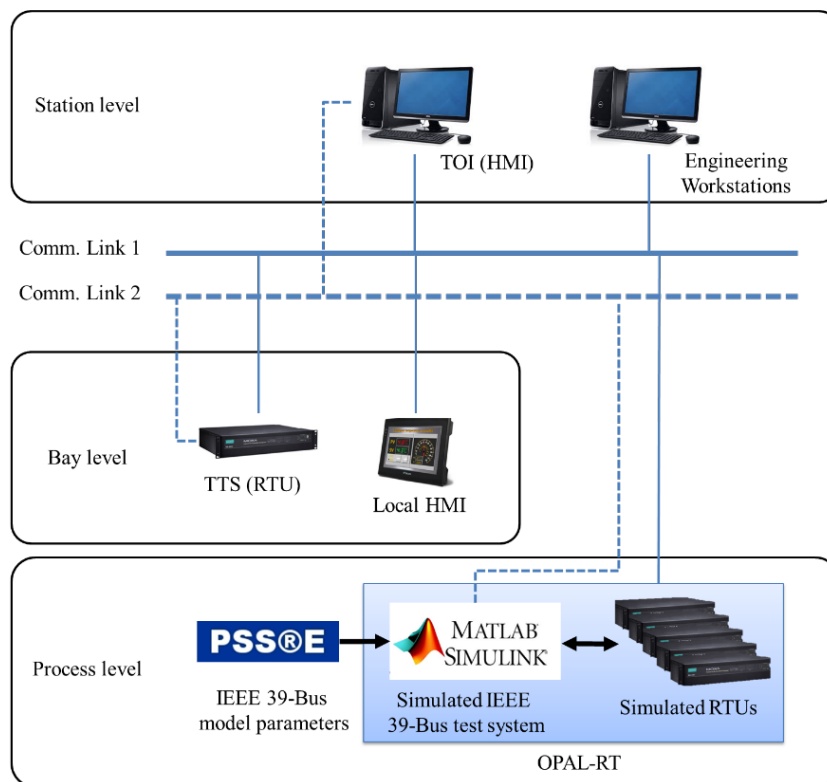


Figure 2: The implemented SCADA testbed

analoge and digital I/O that provides the capability to perform real-time experiments by interfacing with real hardware devices such as RTUs, IEDs and protection relays. In demonstrating this testbed, the IEEE 39-bus electrical network test system is used to validate the implemented testbed and provide a meaningful example for other researches, experiments and applications of this infrastructure. The IEEE 39-bus electrical network is simulated by OPAL-RT to offer a real-time response and feedback.

The IEEE 39-bus network consists of 10 generators, and 46 transmission lines as shown in Figure 3. This particular network is chosen because it is a standard based network used by power system researchers all over the world and serves as a benchmark in result comparison. The electrical network is modelled in Power System Simulator for Engineering (PSS/E or PSSE) software to obtain a simulation parameters to be loaded into MathWorks Simulink software application. PSS/E is a software tool used by power system engineers to simulate electrical transmission networks in steady-state conditions as well as over timescales of a few seconds to tens of seconds. PSS/E is used by most power system utilities around the world to run and test their power system network. It has the ability to run from basic power flow simulations right up to complex dynamic simulations. However, the software requires huge computing power to run large scale dynamic simulations. Hence the reason why OPAL-RT was chosen as the computing engine to enable this.

For the IEEE 39-Bus network to run in the OPAL-RT environment, the configuration files are loaded up into the OPAL-RT Solver which is integrated with MathWorks Simulink software application. The OPAL-RT Solver serves as the engine to run the power system network. Since

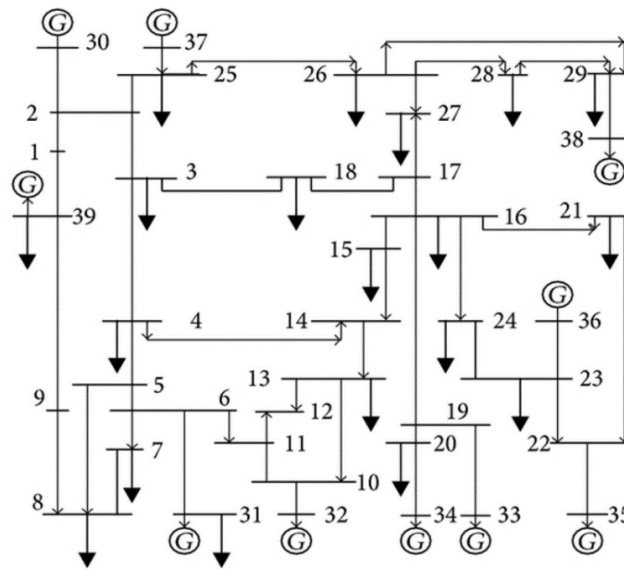


Figure 3: IEEE 39-bus test system network

OPAL-RT is configured to run on multiple processors, this enables it to run the real-time digital simulation with ease.

5.2 Implementation of bay level

For the SCADA testbed, a new proprietary software application were developed called Tenaga Terminal Simulator (TTS). The TTS is used for modelling and simulating an industrial control system device such as an RTU. Instead of having a physical RTU to be attached to the testbed, the RTU will be replaced with the TTS software which can produce data as if it is an RTU. At the heart of the TTS architecture is a Real-time Database (RTDB), which is an in-memory, event-driven, real-time database. The RTDB can run on any Portable Operating System Interface (POSIX) compliant operating system. RTDB consists of freely configurable tag items that can be associated with different drivers for communication and applications usage. Among the communication protocols supported by RTDB are:

- C37.118 (for communication with synchrophasor)
- IEC 60870-5-101 (for communication with SCADA devices such as RTU)
- IEC 60870-5-104 (for communication with SCADA devices such as RTU over TCP/IP network)
- IEC 61850 (for communication with IEDs in substation automation environment)

5.3 Implementation of communication link

The communication network for controlling and monitoring the power system, namely the SCADA network has to be simulated as close as possible to the real substation SCADA network. In this work, the SCADA testbed was designed to simulate the IEC-60870-5-104. IEC 60870-5-104 (also known as IEC104) is one of the IEC 60870 set of international standards released by the IEC (International Electrotechnical Commission) which define systems used for tele-control in electrical engineering and electrical power system automation applications (Chalamasetty et al., 2016). It specifies a communication profile for sending basic tele-control messages between two systems over standard TCP/IP network. The usage of TCP/IP network offers simultaneous data transmission between several devices and services (Lee et al., 2014). Apart from this, the security of IEC 60870-5-104 has been proven to be problematic, according to recent security advisories (Laboratory, 2011, Ralston et al., 2007), multiple issues in this protocol such as the lack of proper data encryption could allow an unauthenticated, remote attacker to spoof network communications or exploit input validation flaws on vulnerable systems using the affected protocol. IEC has published a security standard (IEC 62351), the security of IEC tele-control protocol series that implements authentication of data transfer and end-to-end data encryption. The implementation of IEC 62351 would prevent common cyber-attacks such as replay, man-in-the-middle and packet injection attacks. However, due to the increase in complexity and the limited processing capabilities of existing SCADA devices, vendors are reluctant to employ these countermeasures on their devices and/or networks.

5.4 Implementation of station level

At the station level, a web-based application platform have been developed to simulate the function of a SCADA HMI. Tenaga Operator Interface (TOI) is the codename of our HMI software application. TOI is a real time interface system that collects and displays information about power system parameters to the substation operators. It also offers control capability whereby immediate action can be taken should there be anomalies detected in the power system. TOI provides various communication ports for fast communication and convenient control of a diverse range of machines, systems and facilities. The TOI was developed for real-time visualisation and control of power system network. Warning alarms can be visualised in dynamic symbols or retrieved in tabular format. TOI increases operational awareness of substation operators in a manned substation allowing pre-emptive corrections to be performed to avert catastrophe. TOI complements the SCADA network control centres capability by monitoring operations of substation equipment remotely as well as locally.

5.5 Simulation process

The integration of OPAL-RT, TOI and TTS, which involves ensuring data are communicated properly between all three systems, is tested by simulating a case following these steps: The IEEE 39-bus model is loaded into OPAL-RT.

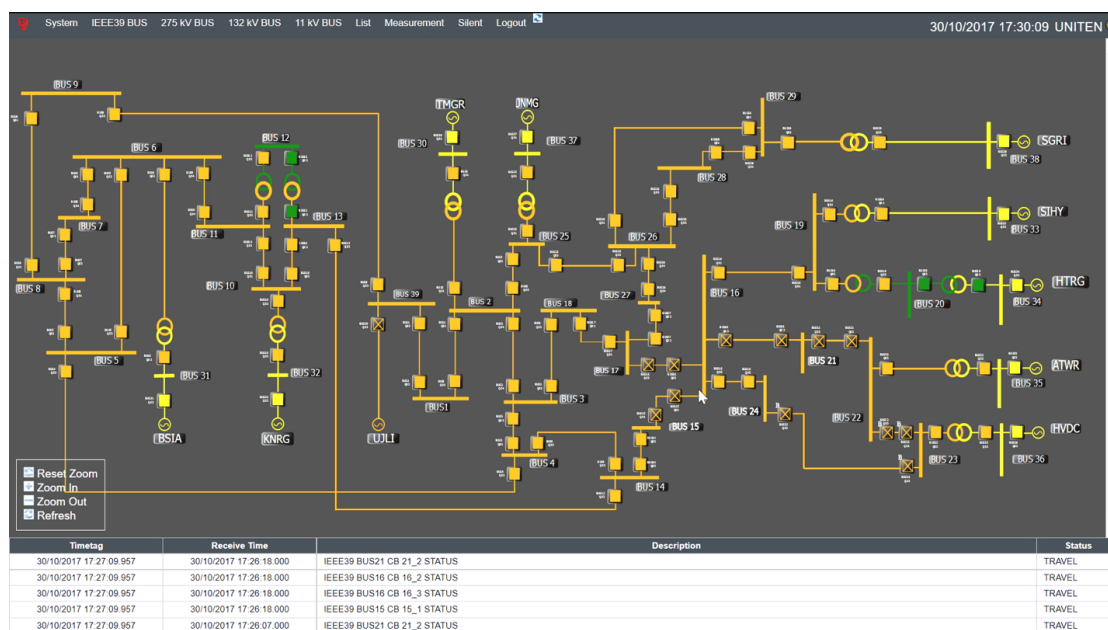


Figure 4: Operator Interface

1. Once the loading and initialization process is successful, simulation is executed which will generate all the system parameters reflecting the power system condition.
2. These parameters are then sent to TTS and TOI machine via TCP/IP protocol.
3. The operator interface will show the status of the power system i.e. breaker open, breaker close and the analog parameters such as voltage, frequency etc.
4. Any changes made either in OPAL-RT itself, or through system control in TOI, or external injected data via TTS will be reflected in TOI display and OPAL-RT console. For example, if the breaker is closed, the box in Figure 4 at the affected bus will turn to empty box, highlighted box or crossed-box to notify status of open, closed and traveling (error status), respectively.

6 CONCLUSION

The proposed testbed in this work offers highest fidelity and accuracy. This is because the real-time simulator offers an excellent representation of the actual power system with the support of a reconfigurable environment. Moreover, the testbed has been designed to host several substations to accurately model the scale of real electric power grid system. An operator shall control the emulated substations from separate computer system through Ethernet based network to mimic real SCADA system implementation. The proposed testbed supports a wide range of experimental applications such as security related studies including vulnerability assessment, risk analysis, cyber-security evaluation and digital forensic investigation as well as other control related applications. This is due to the scalability feature of the proposed testbed environment. The proposed testbed also supports repeatability in order to obtain the same or statistically similar results. A

researcher/tester could easily reproduce a previously tested experiment since the initial state and set up of the testbed environment is fixed at the start of each simulation and is a user-defined choice. In addition to the power system security, the proposed testbed is capable of conducting control experiments in real time. Therefore, it is a valuable tool for simulating different events in the realistic cyber-physical environment. In future works, the implemented testbed will be used to demonstrate the impact of the cyber-attack on the physical power grid in terms of voltage stability and loss of generation.

ACKNOWLEDGMENTS

This research study is supported by Ministry of Science, Technology and Innovation, Malaysia and CyberSecurity Malaysia through DSTIN project 2016-2018.

REFERENCES

- Adhikari, U., Morris, T. H., Dahal, N., Pan, S., King, R. L., Younan, N. H., and Madani, V. (2012). Development of power system test bed for data mining of synchrophasors data, cyber-attack and relay testing in RTDS. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–7. IEEE.
- Al Baalbaki, B., Al-Nashif, Y., Hariri, S., and Kelly, D. (2013). Autonomic critical infrastructure protection (acip) system. In *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, pages 1–4. IEEE.
- Chalamasetty, G. K., Mandal, P., and Tseng, T.-L. (2016). Secure SCADA communication network for detecting and preventing cyber-attacks on power systems. In *Power Systems Conference (PSC), 2016 Clemson University*, pages 1–7. IEEE.
- Chikuni, E. and Dondo, M. (2007). Investigating the security of electrical power systems SCADA. In *AFRICON 2007*, pages 1–7. IEEE.
- Gao, H., Peng, Y., Dai, Z., Wang, T., and Jia, K. (2013). The design of ics testbed based on emulation, physical, and simulation (eps-ics testbed). In *Intelligent Information Hiding and Multimedia Signal Processing, 2013 Ninth International Conference on*, pages 420–423. IEEE.
- Hahn, A., Kregel, B., Govindarasu, M., Fitzpatrick, J., Adnan, R., Sridhar, S., and Higdon, M. (2010). Development of the PowerCyber SCADA security testbed. In *Proceedings of the sixth annual workshop on cyber security and information intelligence research*, page 21. ACM.
- He, H. and Yan, J. (2016). Cyber-physical attacks and defences in the smart grid: a survey. *IET Cyber-Physical Systems: Theory & Applications*, 1(1):13–27.
- Holm, H., Karresand, M., Vidström, A., and Westring, E. (2015). A survey of industrial control system testbeds. In *Secure IT Systems*, pages 11–26. Springer.

- Karnouskos, S. (2011). Stuxnet worm impact on industrial cyber-physical system security. In *IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society*, pages 4490–4494. IEEE.
- Laboratory, I. N. (2011). Vulnerability Analysis of Energy Delivery Control Systems (Report No. INL/EXT-10-18381). Technical report, Idaho National Laboratory.
- Lee, R. M., Assante, M. J., and Conway, T. (2014). German steel mill cyber attack. *Industrial Control Systems*, 30:62.
- Liang, G., Weller, S. R., Zhao, J., Luo, F., and Dong, Z. Y. (2017). The 2015 ukraine black-out: Implications for false data injection attacks. *IEEE Transactions on Power Systems*, 32(4):3317–3318.
- Lin, C.-T., Wu, S.-L., and Lee, M.-L. (2017). Cyber attack and defense on industry control systems. In *Dependable and Secure Computing, 2017 IEEE Conference on*, pages 524–526. IEEE.
- Liu, R., Vellaithurai, C., Biswas, S. S., Gamage, T. T., and Srivastava, A. K. (2015). Analyzing the cyber-physical impact of cyber events on the power grid. *IEEE Transactions on Smart Grid*, 6(5):2444–2453.
- Mallouhi, M., Al-Nashif, Y., Cox, D., Chadaga, T., and Hariri, S. (2011). A testbed for analyzing security of SCADA control systems (TASSCS). In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1–7. IEEE.
- Maynard, P., McLaughlin, K., and Haberler, B. (2014). Towards Understanding Man-in-the-middle Attacks on IEC 60870-5-104 SCADA Networks. In *ICS-CSR*.
- Parks, R. C. (2007). Guide to Critical Infrastructure Protection Cyber Vulnerability Assessment (Report No. SAND2007-7328). Technical report, Sandia National Laboratories, Albuquerque, New Mexico.
- Pidikiti, D. S., Kalluri, R., Kumar, R. S., and Bindhumadhava, B. (2013). SCADA communication protocols: vulnerabilities, attacks and possible mitigations. *CSI transactions on ICT*, 1(2):135–141.
- Poudel, S., Ni, Z., and Malla, N. (2017). Real-time cyber physical system testbed for power system security and control. *International Journal of Electrical Power & Energy Systems*, 90:124–133.
- Qassim, Q., Jamil, N., Abidin, I. Z., Rusli, M. E., Yussof, S., Ismail, R., Abdullah, F., Ja'afar, N., Hasan, H. C., and Daud, M. (2017). A survey of scada testbed implementation approaches. *Indian Journal of Science and Technology*, 10(26).
- Queiroz, C., Mahmood, A., and Tari, Z. (2011). SCADASim – A framework for building SCADA simulations. *IEEE Transactions on Smart Grid*, 2(4):589–597.
- Ralston, P. A., Graham, J. H., and Hieb, J. L. (2007). Cyber security risk assessment for scada and dcs networks. *ISA transactions*, 46(4):583–594.
- Shaw, W. T. (2006). *Cybersecurity for SCADA systems*. Pennwell books.

- Siaterlis, C. and Genge, B. (2008). Cyber-physical testbeds: Scientific instruments for cyber security assessment of critical infrastructures.
- Siaterlis, C. and Genge, B. (2014). Cyber-physical testbeds. *Communications of the ACM*, 57(6):64–73.
- Singh, H. P. (2013). Cyber security trend in Substation Network for automation and control Systems. In *Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on*, pages 1–3. IEEE.
- Singh, P., Garg, S., Kumar, V., and Saquib, Z. (2015). A testbed for SCADA cyber security and intrusion detection. In *Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC), 2015 International Conference on*, pages 1–6. IEEE.
- Spellman, F. R. (2016). *Energy Infrastructure Protection and Homeland Security*. Bernan Press.
- Sridhar, S., Hahn, A., and Govindarasu, M. (2012). Cyber attack-resilient control for smart grid. In *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES*, pages 1–3. IEEE.
- Stefanov, A. and Liu, C.-C. (2014). Cyber-physical system security and impact analysis. *IFAC Proceedings Volumes*, 47(3):11238–11243.
- Stefanov, A., Liu, C.-C., Govindarasu, M., and Wu, S.-S. (2015). SCADA modeling for performance and vulnerability assessment of integrated cyber-physical systems. *International Transactions on Electrical Energy Systems*, 25(3):498–519.
- Stoian, I., Ignat, S., Capatina, D., and Ghiran, O. (2014). Security and intrusion detection on critical SCADA systems for water management. In *2014 IEEE International Conference on Automation, Quality and Testing, Robotics*, pages 1–6.
- Stouffer, K., Falco, J., and Scarfone, K. (2011). Guide to industrial control systems (ICS) security. *NIST special publication*, 800(82):16–16.
- Stouffer, K., Scarfone, K., and Falco, J. (2008). Guide to Industrial Control Systems (ICS) Security. Technical report, Technical report, September.
- Urias, V., Van Leeuwen, B., and Richardson, B. (2012). Supervisory Command and Data Acquisition (SCADA) system cyber security analysis using a live, virtual, and constructive (LVC) testbed. In *Military Communications Conference, 2012-MILCOM 2012*, pages 1–8. IEEE.